Metaethics as Conceptual Engineering

Knut Olav Skarsaune

Metaethics has usually been conceived as a *descriptive* project: its goal to clarify what our normative judgments mean, explain how they can be epistemically justified, and describe the nature of the properties (if any) they pick out. Some authors move the goalposts ever so slightly: finding ordinary normative thought metaphysically problematic, they put their theories forward as *revisions* of our existing concepts. But they stress that they are only proposing a *gentle* reform, lest they be accused of changing the subject. Their goal is to defend concepts that are *a*) metaphysically sound, while otherwise *b)* remaining as faithful as possible to ordinary meanings.[1]

I'll call these "the descriptive project" and "the tethered revision project", respectively. There are of course differences between them, but in this paper those differences will matter less than what they have in common: that they take faithfulness to our existing discourse to be a strong desideratum on theory choice, whether it can be satisfied completely or only partially. They try to give the semantics, metaphysics, etc., of our *existing* normative discourse, or, if that is found to have unacceptable metaphysical commitments, of the nearest thing that doesn't.

In this paper I will lay out a third kind of metaethical project. This project is suggested, with varying degrees of detail, in the works of Frankena, Stevenson, Rawls, Brandt, Kahane, Eklund and – outside of metaethics – in the general approach to philosophy of Carnap and Quine.[2] Throwing in with an increasingly common terminology, I will call it "conceptual engineering".

There is one downside to using this name. In the literature on conceptual engineering, it is sometimes understood as an inherently revisionary project, such

---

[1] This paper is a descendant of a part of my PhD dissertation from 2013. Over the decade it has spent in peer-review hell since then, I have received input and encouragement from too many people to list, but I want to thank them all, and especially Matti Eklund and David Plunkett.

[2] Frankena (1958: 80), Stevenson (1963: 11), Rawls (1971: 111), Brandt (1979: ch. 1), Kahane (2012), Eklund (2012), (2015) and (2017), Carnap (1950: ch. 1) and Quine (1960: §53).

that taking an engineering approach to the philosophy of X by definition means that you'll be proposing revisions to our concept of X, in order to promote intellectual progress, or social justice, or whatever. I'd like to plant a big red flag and emphasise right away that that is not how I use the term. The way I intend it, it stands instead for a way of thinking about the *question* we are asking. As I will defend it, the engineering approach is *neutral* between descriptive and revisionary theories. Indeed, that is its main attraction: it provides a unifying conception of the question metaethics is asking, such that descriptive views, gently revisionary views, and radically revisionary views can all serve as competing answers to it.

I will develop conceptual engineering as an approach to metaethics on which the present meanings of our normative terms (and the properties they pick out) have no privileged status, neither as targets, simply, of the enquiry, nor as tethers on how far we can go without changing the subject. To be sure, we will want to understand, as deeply as possible, the psychology and semantics of our present normative concepts, the metaphysics they commit us to, and their means of epistemic justification. But we will also want to understand the psychology, semantics, metaphysics and epistemology of other possible normative concepts. The ultimate point of all this is to choose between the available alternatives: to *decide* what kinds of normative concepts to use, going forward.

My goal is to make this project precise, and thereby show that it is coherent. I will also try to explain how it is suited to address the kinds of questions and concerns that lead a lot of people to engage in metaethical enquiry in the first place, and I will point out some limitations of the two traditional projects. But I will not argue that they are mistaken or misguided. The engineering project need not be a *competitor* to the traditional projects, since it is possible to find both them and it important and worthwhile. I see no point in fighting over which of them is *the* metaethical project, or which has a better claim to the name "metaethics" (all three are represented in the literature a good librarian would sort under that label).[3] What I want to accomplish is to distinguish the three from each other, and to show that the engineering project is coherent and well-motivated. Not to knock the others.

Before we start, a brief defence of talking about concepts in the first place. There is a popular view on which the best way to think of philosophy is simply as asking *object-language* questions, like "What is knowledge?" or "What is

---

[3] See e.g. the references in footnotes 2, 7, and the references to Lewis, Gibbard and Railton in section I. See also footnote 21.

time?".[4] Philosophers investigate this or that part of reality, not our words or concepts. Can we not likewise think of metaethics as simply asking "What is value?" or "What are reasons?".

Perhaps we can, if we take a certain kind of realism for granted. But the metaethical debate includes views like prescriptivism and expressivism, and certain kinds of subjectivism, relativism and contextualism, which cannot be fully stated in the object-language. My aim is to develop a conception of metaethical enquiry that is neutral between a wide variety of views, including these. For this reason, it seems to me useful to make the "semantic ascent" and talk about concepts, as well as the properties (if any) that they pick out. This should be acceptable to all sides. For a realist, the questions "What is value?" and "What is the nature of the property (if any) picked out by the concept VALUE?" are equivalent.[5] So she should have no objection to making the semantic ascent, and she should be willing to make it, out of courtesy to the expressivist.

Even if a non-question-begging discussion of the aims of metaethics requires making some kind of ascent, why an ascent to talk of *concepts*, in particular? Why not talk about *claims*, or *terms*, or *utterances*, or *thoughts*, or *mental states*? Different philosophers will have different preferences about what to ascend to, when a dialectical situation forces an ascent from the object language. I have no particular bone in such fights, but to keep this paper readable, a choice must be made; and any choice one might make here would irritate somebody. There is a tradition in metaethics for using "judgments" in an ecumenical sense, a sense in which everyone (including non-cognitivists) can agree that we make normative judgments. My talk of concepts is intended in a similar ecumenical sense, a sense in which everyone can agree that *a*) we make normative judgments and *b*) they have something that at least looks like a predicate position. By "normative concept" I'll mean whatever goes in there. Readers who think "concept" is an ill-chosen word for it can translate into their preferred idiom throughout.

I.   MOTIVATING THE ENGINEERING PROJECT

The opening pages of textbooks in metaethics tend to be remarkably similar: we start with some examples of normative disputes, say about abortion or severe interrogation of people who know the locations of ticking bombs. As we dig deeper into such questions, the author suggests, it is natural to start to wonder

---

[4] See e.g. Williamson (2007, ch. 2)

[5] C.f. Schroeter and Schroeter (2015: p. 419).

about the nature of the questions themselves. In this way, we are introduced to a list of "metaethical questions", which usually includes the following:[6]

Semantics: What do normative words like "good", "right", "ought" and "reason" mean?

Psychology: What kind of mental state are normative judgments?

Metaphysics: Do normative properties and facts exist, and if so, what is their nature?

Epistemology: How can normative judgments be justified, or constitute knowledge?

These four questions also tend to come up, outside of textbooks, when metaethical writers explain how they understand their field.[7] Let us call them, collectively, the "textbook questions".

I use "normative" and not just "moral" or "ethical" in stating these questions, because most contemporary writers think moral, prudential and epistemic normativity give rise to similar philosophical questions, making it increasingly common to demarcate the field around normativity, not morality. But there are also writers whose theories concern morality alone.[8] Some of these, in turn, primarily mean, by "morality", the *social practice* and the associated social norms actually found in society (as opposed to, say, "the part of the normative that has to do with how to treat others"). I should warn that readers who think of metaethics as the meta-theory of morality-the-social-practice may find much of what I will say strange or irrelevant. This is simply a decision point: this paper is about metaethics understood as the meta-theory of normativity, and not just of morality (I return to this issue in section VI).

The textbook questions are descriptive, and present tense. They ask about the semantics, metaphysics, etc., of our present, actual normative discourse. Now one may wonder whether writers who officially present their views as answers to

---

[6] Miller (2003: 2 and 2010: Introduction); Fisher (2011: 2); Davis (2007: 9-10); Kirchin (2012: 6-19); Van Roojen (2015: 2-5).

[7] Timmons (1999: 11); Darwall, Gibbard, Railton (1992: 125-6); Schroeder (2008: 6); Hussain and Shah (2006: 268), Berker (*manuscript*: 48-9). McPherson and Plunkett (2017) suggest that what *unites* these four questions is the project of "explaining how actual ethical thought and talk – and what (if anything) that thought and talk is distinctively about – fits into reality" (*ibid*: pp 3, 19). This characterisation also fits well with what I call "the descriptive project" in the text.

[8] See e.g. Richard Boyd (1988), Frank Jackson (1998: chs. 5-6) and David Brink (1989).

these questions really intend their views *only* as descriptions of our current discourse and its subject matter, or *also* as recommendations about what kinds of concepts to use, going forward. To my eye, many writers who present their views as descriptive of ordinary normative thought are also engaged in *defending* ordinary normative thought, *as they have described it*, against philosophical objections.[9] This suggests that they not only hold that their theories are true of our *present* discourse (the descriptive project), but also that we should *continue* to use the concepts their theories describe (the engineering project).

