Routledge
Taylor & Francis Group

Inquiry
An Interdisciplinary Journal of Philosophy

ISSN: 0020-174X (Print) 1502-3923 (Online) Journal homepage: https://www.tandfonline.com/loi/sinq20

Conceptual engineering via experimental philosophy

Jennifer Nado

To cite this article: Jennifer Nado (2019): Conceptual engineering via experimental philosophy, Inquiry, DOI: 10.1080/0020174X.2019.1667870

To link to this article: https://doi.org/10.1080/0020174X.2019.1667870

INQUIRY
https://doi.org/10.1080/0020174X.2019.1667870

Routledge
Taylor & Francis Group

Conceptual engineering via experimental philosophy

Jennifer Nado

Department of Philosophy, University of Hong Kong, Pokfulam, Hong Kong

ABSTRACT

Conceptual engineering provides a prima facie attractive alternative to traditional, conceptual analysis based approaches to philosophical method – particularly for those with doubts about the epistemic merits of intuition. As such, it seems to be a natural fit for those persuaded by the critiques of intuition offered by experimental philosophy. Recently, a number of authors [Schupbach, J. 2015. "Experimental Explication." *Philosophy and Phenomenological Research* 94 (3): 672–710; Shepherd, J., and J. Justus. 2015. "X-Phi and Carnapian Explication." *Erkenntnis* 80 (2): 381–402; Fisher, J. 2015. "Pragmatic Experimental Philosophy." *Philosophical Psychology* 28: 412–433; Machery, E. 2017. *Philosophy Within its Proper Bounds*. Oxford University Press] have suggested that experimental philosophy might be employed in service of conceptual engineering. In this paper, I provide a novel argument for x-phi's relevance to conceptual engineering, based on a 'functionalist' approach to conceptual engineering. In short, I argue that experimental philosophy is distinctively well-suited to investigation of the purposes or functions which our concepts serve, and the means by which they fulfil (or fail to fulfil) those functions. Experimental philosophy thereby uncovers potential engineering solutions that may serve as models for the conceptual engineer.

ARTICLE HISTORY Received 4 June 2019; Accepted 22 July 2019

KEYWORDS Conceptual engineering; experimental philosophy; metaphilosophy; intuition

There is a familiar characterization of analytic philosophy which holds its core project to be 'conceptual analysis'. Conceptual analysis, according to traditional pictures of the practice, is an attempt to provide a precise description of the conditions under which an entity falls under a given philosophically interesting, generally pre-theoretic or 'common-sense', concept. In other words, the aim is to do something like 'elucidate the meanings' of the everyman's concepts of knowledge, goodness, causation, beauty, and the like. On this 'traditional' take on philosophical method, the ultimate goal of philosophical inquiry is to reveal the natures of our concepts.

CONTACT Jennifer Nado jennifernado@gmail.com Department of Philosophy, University of Hong Kong, Pokfulam, Hong Kong.

But there is an alternative view that we might hold of the goal of philosophical inquiry. Rather than trying to study the concepts we *currently* possess, a philosopher might instead try to determine what concepts we *should* possess. She might attempt to make improvements – to improve clarity or reduce vagueness, to remedy various confusions and inconsistencies in our current concepts, or even to recommend wholesale replacement with a concept that is in some sense superior. Such revisionary projects are instances of what is now commonly referred to as 'conceptual engineering'. If we view conceptual analysis as a descriptive enterprise, then we may define conceptual engineering as a prescriptive one. Conceptual analysis describes the concepts we have; conceptual engineering makes recommendations.

Since conceptual engineering is inherently revisionary, it looks to be a *prima facie* attractive project for those with doubts about conceptual analysis – and those with doubts about intuition. Conceptual engineering permits the rejection of certain intuitions when those intuitions are held to reflect non-optimal aspects of our concepts. When an engineered concept conflicts with intuition, it is open to the engineer to deem the conflict a design feature, rather than a bug. For those of us with sympathies with the 'negative program' of experimental philosophy, conceptual engineering seems a natural candidate for replacing the ashes of the analyst's armchair.