I suspect, in other words, that a considerable amount of what I will call "conceptual engineering" has been going on in metaethics all along, especially if one is willing to read a bit between the lines. *To the extent that this is so*, this paper's proposal will be to understand and describe what we are doing in a new way, not to do something new. That does not seem to me to make the issue less important. But in any case, I do not want to argue, one way or the other, about the degree to which actual, descriptive metaethical theories are meant *only* as answers to the textbook questions, or also as endorsements or recommendations going forward. That would require a lot of exegesis and explode the length of the paper. I will therefore leave this interpretive issue to the reader's judgment, and proceed to discuss three idealized, clear-cut metaethical projects, in abstraction from the question of how exactly they distribute in the literature.

Let us, then, consider the descriptive project in its pure form. It seeks to answer the textbook questions, read literally, concerning our existing normative concepts and the properties they pick out. It does *not* ask whether we should continue to use those concepts.

I will not argue that there is anything wrong with this project in itself. I see nothing wrong with wanting to understand the semantics, metaphysics and so on of our current normative discourse, either for its own sake, or indeed as a precursor to the engineering project. My concern is rather that the descriptive project, on its own, leaves out something important. The point is not original: one could even call it the standard motivation for conceptual engineering. Here is how Matti Eklund puts it, his target being the corresponding approach to philosophy in general:

> I think that already a little reflection shows how – what's the best
> way to put it? – *navel-gazing* this kind of study is. The concepts we
> have are the ones we have ended up with due to various biological

---

9 See e.g. Nagel (1997: ch. 6), Parfit (2011: chs. 31-4), Scanlon (2014: chs. 2,4), Blackburn (1998: Appendix), Enoch (2011, chs. 6-9).

and cultural factors. By some measure they have proven themselves. But still, why should the concepts we actually have be the best conceptual tools for describing and theorizing about the relevant aspects of reality?

[…] the point is easily restated in terms of properties and relations: why study the properties and relations that our ordinary concepts stand for rather than the most theoretically important properties and relations in the vicinity? (Eklund 2015, 376)

Adapting this general point to the case at hand: our normative concepts are hand-me-downs from previous generations. They are shaped by our biology and by the material and social conditions under which our predecessors have sought to keep the peace with each other and make the grain last through the winter. By some measure they have proven themselves. But still, I can see no reason to assume, at the *outset* of enquiry, that they cannot be improved upon.

To be sure, there are difficult questions about what exactly it would mean to "improve upon" our normative concepts, especially the "thin" ones (like IMPROVE UPON or BETTER). This will occupy us in sections II through V. But grant for now that the idea of improving our normative concepts makes sense. In that case it seems strange to decide in advance that all we want from metaethics is to understand the psychology, semantics and epistemology of our *present* normative concepts, and the metaphysics of the properties they pick out. Why shouldn't we also be at least as interested in knowing which normative concepts we *should* be using, *their* metaphysics, and so on?

So far, then, for situating and motivating the engineering project vis-à-vis the descriptive project. Now a complication. The literature also contains a second kind of approach, which is a kind of halfway house between the descriptive project and conceptual engineering. It is to propose a conceptual revision, while insisting that the revision is only a gentle one, and indeed that it is *small enough* for the post-revision discourse to count as a continuation of the pre-revision discourse. Here, for example, is David Lewis:

There are no perfect deservers of the name ["value"] to be had. But there are plenty of imperfect deservers of the name, and my [theory] is meant to capture what it takes to be one of the best of them. […] Strictly speaking, Mackie is right: genuine values would have to meet an impossible condition, so it is an error to think there are any. […] Then is my position a form of realism about values? – Irrealism about values strictly speaking, realism about values loosely speaking. The

former do not exist. The latter do. (1989: 136-7)

Allan Gibbard:

> Norm-expressivism is meant to capture whatever there is to ordinary
> notions of rationality if Platonism is excluded. […] My hope, then, is
> to save what is clear in ordinary thought about rationality, and to find
> our reflective thinking about rationality reasonably clear and fully
> rectifiable, with one exception: our wavering penchant for Platonism.
> (Gibbard, 1990. pp. 154-5)

Both Lewis and the Gibbard of *Wise Choices, Apt Feelings* propose gentle revisions motivated by metaphysical concerns. They think ordinary normative though is committed – at least sometimes, or with some part of itself – to properties that do not in fact exist. But they think they can defend something *quite close* to ordinary normative thought that lacks commitment to those properties.

Once we propose a revision, though, why is it important that it be gentle? Why is a smaller revision preferable to a bigger one? Here is Peter Railton:

> Revisionism may reach a point where it becomes more perspicacious
> to say that a concept has been abandoned, rather than revised. No
> sharp line separates tolerable revisionism and outright abandonment,
> but if our naturalist wishes to make his case compelling, he must
> show that his account of a person's good is a rather clear case of tolerable revision, at worst. (1989: 159)

He explains what he means by "tolerable revision" via an example: a proposed naturalistic reduction of "seaworthiness" in terms physical properties, i.e. the disposition to stay afloat and manoeuvrable in foul weather. Such a reduction would capture *most* of the traditional notion, but some of it – such as having been properly blessed – would be left out.

> Would such a revision change the subject? That seems unlikely. The
> sort of revision involved is typical of changes that have taken place
> in a very wide range of properties that figure in commonsense theorizing. (*ibid*. 165-6)

The thought is that a revisionary proposal should stay close enough to ordinary discourse to avoid "changing the subject".[10] The existing concepts act as a kind

---

[10] Railton's discussion of "tolerable revision" sometimes collapses this requirement and another

of *tether* on how radical a revision can be. Otherwise, it will no longer count as a revision of the target discourse, but rather as a recommendation to abandon it and replace it with a new one.

Like before, I will not argue that there is anything wrong with this "tethered revision" project in itself. If one thinks ordinary normative discourse (at least sometimes, or with some part of itself) is committed to properties that do not exist, I see nothing wrong in wanting to understand what the smallest revision would be which is not so committed. But again I will suggest that it seems strange to think that this is all we should want from metaethical enquiry.

My worry is a simple extension of the corresponding point about the descriptive project. Once we are clear that we are not just investigating our present concepts (or the properties they pick out), but advocating for concepts to use going forward, why should we still insist that they must be *almost* the same as the old ones? Why assume that the right concepts *must* be found *nearby* to our navel? Suppose a more radical revision – even so radical as to count as an abandonment-and-replacement – would have other advantages over a smaller revision, why should we prefer the smaller one?

However, the worry about changing the subject remains. As Strawson (1963: 505) emphasized, philosophical questions arise out of ordinary, everyday thought. We might be puzzled, for example, about how moral responsibility is possible, or about what to do in cases where duty seems to conflict with self-interest. If a philosopher responds by suggesting that we drop that kind of talk, and to speak a new language she has invented instead, we might feel our questions have been avoided rather than answered.

To put the point in a different way, we expect a metaethical theory to be *a theory of value* (or reasons, or oughts, etc.). If someone proposes something completely different from our present concept of value (or reasons, etc.), then haven't they just offered us a (potentially interesting) theory about *something else*? In other words, if our existing normative language does not even act as a tether on metaethical theories, then what does?

The answers to these questions will lie in offering an alternative account of

---

one: that of being able to take up the *role* of the existing concept – in the evaluative case, of being able to serve the action-guiding role of evaluative concepts (1989: 162). But officially this is treated as a separate requirement called "normative role". As we shall see, I completely agree with *this* part of Railton's view. But the "normative role" requirement does not support the desideratum that the revision must stay close to the original, which is at issue in this section. In effect, my view will be that Railton should drop the "tolerable revision" requirement and make the "normative role" requirement do all the work.

what it is for something to be a metaethical theory, one that is not anchored in our present normative words or concepts, but something else. In other words, it will be to identify, in some other way, the "subject" that is not to be changed.

II. THE CONCEPTUAL ENGINEERING APPROACH

Suppose we say that to take an engineering approach to the philosophy of X is to ask, not what our concept of X is (or what X-ness is), but what concept of X we *should* be using (or which property in the X-vicinity is most worthy of study). Metaethics will then stand out as a rather special case, for the question "Which normative concepts should we use?" exhibits a peculiar circularity. Since the question contains a normative term ("should"), the answer may be hostage to the term we use in stating it. Suppose there are, in the vicinity of "should", two candidate concepts SHOULD$_1$ and SHOULD$_2$, belonging to the normative$_1$ and the normative$_2$ family of concepts, respectively. Then it seems possible that we should$_1$ use the normative$_1$ concepts but should$_2$ use the normative$_2$ concepts.

Such possibilities lead Matti Eklund (2015) to worry that the would-be conceptual engineer in metaethics will need to posit an "ineffable further question": it is as if we want to ask which normative concepts we **should** use, in some "privileged" sense of "should". But we don't know what this privileged sense is, and hence how to state the question: in fact, which sense is the "privileged" one is exactly the question we are trying to ask.

We should distinguish two circularity worries. One is *evaluative*, the worry being that how we evaluate different options may depend on which concepts we start with. I will discuss this issue in section IV. But first we must address a prior, *constitutive*, challenge about the question we are asking in the first place. The worry is that, if we think metaethics is about which normative concepts we *should* use, then thinkers who begin the enquiry with different normative concepts are not even asking the same question.

My response to this challenge will have two steps. In this section, I will describe an engineering question that is not circular. To preview a bit, I will suggest that we simply ask which normative concepts to use. A question to which the answer is not a proposition, of the form "We should use such-and-such concepts", but rather a *decision* to use one or another set of concepts. This avoids circularity, but might look deflationary instead: as if it is just a matter of *picking* a set of concepts. That brings us to the second step, in the next section, where I will try to describe – in a way that gives us back most of the traditional debates in metaethics – some ways a choice of concepts can be correct or mistaken.

But let us first take a step back. The engineering approach will make better

sense if we first consider the point of doing metaethics to begin with. What do we need this enquiry for?

Everyone should speak for themselves here, of course. But I will take the liberty of describing some motivations one might have. The common denominator is that they involve finding our normative judgments problematic or puzzling in some respect. Exactly what the source of worry or puzzlement might be is not important for my purposes, but it does matter whether one finds *something* about our normative judgments puzzling or problematic. If not, one will probably not see any point in the conceptual engineering approach.

One possible source of puzzlement concerns the meanings of normative terms. We think that some things are good and that some other things are bad, but on reflection, it is not evident exactly what we are thinking when we think that. What is this goodness and badness supposed to be? Unless they accept a certain kind of religious view, most people will be hard pressed to say. *Definitions* are of course often hard to come by, but when we say that something is square, or Italian, or hypoallergenic, we can if asked provide at least a rough explanation of what we mean. But we may be hard pressed to provide even a rough explanation of what we mean by "good" or "bad". The concern, in other words, is that our normative terms are obscure or opaque to us.

A related possible worry is metaphysical. It is unclear how normative properties might fit in our overall worldview. This is especially true if we aspire to a *naturalistic* worldview, which takes its cues from empirical science. Where are values, rights and reasons in the world that science describes?

Another possible worry is epistemological. It is clear that our normative beliefs are strongly shaped by the family and culture we grew up in, and also by our biological nature, which is shaped by Darwinian natural selection. Unless we are prepared to accept some form of relativism, the worry arises that we may have the normative beliefs we have for reasons that have nothing to do with their being true.

Furthermore, suppose we are alert to the influence of upbringing, culture, and biology on our normative outlook. Now we are ready to turn the page and critically assess the beliefs these contingent factors have left us with. But how, exactly, are we to do that? What *other* ground do we have to stand on, from which to criticise our initial normative beliefs?

These, then, are examples of motivations for metaethical enquiry. One might worry that our normative concepts are obscure; that their place in our (perhaps naturalistic) overall worldview is unresolved and problematic, or that their epistemology is a riddle. Think of this as the "before" state.

The goal of metaethics, on the engineering approach, is to replace this state with one where we have a clear understanding of what our normative judgments mean, their place in our total theory (naturalistic or otherwise) is resolved, and we see how they can be epistemically justified. Think of this as the ideal "after" state.

Importantly, this could be achieved by either a descriptive or a reforming theory. The conceptual engineering approach is *neutral* between these options.[11] We could develop a theoretical understanding of the concepts we already have, work out their place in our total theory (naturalistic or otherwise), and find that we can indeed see how to provide epistemic justification for them. Or we could replace our old concepts with new ones that fit nicely in our total theory, that have an intelligible epistemology, and whose meaning is clear to us. These possibilities are end-points of a spectrum; in between are various degrees of revision, where we reach our goal working largely with the concepts we already have, but clarify or sharpen them in some way.

On the engineering approach, then, the point of metaethical enquiry is to achieve "clarity and decision" in our normative thought, as Frankena (1958: 80) suggests. It is not to answer a set of "second order" questions about our present normative judgments, holding them constant, but to challenge them and decide whether to keep or to change them.

But wait. Isn't that the job of first-order ethics? Does the engineering approach collapse the distinction between metaethics and first-order ethics?

Yes and no. Let us pretend that all normative judgments are of the simple form *x is φ*. "*x*" represents the subject position; the position where we refer to the various things we have normative beliefs about. So *the Iraq war*, *winning that research grant*, *progressive taxation*, and so on, might take the place of *x*. "*φ*" represents the predicate position; the various normative concepts we apply. So *good*, *bad*, *right*, *wrong*, *just*, *unjust* etc. might take the place of *φ*.

The difference between ethics and metaethics, on the engineering approach, is the difference between worrying about the subject, *x*, and worrying

---

[11] This might seem to be in tension with the engineering metaphor. But in fact, engineers do not always build or change things. A road engineer might investigate, say, a bridge, and conclude that "This thing is fine and does not need repair". If the *question* she asked was "Is this bridge safe?", and the work she did was to examine its foundations, calculate material fatigue, etc., then it seems natural to say she was doing an engineer's job.

about the predicate, φ, in this schema. In first-order ethics, we put whatever worries we might have about our normative concepts on hold, and put them to use, asking which things to apply them to. Is progressive taxation just, or is linear taxation, or a poll tax? Is it right to pull the trolley's lever?

Metaethics as conceptual engineering is the reverse enquiry. We put on hold questions about which things to evaluate positively and negatively, which actions to proscribe, prescribe and permit, and ask which concepts to do it with.[12]

I don't mean to suggest that these enquiries can be kept completely apart. There seem to be relations of mutual attraction and repulsion between views in the two areas. Some first-order views seem to require a heavily intuitionistic epistemology, others less so. Some metaphysical views create pressure in the direction of value monism, others sit comfortably with value pluralism. And so on. On the engineering approach, ethics and metaethics will be continuous enquiries that can sometimes be kept apart, sometimes not.

Let's return to the circularity issue. The constitutive challenge was that, if we take the engineering question to be of the form "What kind of normative concepts should we use?" then we court circularity and the implication that thinkers who begin the enquiry with different concepts are asking different questions. The solution I have proposed is that we simply ask what kind of normative concepts to use. But of course, we do not just want to conclude with a decision to use this or that set of concepts, we also want philosophical understanding of the concepts we'll be using. So a fuller statement of the question would be: "What kind of normative concepts to use, and what are their semantics, psychology, epistemology and metaphysics?".

That question is grammatically clumsy, so it may be clearer to say instead what *theory acceptance* comes to, on the engineering approach. On this approach, to accept metaethical theory T is to *decide* to make first-order normative judgments of the kind that T describes (whether one accepts T as descriptive or reforming), and to make them in full consciousness, that is, with the second-order belief that they have the psychology, semantics, metaphysics and epistemology T says they have.

But a crucial part of the plot is still missing. If we approach metaethics as conceptual engineering, then it becomes a bit of a puzzle how the different theories would be in competition with each other. If we evaluate metaethical theories, in effect, as policies for first-order belief, then why not accept all of them: why

---

[12] I will provide a concrete example to illustrate this shortly (five paragraphs down).

not have first-order beliefs of all these kinds?

Consider an example. Suppose we think suffering is bad. I pick this example because it seems safe to assume that we will evaluate suffering negatively. So we can hold fixed that we are going to make a judgment of the form *suffering is φ*, where φ is a concept of negative appraisal. But we are at a loss to explain exactly what the content of our original judgment, *suffering is bad*, really is; or we worry about how to fit it into our overall metaphysics, or we worry about its epistemic status.

So we turn to metaethics. We study the different views, we study all the arguments for and against them, and arrive at a conclusion about which theory is overall most credible (more on this in section V). Then we adopt the concept described by that theory, and use it to evaluate suffering negatively. We might do this while thinking the chosen concept is the same as, or almost the same as, or quite different from, our old concept BAD.

For example, we can believe, with David Lewis (1989), that suffering is bad$_{Lewis}$: that if we were to vividly imagine suffering, we would desire to desire not to suffer.

Or we can believe, with Sharon Street (2008), that suffering is bad$_{constructivist}$, which means that our evaluative attitudes, if made fully coherent and informed of all the empirical facts, would condemn it.

Or we can judge, with Allan Gibbard (2003), that suffering is bad$_{expressivist}$. This judgment is really a *plan*; a plan to count suffering as a drawback when we deliberate what to do. In other words, whenever a possible course of action would produce suffering, we plan to count that as an argument against that course of action.

Or we can believe, with G. E. Moore (1903), that suffering is bad$_{Moore}$; that suffering has the nonnatural, irreducibly normative property badness.