Yet experimental philosophers have only just begun to examine how conceptual engineering might interface with the experimentalist programme. Both Schupbach (2015) and Shepherd and Justus (2015) have explored whether experimental philosophy might be recruited in service of Carnapian explication; Fisher (2015) has explored how x-phi might contribute to a potentially revisionary project he labels 'Pragmatic Conceptual Analysis'; and Machery (2017) explores conceptual engineering as one possible response to the troubling implications of negative x-phi. Here, I aim to throw my own hat into the ring, providing what I take to be a somewhat different account of x-phi's relevance to revisionary projects, based on a 'functionalist' approach to conceptual engineering.

1. Conceptual engineering and the limits of revision

My primary purpose in this paper will be to argue in favour of experimental philosophy's relevance to conceptual engineering projects. As such, I won't spend much time touting the virtues of conceptual engineering itself. The departures from intuition licensed by an engineering-based

approach to philosophical inquiry seem to me to be enough, on their own, to merit serious consideration of the method by those who are swayed by recent critiques of the use of intuition in philosophy. Rather than further pushing the case for conceptual engineering generally, I'll instead begin by arguing for a certain view on *how* to engineer concepts – one that will later serve as the basis for motivating x-phi's role in successful conceptual engineering.

A central question for all proponents of conceptual engineering concerns the limits of revision. Most would agree that not just any old change to a concept is permissible – some potential changes would seem to alter the pre-engineering concept so drastically that, post-revision, we are no longer even speaking of the 'same thing'. For instance, dropping both the truth and the justification conditions on knowledge to generate a new concept KNOWLEDGE* would seem to just utterly change the subject; such a proposal goes beyond any reasonable bounds for revision.

The natural end-point of this line of thought is the suspicion that conceptual engineering *inherently* changes the subject; that the entire enterprise reduces to a misguided exercise in fishing up red herrings. This sort of objection to conceptual engineering goes back at least as far as 1963, to a famous exchange between Rudolf Carnap and P.F. Strawson concerning Carnap's own version of conceptual engineering: explication. It will be helpful to consider that exchange, as well as Carnap's view itself, in some detail. The functionalist take on engineering that I'll be advocating can be viewed as one potential response to the Strawsonian critique, in that it provides a principled method for determining the boundaries of permissible revision in a way that is consistent with even very radical changes to a pre-engineering concept.

Explication can be characterized as the process by which an ordinary, everyday concept is transformed into – or replaced by – a more explicit, exact concept which is better suited for use in rigorous forms of inquiry. Carnap offers the example of the development of the quantitative, precise concept of 'temperature' from the pre-theoretic notion of warmth. The explicatum – 'temperature' – is better suited than the vague, subjective explicandum – 'warmth' – for the purposes of scientific inquiry.

It's worth noting that explication, as Carnap describes it, is primarily tailored towards improvements appropriate to the languages or conceptual schemes of the 'exact' sciences – physics, mathematics, logic, and the like. By contrast, 'conceptual engineering' (at least, as I'll use the term) covers

any form of conceptual improvement – even, potentially, improvements accompanied by a decrease in exactness. Explication is, then, only one possible approach to conceptual engineering. Though the idea of conceptual engineering is thus broader than that of explication, to date the work which focuses on experimental philosophy's possible role in conceptual engineering has largely used Carnapian explication as a model. So I'll continue to use Carnap as a jumping-off point – but I'd urge readers to keep in mind that we are not bound to the details of his approach.

The fullest presentation of the method of explication occurs at the beginning of Carnap's *Logical Foundations of Probability*. There, Carnap outlines four desiderata which a successful explicatum must fulfil. The explicatum must be *similar* to the explicandum; it must be *exact*; it must be *fruitful*, in the sense of enabling the formulation of laws or theorems; and it must be *simple*, or at least as simple as fulfilment of the first three desiderata permits. Though it is by no means the most important desideratum for Carnap, the similarity desideratum gives rise to Strawson's famous worry, and it will thus be our focus in what follows. Here is Carnap's statement of the desideratum:

> The explicatum is to be similar to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted. (Carnap 1950, 7)

Carnap is perhaps not as explicit as one might hope, in this passage, about what precisely this 'similarity' consists in. The most obvious interpretation, however, would seem to be similarity of meaning or conceptual content, or perhaps just similarity of extension. One interpretation, then, might be that it is required that the extension of the explicatum overlap sufficiently with the extension of the explicandum. Elsewhere, Carnap notes that that sufficient similarity is compatible with substantial narrowing of extension, as in the case when whales and other cetacea were removed from the scientific category that explicates 'fish'.