All these options, and more, are available. But now we might scratch our head and wonder: why do we have to choose just one of them? Why not believe that suffering is bad$_{Lewis}$, *and* believe that it is bad$_{constructivist}$, and bad$_{expressivist}$, and bad$_{Moore}$? In fact, now that they have been brought to our attention, it seems quite obvious that we *should* believe all these things, except perhaps the last one. Of course we desire to desire not to suffer. Of course our idealised evaluative attitudes condemn suffering. And of course we share Gibbard's anti-suffering plan.

It would seem that, if we evaluate metaethical theories as policies for first-order judgment, then there is no general reason why we could not accept all of

13

them. And yet we have a strong sense that metaethical theories are in competition. They are in some sense supposed to be theories *of value* (or reasons, etc.), and there cannot be seventeen different, true theories of value. We are thus brought back to the "changing the subject" worry that we left outstanding at the end of section I. The conceptual engineer needs to explain what, (if not its relation to our present normative discourse), distinguishes a given theory *as a theory of value*, as opposed to just a theory about a certain kind of second-order desire, or evaluative attitude, or planning state, or something else. I believe this is why we still cling to the textbook questions, and in particular the semantical one. It can seem like metaethical theories *must* claim to be analyses of ordinary normative thought and talk (the descriptive project), or to *almost* be that (tethered revision), otherwise triviality threatens.

The challenge, then, if we leave the textbook questions behind and become conceptual engineers, is to explain exactly how metaethical theories are in competition with each other, and what distinguishes them as metaethical theories. I suggest the answer is this: *They are competing designs for a limited area of psychological real estate*. We can identify – in a way that is neutral between theories – a distinctive set of psychological roles, and evaluate the different theories as competing proposals about which concepts to use in these roles.

Which roles? Different normative concepts have different roles, of course, but the common denominator is that they serve to regulate action, emotion and attitude. Guy Kahane explains it as follows:

> What is this role? It is not easy to spell out in entirely neutral terms, but the basic idea is simple: it's the role of setting a standard by which attitude and action can be made intelligible and justified, and in light of which we deliberate (in the first-person), and give advice and criticise (in the second- and third-person). (Kahane 2012: 153)

Value judgments supervise desire and preference: if I want something but on reflection think it is no good, I think my desire is misguided. Likewise with moral judgments and moral emotions. If I am angry at someone, or feel guilty for something, but also think that what they or I did was not wrong, I will think my anger or guilt is uncalled for. And likewise with ought- and reason-judgments and intentions. If I think I all-things-considered have most reason to do something, I will try to get myself to do it. If I think you all-things-considered ought to do something, and I am an honest advisor, I will advise you to do it. And so on.

Competition for these psychological roles, then, is how the various metaethical theories conflict, on the engineering approach. There is nothing to

stop us from believing that a state of affairs, say that some swamp is drained and used for residential construction, is bad$_{\text{Lewis}}$, but good$_{\text{constructivist}}$, and indifferent-$_{\text{Moore}}$. But all these beliefs cannot regulate our preference about what is to happen, or what advice we give our friend on the zoning board. Whichever does that is the one that plays the relevant psychological role.

Since there is more than one of these roles (guiding action, preference, guilt and anger, and so on, respectively), one could conceivably decide to mix and match concepts from different metaethical views to play the different roles. For example, it is a conceivable policy to conduct one's practical reasoning in terms of GOOD/BAD$_{\text{CORNELL}}$, RIGHT/WRONG$_{\text{MOORE}}$, and REASON$_{\text{EXPRESSIVIST}}$. Since my proposal is to evaluate metaethical theories as policies for first-order judgment, I am committed to this combination of concepts corresponding to a metaethical view. And indeed it does. It corresponds to *Cornell-realism-about-value-and-nonnaturalism-about-morality-and-expressivism-about-reasons*, a view still waiting for its first defender.

Thinking of metaethics as conceptual engineering brings to light further possible views that have escaped attention so far. For example, it is a conceivable policy to conduct one's practical reasoning with expressivist concepts Monday through Friday and with realist concepts on the weekend. As before, I am committed to this policy corresponding to a metaethical view. And it does: we can call it *workday-expressivism-and-weekend-realism*.

What the conceptual engineering approach implies is that these and other exotic policies for action-and-attitude-guiding judgment correspond to logically possible metaethical views. That is no objection. They are logically possible views. A separate question is whether they are any plausible. I will briefly suggest why I suspect the answer is "no", but the case for the engineering approach does not depend on this answer. On the contrary: if approaching metaethics as conceptual engineering were to lead to the discovery of new and promising views worth exploring, so much the better.

My main worry about mixing and matching is that it is hard to see what the motivation for such a solution would be. Metaethical enquiry, I suggested, is driven by certain worries about our normative judgments. And whether one would emphasise the metaphysical, epistemological or opaqueness worries that I raised, or yet other ones, it seems that such worries arise in much the same way for all the different kinds of normative judgments. Likewise, it seems that the considerations in favour of making up one's mind one way or the other are similar in the different cases (and constant throughout the week). Why would one conclude, for example, that there are mind-independent facts about what to do (and use concepts that reflect that), but that there are no mind-independent facts about

how to feel (and use concepts that reflect this), or vice versa?

There is more to say about this, but since the case for the conceptual engineering approach does not turn on the issue, we can leave it for another time. In what follows, I will for simplicity's sake speak as if the main options are to adopt concepts of the same metaethical kind to play all the action-and-attitude-guiding roles.

Another interesting implication of the engineering approach is that an Error Theory, on its own, will not count as a metaethical theory. To have a complete metaethical view, the error theorist must also say what she proposes we do once we have discovered the error. And to the best of my knowledge, all error theorists do indeed proceed to do just that.[13]

III.   CAN METAETHICAL THEORIES BE TRUE OR FALSE?

The core of the engineering proposal is now on the table. In the rest of the paper, I will develop it further by responding to six worries.

The first has to do with truth and objectivity. In the previous section, I proposed that we simply ask which normative concepts to use. A question to which the answer is not a proposition, but a decision to use a certain set of concepts. While this avoids circularity, it might look deflationary instead: as if the goal is just to achieve "clarity and decision" by plumping for some set of concepts or other. In the same vein, it might seem to make metaethics a subjective affair. If concept family A looks good to me, while B looks good to you, it is not obvious why there would be any pressure for either of us to think that the other is making a mistake, or to debate the issue.

In fact, however, we tend to think of metaethics as a search for truth, and of metaethical theories as things that can be true or false. And the way we debate

---

[13] Most error theorists are only *moral* error theorists. Of these, Richard Joyce (2001) recommends that we continue using moral concepts in a *fictionalist* way; John Burgess (2007) recommends that we stop using moral concepts and instead talk about our subjective or communal standards and ideals. John Mackie (1977) is open to interpretation on this point, but he seems to recommend that we continue to use moral language in a relativist sense.

Jonas Olson (2014) and Bart Streumer (2017) defend full-blown *normative* error theory. Streumer thinks it is impossible to fully believe the error theory; the best we can achive is a partial belief, with a corresponding lowering of our credences in our first-order beliefs. Olson recommends that we combine belief in the Error Theory in "critical contexts" (such as philosophy seminar rooms) with a continued use of traditional normative concepts in everyday contexts. He calls this view "Moral Conservationism", but on my approach it would be more perspicuously called "Seminar-room-Nihilism-and-everyday-perfunctory-Realism".

suggests that we think this truth is objective, and thus the same for everyone. Can we make sense of this, on the engineering approach?

I will suggest that the answer is largely Yes. And that in the limited respect in which the answer is No, that is a plausible answer. Even though concept choices cannot themselves be true or false, there are other ways in which they can be correct or mistaken. Reflection on these ways will, as it were, give us back most of the traditional debates in metaethics.

One way a choice of concepts can be a mistake is if it traps us in superstition: belief in non-existent things. Suppose for example that we accept something like G. E. Moore's view. Whether or not we think common sense is so committed, we achieve clarity and decision in our normative thought by deciding that *we* accept the existence of non-natural normative properties. We go on to form action-and-attitude-guiding beliefs that ascribe such properties. But suppose, further, that reality does not cooperate: there aren't any. Our normative beliefs are then bound to be systematically false, just like Mackie said.

Conversely, a choice of concepts can be a mistake if it leads us miss out on important truths. Suppose for example that we accept a constructivist view. Whether or not common sense agrees, we make up our minds that *we* think the highest authorities there are for our normative beliefs to answer to are some suitably idealized versions of our own normative attitudes. We go on to conduct our practical reasoning using concepts defined in these terms. But suppose, further, that some kind of normative realism is true. Perhaps Moore was right after all, or perhaps Aristotelian teleology is true, or Divine Command Theory. Whichever way, reality itself supplies a standard for our actions and attitudes.[14] In that case,

---

[14] Eklund (2017, ch. 1) challenges this and other common ways of defining metanormative realism. Adapted to the terms I am using, his challenge would be this: the suggestion of reality supplying a "standard" for actions and attitudes uses *our* concept of a standard. And it does not seem like we are entitled to assume that our concept (and the property it picks out) is "privileged" over other possible concepts of a standard (and the properties they pick out). What if reality supplies a standard, but also supplies a standard* and a standard**, for actions and attitudes?

This is primarily a challenge for the would-be realist. But since I am appealing to the possibility that realism might be true (or false) in order to explain how concept choices can be correct or mistaken, it is derivatively also a challenge for me. Very briefly, I would suggest that the realist should take up the solution Eklund offers her, and include some kind of "sparseness" assumption in her realist view (*ibid*, p. 59). The realist's metaphysical claim must include that the standard she posits for actions and attitudes has no relevant rivals. She asserts the existence of exactly one standard, and the non-existence of any other, similar things.

That normative realism needs some kind of sparseness assumption is not a new observation. One of the first points made in one of the first works we have in metaethics, the *Euthypro*, is

---

17

we would be conducting our practical reasoning with concepts that fail to pick out the properties that are the actual standards of practical reasoning. Our action-and-attitude-guiding beliefs would not be *false* (recall the point about suffering being bad$_{Lewis}$, bad$_{constructivist}$, etc.), but they would miss the point.

The possibility of these two kinds of error – superstition on the one hand, and missing out on important truths on the other – means that metaethical enquiry will in large part be a search for truth, also on the engineering approach. We can think of theory choice in two stages. First sort the metaethical views into groups closed under agreement about what reality is like. Between these groups, the engineering approach will see theory choice as a matter of trying to get things right, of choosing concepts that fit reality better.

However, there are also metaethical views that *agree* about what reality is like. For the conceptual engineer, the choice *between such views* will look different than it does on the traditional approaches. For example, there is no disagreement about what reality is like between subjectivism, constructivism, dispositionalism, and other views that in one way or another take value to be dependent on our attitudes. Since they agree about what reality is like, the only thing left for them to disagree about would seem to be semantics, the meaning of normative language. On the traditional approaches, the correct view is the one that is faithful to ordinary meanings (the descriptive approach), or at any rate less unfaithful (tethered revision).

But if we are asking which normative concepts to use going forward, then it is not clear why the least revisionary of these views would have to be the best. There may be arguments or considerations that favour a more revisionary view. Suppose for example that, among these views, the one that departs least from ordinary meanings is also vulnerable to David Enoch's argument from disagreement or Sharon Street's Darwinian Dilemma,[15] while the runner-up is not. Then the engineering approach implies that we may well have reason to accept the more revisionary view. This will not, however, be a matter of one view being false and the other true; rather it will be a matter of one view offering a better way forward than the other.

If reality itself does not set a standard for our actions and attitudes, then

---

that Divine Command Theory crashes if there are *many* gods, who love conflicting things. And it seems important both to Plato and to Aristotle, in developing their respective theories of Forms, that the Forms be sparse: otherwise they could not do the explanatory work they do.

[15] Street (2011) argues that Blackburn and Gibbard's "Quasi-Realism" is vulnerable to the Darwinian Dilemma. For Enoch's argument, see his (2011, ch 2).

there will not on the engineering approach be one uniquely correct metaethical theory. Realist views that say there is such a standard will be mistaken, of course, and all the non-realist views will be correct in the sense that they are right about what reality is like. But there can still be reasons to prefer some of them over others – reasons having to do with epistemology, first-order implications, internal coherence, and so on. But these will be considerations about *how to move forward*, not evidence of being the correct answer to some descriptive question.

This may be unsettling, but to me it seems like a plausible prediction. If there is no external, objective standard for our actions and attitudes; if it is up to us to "create" value – to "gild or stain" the world with our sentiments – then I can see no reason to expect there to be a unique, objective answer about how we should go about it, or which concepts we should do it with.

IV.   EVALUATIVE CIRCULARITY

This talk of ways in which concept choices can be better or worse may have set off some alarm bells. For what does "better" mean in that sentence? Does it mean *better$_{Moore}$* or *better$_{constructivist}$* or…?

I distinguished earlier between a constitutive and an evaluative circularity worry. We have now come to the latter. As discussed in the last section, and to be further discussed in the next, the engineering proposal is not to just plump for some set of concepts that strike our fancy; it is to seek a *warranted* decision, the one that has the *best arguments* and the *strongest reasons* in its favour. But these evaluative and epistemic concepts are of course themselves normative. And suppose that there is more than one possible family of normative concepts; that more than one metaethical theory describes concepts that are capable (psychologically speaking) of playing the action-and-attitude-guiding roles (more on this in section VII). Then it seems possible that different metaethical theories might be warranted, depending on which concepts we start out with. Perhaps constructivism has the best$_{constructivist}$ arguments, while expressivism has the best$_{Lewis}$ arguments, and so on; every permutation is conceivable. But then it seems possible that two thinkers entering the enquiry with different normative concepts, and who each reason flawlessly, assessing all the same arguments, can be justified in reaching different conclusions.

The engineering approach has this implication, but I do not see that as an objection, because I think this is a predicament we happen to be in. We are indeed, to use a fine cliché, like sailors on Neurath's boat, who must repair their ship at sea.

This is to be expected when we put our most basic concepts in play, and

entertain the possibility of revising them. Extreme versions of this phenomenon can arise in basic logical debates.[16] Suppose, for example, that a defender of logic A has arguments for A that are valid in A but not in logic B, while a defender of B has arguments for B that are valid in B but not in A. Given their different starting points, it seems they are doomed to evaluate the arguments differently and stay in their respective corners.

Fortunately, the situation does not seem to be as dire as that in the metaethical case. I will discuss methodology and theory choice further in the next section, but for now we can note that the argumentative force of many arguments in the literature does not seem to hinge on which normative concepts a given thinker employs. Consider for example David Enoch's argument that a range of subjectivist, response-dependence and expressivist views face an unpleasant choice between moral spinelessness and denying the equal importance of others (2011, ch. 2). Suppose a given thinker assesses this argument as valid, in the sense that Enoch's reasoning from subjectivist, response-dependence and expressivist premises to the unpleasant choice goes through. In that case the argument will, for her, have force against the targeted theories just in case she rejects both spinelessness and denying equal importance, but it does not seem to matter whether she rejects them by taking them to be bad$_{Moore}$, or bad$_{constructivist}$, or bad$_{Gibbard}$, etc.

Likewise with arguments that appeal to things like metaphysical parsimony, the ability to explain how normative judgments can be reliable or justified, the ability to explain the supervenience of the normative on the descriptive, *et cetera*. In such cases too, so long as a given thinker holds that parsimony, the ability to explain reliability, supervenience, etc., are desiderata on metaethical theory choice, it does not matter which normative concepts she herself uses. What matters for theory choice are the desiderata she accepts, not which concepts she uses to accept them.

Granted, there are also cases where conceptual starting points do make a difference. But I do not think this can be held against the engineering approach, because the same will be true on any methodology. If people start with different normative concepts, they are likely to (even supposed to) have different linguistic intuitions, and will come to different verdicts about which view fits best with naive normative thought and talk. So on the descriptive and tethered revision approaches, which both emphasise this desideratum, differences in conceptual starting points is likely to have a strong effect on theory choice. If anything, this effect may be weaker on the engineering approach.

---

[16] C.f. Burgess and Plunkett (2013: 1107).

V.   DOES THE PROPOSAL MAKE ANY METHODOLOGICAL DIFFERENCE?

Which leads to the following question. Suppose we take up the engineering pro-ject. We think of theory choice as a choice of concepts to use in practical reasoning. Does this shift in approach have any implications for how we should actually conduct the enquiry?

Here is a brief description of what seems to be the standard view about methodology. Like elsewhere in philosophy, the language of "benefits" and "costs" has taken root in metaethics. What we need, an emerging consensus says, is a *holistic* methodology.[17] Since every known view has some weakness or other, we should examine how each view performs with respect to a wide range of de-siderata, and see which one does the best job overall. Which desiderata? That is in principle open: what we should want from metaethical theories is itself open to debate, but in practice there is considerable agreement on the desiderata (though disagreement about their respective weights). It would be hard to com-pile a comprehensive list, but the following are widely recognized desiderata:

- Fit with linguistic evidence.
- Fit with moral phenomenology, and other psychological evidence.
- The ability to explain the motivational role of normative judgments.
- Ontological parsimony in general, and (for many) compatibility with met-aphysical naturalism in particular.
- The ability to explain how normative judgments can be epistemically jus-tified.
- The absence of problematic first-order normative implications.
- General theoretical virtues: explanatory power, coherence and simplicity.