It is Carnap's flexibility regarding meaning or extension that seems, at least at first glance, to prompt Strawson's indignation. Explication, Strawson objects, changes the subject – and consequently, it cannot resolve the central questions with which philosophers are concerned. In Strawson's well-worn words:

> To offer formal explanations of key terms of scientific theories to one who seeks philosophical illumination of essential concepts of non-scientific discourse, is to do something utterly irrelevant – is a sheer misunderstanding, like offering a

> textbook on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Strawson 1963, 504)

The implication seems to be that the process of explication results in an unwanted change of subject; and thus, answers phrased in terms of the explicatum fail to answer our original questions regarding the explicandum.

As we've already noted, today's would-be conceptual engineer need not adhere to an orthodox Carnapianism. Yet there is a lingering sense that Strawson's worry poses a challenge to any enterprise that aims at conceptual revision. An engineer, the thought goes, must take pains to ensure that her invention bears a suitable degree of similarity of meaning to the concept which inspired it. Else, she merely changes the subject. This same worry crops up repeatedly in contemporary discussions of whether philosophers are at liberty to revise 'pre-theoretic' concepts. Note, for instance, Goldman, discussing the idea of revising our epistemic concepts:

> Whatever else epistemology might proceed to do, it should at least have its roots in the concepts and practices of the folk. If these roots are utterly rejected and abandoned, by what rights would the new discipline call itself "epistemology" at all?. (Goldman 1993, 272)

And another take on the worry, from Frank Jackson:

> [I]f we give up too many of the properties common sense associates with belief as represented by the folk theory of belief, we do indeed change the subject, and are no longer talking about belief. The role of the intuitions about possible cases so distinctive of conceptual analysis is precisely to make explicit our implicit folk theory and, in particular, to make explicit which properties are really central to some state's being correctly described as a belief. For surely it is possible to change the subject, and how else could one do it other than by abandoning what is most central to defining one's subject?. (Jackson 1998, 38)

On such views, a revisionary account which gives insufficient weight to the 'similarity' desideratum is thereby deemed a failure.

This might suggest a fairly direct role for 'positive' experimental philosophy – x-phi might contribute to clarification of our pre-theoretic concepts, enabling us to ensure that sufficient continuity of meaning is maintained during revision. Indeed, two extant proposals for x-phi's role in engineering essentially take this approach. Jonas Schupbach is quite explicit about this, writing that 'empirical research provides us with crucial information for assessing more directly just how well a particular explication does with regards to explication's similarity desideratum' (Schupbach 2015, 689). On Schupbach's view, then, the role of x-phi is

essentially to provide preparatory conceptual analysis – to clarify the meanings of the explicanda. It is especially worth noting that Schupbach envisions x-phi as being employed in service to what he calls 'Oppenheimian' explication – a variant of Carnap's original methodology which subordinates the fruitfulness desideratum to that of similarity.

Joshua Shepherd and James Justus, meanwhile, defend a role for experimental philosophy in 'explication preparation', which consists in clarifying the content of the explicandum 'to pinpoint the content that merits attempted preservation and the content that should be abandoned' (Shepherd and Justus 2015, 389). Again, the role this assigns to xphi seems to be that of uncovering information about the *meanings* or *contents* of the pre-engineered concepts. Shepherd and Justus give us a further hint to what they have in mind by noting that Carnap himself held something like explication preparation to be a crucial part of the explication process. Here is Carnap's own description in the Logical Foundations of Probability:

> There is a temptation to think that, since the explicandum cannot be given in exact terms anyway, it does not matter much how we formulate the problem. But this would be quite wrong. On the contrary, since even in the best case we cannot reach full exactness, we must, in order to prevent the discussion of the problem from becoming entirely futile, do all we can to make at least practically clear what is meant as the explicandum. What X means by a certain term in contexts of a certain kind is at least practically clear to Y if Y is able to predict correctly X's interpretation for most of the simple, ordinary cases of the use of the term in those contexts. (Carnap 1950, 4)