It is generally agreed that no known theory performs perfectly in all these re-spects, so theory choice depends on a holistic assessment of the overall benefits and costs of the available alternatives.

I fully endorse this holistic methodology. But if we approach metaethics as conceptual engineering, it seems to me that there are implications for how we should think about "benefits" and "costs". To the degree that these implications are independently plausible – as it seems to me that they are – we can consider that a further consideration in favour of the engineering approach.

First we can look at the use of linguistic evidence. It is generally accepted in the current debate that theories lose plausibility in proportion to how far they

---

[17] Frankena (1958: 80-81), Shafer-Landau (2003: 234), Street (2009: 294), Enoch (2011: 267), McPherson (2012: 232), Väyrynen (*forthcoming*), Suikkanen (*manuscript*).

are from being analyses of ordinary normative discourse. Therefore, for example, much recent discussion of expressivism has focused on the "negation problem".[18] It is assumed that, if it can be shown that negation would work differently in an expressivist language than it in fact does in natural language, then expressivism suffers a huge loss in plausibility; perhaps it has been refuted outright.

But on the engineering approach, this assumption would not necessarily hold. *If expressivism offers a viable way forward* from our current intellectual predicament, then it would not be a big objection if we would need to learn a new negation symbol. It *would* be a serious problem if rigorous reasoning is impossible in an expressivist language (since we presumably want our practical reasoning to be rigorous). But it would be OK if rigorous reasoning is possible, but *different* in some way, from how it works in descriptive language.

In general, the significance of linguistic evidence will be different, and more indirect. It will not be a simple matter of losing plausibility in proportion to the amount of recalcitrant linguistic evidence. On the engineering approach, the big question for any metaethical theory is whether it offers us a form of action-and-attitude-guiding belief we can endorse in full consciousness, going forward. The relevance of linguistic evidence is in helping us to see the difference (if any) between our current beliefs and prospective future beliefs, thus helping us see what (if anything) we would be giving up. In some cases like, say, the univocality of negation, I suggest that even a big linguistic difference would not be much of a problem. In other cases, a seemingly small difference, for example whether it is conceptually possible that one's most stable normative beliefs are mistaken (see Egan (2007)), may bring out differences we cannot live with.

Secondly, psychology would play a different role than it does in the descriptive and tethered revision approaches. Again the common assumption seems to be that theories win plausibility in proportion to how faithful they are to the phenomenology of normative judgment, and other psychological evidence. But on the engineering approach, as with linguistic evidence, the important thing will not be the size of any psychological differences, in itself, but rather whether or not we are willing to give up the relevant features of our current judgments.

There is also the question of what kinds of changes we are *able* to make in the first place. On the engineering approach, metaethics should be equally concerned with which moral psychologies are *possible*, as with which are actual. In this role, psychology is a source of absolute constraints, not just plausibility. If a

---

[18] Unwin (1999); Dreier (2006)

22

given metaethical theory describes a form of judgment that cannot play the action-and-attitude-guiding roles for creatures with our biology, then the view is unavailable, regardless of its attractiveness in other respects. For example, suppose our present normative judgments are straightforward beliefs, with representational content. We might have some control over these contents, in the sense that we can revise our concepts, but it would seem more difficult to replace them with mental states of a different *kind*, like motivational states. Or vice versa: suppose normative judgments are in fact desires or preferences; it might be hard to convert them into beliefs. In other words, the cognitivism/non-cognitivism divide might not be one we could reform our way across, even if we wanted to.

Thirdly, if the goal is to achieve clarity and decision in our normative thought, then theories that posit a big gap, between the surface form of normative judgments and their underlying function, have some justificatory work to do. Consider expressivism. If our normative judgments, going forward, are just to be vehicles of expressing our endorsement of certain attitudes, and nothing more, why use linguistic forms that look like they ascribe objective properties? Wouldn't it bring greater clarity to become subjectivists and simply think and say: "I endorse such-and-such attitude"?

To be sure, expressivist have offered explanations of why our normative language, even though its real function is to express endorsement of attitudes, is propositional on the surface. But on the engineering approach, the challenge for expressivists will not be to provide an *explanation* of how this has come about, but to provide a *justification* for embracing it, going forward.

VI. What about motivational externalism?

I have proposed to evaluate metaethical views as competing proposals for which concepts to use in certain action-and-attitude-guiding roles. Can all the traditional theories sensibly be evaluated in this way? Clearly not a stand-alone Error Theory, which says nothing about what to do next, but as mentioned above, actual error theorists do say something about what to do next. But there is another kind of view that deserves separate discussion. Many forms of *naturalist realism* incorporate a doctrine known as *motivational externalism* – that there is no necessary connection between moral judgments and motivation. Which would seem to make it strange to try to evaluate such views as proposals for action-and-attitude-guiding judgments to make going forward.[19]

In order to discuss this, we need to distinguish two kinds of externalism.

---

[19] Many thanks to an anonymous referee for pressing this worry.

The first, which we can call "ordinary moral externalism", says that it is possible to make a sincere moral judgment without being motivated by it.

There is no tension between this view and the proposal I have made in this paper. I have appealed to "action-and-attitude-guiding" psychological roles, but this phrase is not meant to imply a direct link to motivation, of the kind ordinary moral externalism denies. Rather, the roles in question are to subject actions and attitudes to a certain kind of reflective supervision, by expressing standards against which they are assessed as appropriate or inappropriate. Let us call them "the R-roles". For purposes of illustration, consider a thinker who uses the ordinary normative concepts in the R-roles. For example, if she wants something, she may ask: "Would it be good to have this thing?". If her answer is Yes, then she sees her desire as appropriate; if No, she sees it as inappropriate. Or if she feels guilt about something she has done, or anger about something others have done, she may ask: "Was what I/they did wrong?". If her answer is Yes, she will see her guilt/anger as appropriate; if No, as inappropriate.[20] And likewise with all-things-considered ought-judgments and intentions.

The R-roles, then, do not have direct connections to motivation, but rather three-way connections to motivation and a certain kind of reflective self-criticism. If an agent makes an R-judgment but lacks the corresponding attitude/motivation, or has an attitude/motivation but denies the corresponding R-judgment, then her attitudes or motivations are, by her own lights, not appropriate. (This last sentence is meant as a stipulation: it defines what I mean by "the R-roles".)

However, a moral externalist may want to deny also this kind of three-way connection. On this view, an agent can make a sincere moral judgment without seeing it as having any implications about which motivations and attitudes it would be appropriate to have. For example, it is possible judge that something is morally wrong, without seeing that as a reason not to do it, or as making guilt/anger appropriate if it has been done. Let us call this view "strong moral externalism".

The debate having focused on ordinary moral externalism, it is not obvious exactly which authors would defend the stronger version. But it seems like an interesting view, also if we take the forward-looking perspective of the conceptual engineer. As mentioned in section I, one can mean different things by the

---

[20] This is a bit oversimplified, as there are further conditions (beyond wrongness) for guilt/anger to be appropriate. One can judge an act to be wrong but still not judge anger appropriate if the agent had a valid excuse, or if one lacks standing (e.g. because one has recently done the same thing oneself). I suppress these complications in the main text in order to keep it readable.

term "morality". Externalist views tend to take morality-the-social-practice as their object of study, and provide naturalistic accounts of the properties that practice refers to – by way of a suitable metasemantics for moral terms. A prominent example, which I will use for illustration, is Richard Boyd (1988), who takes moral terms to refer to the properties that stand in a causal relation he calls "epistemic regulation" to the community's past and present use of the terms.

Given a view of this kind, it makes good sense that an agent can make a moral judgment without seeing it as implying a reason for action. For example, she can believe that an act has the property that has epistemically regulated her community's use of the word "wrong", without seeing that as a reason not to perform it. And this can continue to make sense also if we take the forward-looking, engineering perspective. It seems a live option to decide to continue using the ordinary moral terms while thinking that their extensions are determined in the ways that Boyd or other externalists describe. And in that case, it makes sense to leave it open whether, in any particular case, our judgment that something is morally wrong (in this sense) will imply for us a reason not to do it, or to feel guilt or anger about it.

There ought to be room, then, for strongly externalist views about morality in the engineering project, but it is not clear how that would work, given the dialectical function of the R-roles in that project. But now we face a second complication. The engineering project is meant as an approach to metaethics as the meta-theory of *normativity*, while externalist views are defended as views about *morality*. So whether or not there is a tension between the two depends on what the externalist has to say about normativity. We will go through the possibilities by considering what my engineer should say to four different externalists:

Externalist 1 defends a strongly externalist view about morality, and has no particular view about normative notions like "reason for action" and "ought, all things considered". Her goal is just to account for morality and not other things.

She is not alone in this: there are august figures in the literature who have offered accounts of morality, full stop, and not of normativity. As mentioned earlier, I accept that such views have every right to the name "metaethics",[21] and I see no point in trying to argue over which philosophical questions are more interesting that which. So my response to Externalist 1 is simply to accept that we

---

[21] I have nothing very interesting to say about what it takes to "have a right to the name 'metaethics'". I take metaethics to be a sub-discipline of philosophy, bound together in the historical-sociological way academic sub-disciplines usually are, by a literature in which authors take themselves to be engaging with each other.

are engaged in different projects, and to wish her success with her project.

Externalist 2 defends a strongly externalist view about morality, which she embeds in some not-strongly-externalist view, perhaps subjectivist, or expressivist, etc., about normative notions like "reason for action" and "ought, all things considered".

This kind of overall view fits well in the engineering project. To adopt such a package of concepts seems like an interesting way to achieve clarity and decision in one's normative thought. Competition with other overall views is secured by the fact that, even though Externalist 2 does not propose concepts for the *moral* subset of the R-roles, she does propose concepts for *other* R-roles. Thus, we can welcome her aboard the engineering project.

Externalist 3 defends a strongly externalist view of morality, and furthermore denies the existence of any "all things considered" normative notions, of the kind just appealed to. Some of the "normative pluralists" may find such an overall view attractive. Pluralists posit a plurality of normative standards, each with its corresponding "ought" (the moral ought, the prudential, the legal, the chess ought, etc.), but deny the existence of any simple, unqualified ("all-things-considered") ought.

It is not always clear whether what is denied is the existence of *facts* about what we ought, simply, to do, or whether we possess such a *concept*. As a limited Error Theory, denying that there are ought simpliciter facts raises obvious engineering questions (what, if anything, to do with this concept if we learn that it does not refer?), and is thus not at odds with the engineering project. The challenge to the engineering proposal comes from the denial, made at least by Tiffany (2007: 233) and Baker (2018: 247-8), that we have any ought simpliciter concept. This denial could, again, be combined with strongly externalist accounts of all the qualified normative terms, resulting in an overall view that is strongly externalist across the board.

My response to Externalist 3 is just to dig in and say that we do have an ought simpliciter concept. I know because I use it every day. If an example is wanted, we can take one from Thomson:

> Suppose that Alfred is ill, and that only a dose of a certain medicine will cure him. It tastes truly awful, however. Alfred asks us "Ought I really take it?" It is a wildly implausible idea that we can reply only: "Well, your taking it would be very unpleasant, so in one sense of 'ought', it is not the case that you ought to take it, namely the 'ough-$_{enjoyable}$' sense of 'ought.' But your taking it would be good for you,

so in another sense of 'ought', you ought to take it, namely the 'ought$_{goodness\text{-}for\text{-}Alfred}$' sense of 'ought.'" It is likely that Alfred will repeat his question: "But ought I take it?" (Thomson 2001: 46)

"Ought simpliciter" and "ought, all things considered" are just terms of art for the "last ought before action": the "ought" that expresses the final conclusions of practical deliberation (in the first person), advice (in the second person) and criticism of action (in the third person). It is hard to see how this could be any sort of restricted "ought", given that there is no restriction in principle on the kinds of normative reasons one can feed into the final conclusions of practical reasoning.

I might of course be wrong. The upshot of our encounter with Externalist 3 is that unrestricted, all-things-considered normative concepts have an indispensable role in the engineering project as developed in this paper. Accordingly, if there are no such concepts, my version of the engineering project is a mistake.

Externalist 4 defends a strongly externalist view of both morality and of unrestricted normative notions like "reason for action" and "ought, all-things-considered".

Such a view would not fit in the engineering project. But Externalist 4 is also, so far as I know, a hypothetical figure. I am not aware that anyone has defended a strongly externalist account of the unrestricted ought. This is not surprising, since it is hard to see how such a view would work. Suppose for example that we tried to extend Boyd′s view, and said that "ought, all things considered" refers to the property that has causally regulated the community's use of that term. Now consider a thinker who knows both what this property is and that some act A open to her has it. If she accepts the hypothetical view considered here, to be consistent she has to grant that "All things considered, I ought to do A" is true. But as it happens, she is not convinced that A is the right decision. She is deliberating – or trying to deliberate – about whether to do A or something else, B.

Given all of this, then, she cannot use the term "ought, all things considered" to express the question she is trying to ask (whether to do A or B). The only solution for her that I can see is to introduce some new term, say "ought*" or a Gibbardian "the thing to do", and ask what she ought* to do, or what is the thing to do. But then, if she is inclined towards metaethics (which we imagine her being), all the metaethics-driving questions discussed in section II will, for her, just shift over to this new term. Are there facts about what she ought* to do, or what is the thing to do? If so, are they non-natural facts, or natural ones? Are they independent of her, or are they determined by her own attitudes, perhaps after some kind of idealization? Is it possible to know what one ought* to do, or what

is the thing to do, and if so how? Or can some kind of expressivist account of "ought*" or "thing to do" be made to work? And so on.

If our enquiry is motivated by these kinds of questions about normativity in the first place, then, it is hard to see how a strongly externalist view about normative notions could satisfy us. Boyd's view is of course just one example, and it is in any case difficult to predict the philosophical future – perhaps someone will propose a view next week that is both strongly externalist across the board and gives satisfactory answers to these questions. In that case also, the engineering proposal as developed in this paper would be shown to be misguided.

To sum up: there is no conflict between the engineering proposal and moral externalism as such. It depends on what kind of view, if any, about the wider issue of normativity that a given moral externalist view is combined with. Further, we have seen that the view developed in this paper needs two background assumptions, namely *a*) that we have unrestricted, all-things-considered normative concepts (such as OUGHT and REASON FOR ACTION), and *b*) that *strong* externalism about them is not plausible.

VII.   IS CONCEPT CHANGE POSSIBLE?

I have written throughout as if it is *possible* to change our normative concepts if we should decide to. That assumption might be naïve or at least problematic, either for general reasons in the philosophy of language, or for reasons specific to the normative case.

Starting with general reasons for scepticism, one might think that individuals do their thinking with words that belong to a public language, and that they are not able to impart idiosyncratic meanings to these words, or introduce new ones for private use. Alternatively, or in addition, one might accept a very strong form of *semantic externalism*, on which language users, individually or collectively, have very little control over the meanings of their terms. They control which terms they use, and which things they apply them to, but after that external factors take over and determine which properties their terms pick out.

I have already touched on a possible limit peculiar to the normative case: our normative judgments' status as cognitive or non-cognitive states might be hard-wired. Another possibility is that a certain, strong form of conceptual role semantics is true for normative concepts. On this view, their semantic properties would be *entirely* determined by the action-and-attitude-guiding roles they play,

making the apparent possibility of different concepts playing a given role illusory.[22] For example, while it *seems* like G. E. Moore and David Lewis describe different concepts, capable of playing the desire-regulating role, that role itself is in fact all that matters: any concept that plays that role will get its content from it and nothing else.

These are just some examples of possible limits to our ability to change our normative concepts. There is much to say about all of them. But fortunately, there is no need to say it here. I do not see any of these possibilities as potential *objections* to the engineering approach. Rather, I see them as research questions to pursue within metaethics as conceptual engineering. To repeat from the introduction, I do not intend "conceptual engineering" to refer to any particular kind of solution to the problems of metaethics, but to a way of thinking about the question we are asking in the first place. Namely: what kind of concepts to use in practical reasoning, going forward? Investigating the limits on which concepts we *can* use in practical reasoning is a natural part of that enquiry. If it should turn out that there are limits, like those mentioned above or other ones, the conceptual engineer should see that as an extremely interesting substantial result in her enquiry, not as an objection to it.

VIII. OTHER PEOPLE

Reverting now to the assumption that some kinds of conceptual revision are possible, a worry arises concerning other people, and how to talk with them. If we approach metaethics as conceptual engineering, and if people who reach different conclusions proceed to engineer their concepts accordingly, then it seems they will thenceforth be doomed to talk past each other when they try to discuss normative issues.

For example, suppose Mr. A is a perfectly coherent nonnaturalist, and Mrs. B a perfectly coherent constructivist. They try to discuss a looping trolley case:

A: In this case it is clearly right$_{Moore}$ to turn the switch.

B: No! Of course that would be terribly wrong$_{constructivist}$.

Since the proposition A asserts is different from the one B denies, have we lost

---

[22] Allan Gibbard (2003), Matti Eklund (2017) and Ralph Wedgwood (2007) propose versions of conceptual role semantics for normative terms. I am not sure about who among them would defend the very strong version of the view discussed above. Gibbard's remarks, quoted in section I, about the "wavering penchant for platonism" in ordinary thought, may suggest that, at least at the time of *Wise Choices, Apt Feelings*, he did not think his expressivist concepts were the only possible normative concepts.

the disagreement between them? How are they to discuss and learn from each other, if they systematically think in terms of different propositions?