I take this passage to indicate that Carnap, to some extent, felt the pull of the worry that Strawson would later make explicit. Thus, though Carnap permits substantial difference in extension between explicandum and explicatum, he does seem to hold that at least some preliminary 'analysis' must be completed before we can consider changes to the extension of the explicandum – we must at least be clear on the usage of the term in 'ordinary' cases. This is remarkably in line with the sentiments noted earlier from Goldman and Jackson.

I have the opposite reaction to Strawson's worry – I feel no qualms whatsoever about 'changing the subject'. Indeed, I'm inclined to take a more radical stance than Carnap here; I think conceptual engineers should not take similarity of meaning, content, or extension to be a desideratum on successful engineering *at all*. As such, there is no need to employ x-phi to clarify the meanings of our pre-engineering concepts.

I'll defend this radical stance in more detail in the next section by appeal to a 'functionalist' take on conceptual engineering. But even at first glance, the meaning-similarity desideratum should cause us to furrow our brows a bit – meaning, after all, doesn't seem to be what most philosophers (outside the philosophy of language, that is) are ultimately interested in. Though the 'traditional' take on conceptual analysis mentioned in the introduction still has its advocates, the so-called 'linguistic turn' has lately fallen increasingly out of favour: many analytic philosophers now object that their real interest is in phenomena in the world, not in the language or concepts used to refer to said phenomena. On this post-linguistic-turn conception of philosophy, epistemologists are interested in knowledge, not 'knowledge'; metaphysicians are interested in causation, not 'causation'.

Of course, we might note that an object falls under 'knowledge' if and only if it is knowledge; but we should not conclude from this that the real philosophical goal is to elucidate the natures of *whatsoever* categories our terms refer to. What motivates philosophers doesn't seem to be a desire to delineate the categories that *happen* to be picked out by our terms, or by our mental representations. What philosophers typically want is to get at the nature of phenomena that we take to be philosophically significant, or interesting, or useful. We want to make the divisions in nature that are worth making. Perhaps the correct metasemantic theory will guarantee that our words do mark out such divisions; but perhaps not. By routing our desire to retain focus on important phenomena through a demand for similarity of meaning in our explications, we appear to be taking an unneeded gamble on how the facts about language ultimately work out. What if 'knowledge' turns out to refer to something utterly *boring*? Wouldn't we then *want* to change the subject?

Nonetheless, one might object, we can't wholly abandon similarity as a desideratum. If we do, the objection might run, the entire practice of philosophical theorizing is potentially trivialized – it will become possible to solve philosophical puzzles via stipulation. We could resolve, say, the millenia-old mystery of free will by simply stipulating that 'free will' is to refer to such-and-so compatibilist notion of action in accord with one's second-order desires (or what have you), and thus conclude that free will is quite obviously both possible and commonplace. Problem solved – we can now all pat ourselves on the back, and take a well-deserved vacation. We needn't worry about counterexamples, after all – if we encounter a case where an agent has a second-order desire that aligns with her act and

yet intuitively seems unfree, so much the worse for intuition. The case falls under our stipulated definition, and is therefore a free act. End of story.

Or why not go further – why not simply stipulate that 'free will' is to refer to H_2O? H_2O is a phenomenon of interest, after all. The claim above – that philosophers are interested in important phenomena in the world, not in meanings *qua* meanings – doesn't provide us with the resources to explain why the above replacement should be a problem. Clearly, this sort of *laissez-faire* approach will not do – we need some sort of 'test' to make sure that the proposed result of an instance of conceptual engineering is, in some way or another, sufficiently continuous with the original target. Now, it is of course true that by requiring that our engineered concept be sufficiently similar in *meaning* to the original, we rule out trivial stipulatory 'successes'. But preserving similarity of meaning is not the only method for preventing arbitrary self-congratulation. We could, alternately, impose a different form of similarity desideratum – one that insists only on similarity or continuity of *function*.