What A and B have in common, however, is that the concepts they use play the same action-and-attitude-guiding roles for each of them. Thus, A endorses turning the switch, while B endorses not turning it, and furthermore endorses guilt/anger if she or others were to turn it (absent a valid excuse), which B rejects.

This generalises. Post engineering, interlocutors may indeed find themselves expressing different propositions. That kind of agreement and disagreement is lost. But their minds can still meet on questions about what to do, what to want, and how to feel. So long as they use concepts that play the same action-and-attitude-guiding roles for each of them, this function of normative discourse can proceed as before.

IX. Conclusion

Not a small part of our intellectual bandwidth is spent thinking about what to do, want, and feel. Of course, emotions, desires and decisions to act are often visceral and spontaneous, rather than results of reasoning. But in humans they have cerebral counterparts: desire can be rationally guided, to some degree at least, by judgments about good and bad; anger by judgments of right and wrong; and actions by judgments about reasons for action.

There seems to be more than one kind of judgment we might use in these roles. We could use concepts that purport to ascribe objective normative properties, properties that would be there even if we didn't think so or if our attitudes were different. Or we could ascribe normative properties in a subjective sense, one that is defined by our attitudes, or by what our attitudes would be, if we were fully coherent and knew all the facts. Or we could simply assert our desires and attitudes; make judgments fundamentally of the form: *I want this to happen, and I don't want that to. I love this, and hate that.* Or we could make judgments that express, rather than describe, our attitudes; judgments fundamentally of the form: *Boo to this, and Hurrah to that!*

It is both coherent, I have argued, and responsive to the concerns that lead many people to the subject in the first place, to approach metaethics as a *choice*, between these and other options, including those we might think of in the future.

References
Baker, Derek (2018); "Skepticism about Ought *Simpliciter*" in *Oxford Studies in Metaethics*, vol. 13, Russ Shafer-Landau (ed.), OUP

Berker, Selim (*manuscript*); *The Unity of Grounding*, downloaded (24. Oct. 2016) at http://scholar.har-vard.edu/files/sberker/files/berker-unity-grounding.pdf

Blackburn, Simon (1998); *Ruling Passions*, Oxford, OUP

Boyd, Richard N (1988). "How to be a moral realist." in *Essays on Moral Realism*, Geoffrey Sayre-McCord (ed.), Ithaca, Cornell University Press

Brandt, Richard B. (1979); *A Theory of the Right and the Good*, OUP

Brink, David (1989); *Moral Realism and the Foundations of Ethics*, Cambridge University Press

Burgess, Alexis and Plunkett, David (2013); "Conceptual Ethics II", *Philosophy Compass*

Burgess, John P. (2007); "Against Ethics" in *Ethical Theory and Moral Practice*

Darwall, Stephen; Gibbard, Allan; Railton, Peter (1992); "Toward fin de siecle ethics: some trends" in *The Philosophical Review*

Davis, Paul (2007); *Metaethics Explored*, Tirill: Humanities-Ebooks

Dreier, James (2006); "Negation for Expressivists: A Collection of Problems with a Suggestion for their Solution" in *Oxford Studies in Metaethics*, vol. 1, Russ Shafer-Landau (ed.), OUP

Egan, Andy (2007); "Quasi-realism and fundamental moral error" in *Australasian Journal of Philosophy*

Eklund, Matti (2012); "Alternative Normative Concepts" in *Analytic Philosophy*, Vol. 53 No. 2 June 2012 pp. 139–157

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

Eklund, Matti (2017); *Alternative Normative Concepts*, Oxford, OUP

Enoch, David (2011); *Taking Morality Seriously*, OUP

Fisher, Andrew (2011); *Metaethics: An Introduction*, Acumen

Frankena, William K. (1958): "Obligation and Motivation in Recent Moral Philosophy" in Melden (ed.) *Essays in Moral Philosophy*, Seattle, University of Washington Press

Gibbard, Allan (1990); *Wise Choices, Apt Feelings*, Harvard University press

Gibbard, Allan (2003); *Thinking How to Live*, Harvard University Press

Hare, R. M. (1952); *The Language of Morals*, Oxford, OUP

Harman, Gilbert (2009); "Guilt-free Morality" in *Oxford Studies in Metaethics*, vol. 4

Hussain, Nadeem and Shah, Nishi (2006); "Misunderstanding Metaethics: Korsgaard?s Rejection of Realism" in *Oxford Studies in Metaethics*, vol. 1

Jackson, Frank (1998); *From Metaphysics to Ethics, a Defence of Conceptual Analysis*, Oxford, Clarendon Press

Joyce, Richard (2001); *The Myth of Morality*, Cambridge University Press

Kahane, Guy (2012); "Must Metaethical Realism Make a Semantic Claim?" in *Journal of Moral Philosophy*

Kirchin, Simon (2012); *Metaethics*, Palgrave MacMillan

Laskowski, N. G. (2019); "The Sense of Incredibility in Ethics" in *Philosophical Studies*, vol. 176: 93-115

Lewis, David (1989); "Dispositional Theories of Value" in *Proceedings of the Aristotelian Society*

Mackie, John (1977); *Ethics; Inventing Right and Wrong*, Pelican Books

McPherson, Tristram (2012); "Ethical Non-Naturalism and the Metaphysics of Supervenience" in *Oxford Studies in Metaethics*, vol 7, OUP

McPherson, Tristram and Plunkett, David (2017); "Introduction: The Nature and Explanatory Ambitions of Metaethics" in McPherson and Plunkett (eds.) *The Routledge Handbook of Metaethics*, Routledge

Miller, Alexander (2003); *An Introduction to Contemporary Metaethics*, Cambridge, Polity Press

Miller, Alexander (2010); Metaethics: Oxford Bibliographies Online

Moore, G. E. (1903); *Principia Ethica*, Cambridge University Press

Nagel, Thomas (1997); *The Last Word*, Oxford, OUP

Olson, Jonas (2014); *Moral Error Theory: History, Critique, Defence*, Oxford, OUP

Parfit, Derek (2011); *On What Matters*, OUP

Quine, W. V. O. (1960); *Word and Object*, MIT Press

Railton, Peter (1989); "Naturalism and Prescriptivity" in *Social Philosophy & Policy*, vol. 7, pp 151-174.

Rawls, John (1971); *A Theory of Justice*, Harvard University Press

Scanlon, T. M. (2014); *Being Realistic About Reasons*, Oxford, OUP

Shafer-Landau, Russ (2003); *Moral Realism - A Defence*, OUP

Schroeder, Mark (2008); *Being For: Evaluating the Semantic Program of Expressivism*, OUP

Schroeter, Laura and Schroeter, Francois (2015); "Rationalizing Self-Interpretation" in *The Palgrave Handbook of Philosophical Methods*, Palgrave Macmillan UK

Stevenson, C. L. (1963); *Facts and Values*, Yale University Press

Strawson, P. F. (1963); "Carnap's views on constructed systems versus natural languages in analytic philosophy" in Schlipp, Paul Arthur (ed) *The Philosophy of Rudolf Carnap*, Open Court, La Salle, IL

Street, Sharon (2006); "A Darwinian Dilemma for Realist Theories of Value" in *Philosophical Studies*

Street, Sharon (2008); "Constructivism about reasons" in *Oxford Studies in Metaethics*, vol. 3, Russ Shafer-Landau (ed.), OUP

Street, Sharon (2009); "In Defence of Future Tuesday Indifference" in *Philosophical Issues*

Street, Sharon (2011); "Mind-Independence without the Mystery; Why Quasi-Realists Can?t Have it Both Ways" in *Oxford Studies in Metaethics*, vol. 6, Russ Shafer-Landau (ed.), OUP

Streumer, Bart (2017); *Unbelievable Errors*, Oxford: OUP

Suikkanen, Jussi (*manuscript*); "Parfit's Non-Realist Cognitivism and Truth", downloaded 14. Sept 2015 at https://www.academia.edu/12999021/Parfits_Non-Realist_Cognitivism_and_Truth

Tiffany, Evan (2007); "Deflationary Normative Pluralism" in *Canadian Journal of Philosophy* vol. 33

Timmons, Mark (1999); *Morality Without Foundations*, OUP

Unwin, Nicholas (1999); "Quasi-Realism, Negation and the Frege-Geach Problem" in *The Philosophical Quarterly*

Van Roojen, Mark (2015); *Metaethics: A Contemporary Introduction*, Routledge

Värynen, Pekka (*forthcoming*); "The Supervenience Challenge to Non-Naturalism" in *The Routledge Handbook of Metaethics*, Routledge

Williamson, Timothy (2007); *The Philosophy of Philosophy*, Oxford, Blackwell