So that's the plan. I'll be arguing that continuity of function is sufficient to prevent philosophy from devolving into a free-for-all – and, moreover, I'll argue that it better captures the interests that lead us to conceptual engineering in the first place. A pleasant upshot of this is that, as we'll see, incorporating continuity of function as a desideratum generates a distinctive role for experimental philosophy – of both positive and negative types – in the engineering process.

2. Functional approaches to conceptual engineering

Preserving semantic similarity is, I'd argue, at best a very indirect route to ensuring that our conceptual interventions don't go off the rails, distracting us from our initial philosophical concerns. And it's not clearly a necessary one. Consider Strawson's lovesick inquirer into the workings of the heart. Resolving this poor fellow's puzzlement, ironically, in no way requires comparing the pre-theoretic meaning of 'heart' with some proposed explication. What the man *wants* is to learn how to woo his beloved; the 'heart' chapter of a physiology textbook won't help him do that, of course, but a psychology textbook might. We could, at least to some degree, further the lover's purposes and interests by providing him with information about all sorts of physical and mental processes described in the precise, technical, scientific language that Strawson decries. We might explain to him the workings of various 'bonding'

hormones such as oxytocin; we could share various research findings regarding the biological and cognitive components of physical attraction. And so forth. And we could do all of it without ever making use of the term 'heart', much less pondering its semantics.

One might think I'm pressing Strawson's off-hand, illustrative example a bit far here. But on the contrary, I think the example points to a generally missed feature of Strawson's objection to Carnap: Strawson very clearly motivates his accusations of irrelevance by appeal to the *functions* or *purposes* of everyday concepts, and the supposed inability of scientifically-refined replacements to properly serve those purposes. The text immediately following the well-known 'heart' quote, for instance, reads as follows:

> The scientific uses of language, whether formal or empirical, are extremely highly specialized uses. Language has many other employments. We use it in pleading in the law courts; in appraising people's characters and actions; in criticising works of art .... (Strawson 1963, 505)

And the crux of Strawson's argument displays this concern with purposes even more clearly:

> And it seems in general evident that the concepts used in non-scientific kinds of discourse could not *literally* be replaced by scientific concepts serving just the same purposes ... in most cases, either the operation would not be practically feasible or the result of attempting it would be something so radically different from the original that it could no longer be said to be fulfilling the same purpose, doing the same thing. (Strawson 1963, 505)

The real concern underlying Strawson's complaint, then, hinges on whether terms that have been developed for use in the sciences would be suitable for fulfilling the additional purposes that pre-theoretic terms serve in ordinary life. Thus, the worry is *not* merely that explication threatens to 'change the subject' – it is that explication threatens to generate concepts which fail to fulfil certain crucial functions, *thereby* leaving the philosophical issues prompted by those functions by the wayside. And that is a worry that seems to call for a similarity desideratum couched in terms of function or purpose.

Carnap, meanwhile – despite the earlier indications that the similarity desideratum is to be understood in terms of something like similarity of meaning – strongly suggests in his reply to Strawson that what he, too, *really* cares about is similarity of function or purpose.

> A natural language is like a crude, primitive pocketknife, very useful for a hundred different purposes. But for certain specific purposes, special tools are

> more efficient … If we find that the pocket knife is too crude for a given purpose and creates defective products, we shall try to discover the cause of the failure, and then either use the knife more skillfully, or replace it for this special purpose by a more suitable tool, or even invent a new one. [Strawson's] thesis is like saying that by using a special tool we evade the problem of the correct use of the cruder tool. (Carnap 1963, 938)

Indeed, the entire exchange is utterly *drenched* in talk of purposes, on both sides – the words 'purpose' and 'purposes' occur no less than forty-four times in the twenty-four pages comprising the two papers. Though Strawson admittedly frames the debate in terms of which methods suffice for *clarification* of philosophical problems, the argument that explication fails to so clarify is explicitly motivated by appeal to the purposes that give rise to said problems. Strawson's objection, I'd argue, is ultimately grounded in the following idea: a successful replacement concept must serve the all the various purposes served by the original concept. But if that's right, then 'changing the subject' is a problem *only* if the result is neglect of the original concept's purposes.

Again, we need not hold ourselves to Carnap, nor to explication per se. Stawson's objection holds some plausibility for the sorts of quantitative, precise, scientifically-calibrated modifications that Carnapian explication focuses on. The vagueness and subjectivity of 'warm', for instance, holds a certain amount of utility in ordinary contexts that would be lost if we demanded its wholesale replacement with precise temperature-speak. But an engineer is not restricted to bespoke scientific concepts; she might perfectly well undertake to modify ordinary concepts in order to increase their efficacy for the day-to-day tasks Strawson mentions. So Strawson's specific worry that scientific concepts will be unemployable in ordinary contexts seems to me to be inapplicable to conceptual engineering as a whole; pointing out the existence of unfulfilled functions in a proposed explicatum simply shows that the engineer has further work to do. Strawson gives us no reason to believe that we *couldn't* engineer some successor concept to an 'everyday' piece of language which better suited the purposes for which the original was used. And so long as said successor was better suited for said purposes, why would it matter one whit whether it possessed *any* similarity in meaning or extension to its predecessor?[1]

[1]It's also worth noting that nothing prevents an engineer from proposing two (or more) replacement concepts – one which improves upon certain 'scientific' functions of the original, and one which improves on whatever functions the concept plays in non-scientific contexts.

My suggestion, then, is that we take continuity of function/purpose to be a desideratum on successful conceptual engineering, rather than similarity of meaning. And indeed, this desideratum fits well with a number of current approaches to conceptual engineering which seem more or less 'functionalist' or 'pragmatic'. The most well-known of these is likely that of Sally Haslanger, who suggests that we ask ourselves:

> What is the point of having these concepts? What cognitive or practical task do they (or should they) enable us to accomplish? Are they effective tools to accomplish our (legitimate) purposes; if not, what concepts would serve these purposes better?. (Haslanger 2000, 33)

Other instances of broadly functionalist approaches to engineering are found in the work of Brigandt (2010), Fisher (2015), Thomasson (2017, forthcoming), and Prinzing (2018).

Let's pause for a moment to consider a bit more carefully what exactly is meant by the idea that concepts have functions. Though different functionalist accounts vary here, I want to make clear that when I speak of concepts having 'functions', I do not mean to imply that concepts have, say, a *telos* – that they possess some inherent, essential aim. Certainly that is not what Strawson or Carnap had in mind, nor is it what I intend. Relatedly, the claim is not that concepts are *individuated* by their functions – I thus patently do not wish to tie the current view to any form of teleosemantic view on concepts or language. Here I depart from another recent attempt to link conceptual engineering and experimental philosophy, that of Justin Fisher. Fisher argues in favour of what he calls 'pragmatic experimental philosophy' – a project in service of 'pragmatic conceptual analysis', which 'seeks an explication that will best preserve the patterns of beneficial usage for a given concept' (Fisher 2015, 414). Yet this version of explication is itself held to be motivated, at least in part, by the plausibility of a teleosemantic/pragmatic theory of reference, upon which 'a concept is correctly applicable to whatever would best sustain existing patterns in its beneficial usage' (Fisher 2015, 414). While the picture I offer here has clear affinities with Fisher's pragmatic conceptual analysis, I would divorce it entirely from issues of reference – I take conceptual engineering to enable us to operate independently of issues regarding natural language, and I take this to be a benefit of the approach.

On my view, by contrast, the idea that concepts possess 'functions' need imply nothing about their meanings – it implies no more than the banal fact that we use concepts to do things. As Strawson notes, concepts are used to appraise characters, to recount states of mind, to get people to

fetch things, and so on. It's not particularly difficult to generate at least a few plausible hypotheses for some functions for some of our core philosophical concepts, as well. The concept of free will, for instance, serves in part to mark out actions that are candidates for moral blame. The concept of consciousness plausibly serves in part to identify entities that deserve to be considered as possible moral patients, or that are potential initiators of intentional action, or the like. Craig (1990) has suggested that the concept of knowledge serves the function of allowing us to label reliable informants. These examples, by the way, should not be taken to suggest that concept 'functions' always need be in service of some normative end – many of our scientific concepts serve primarily to mark divisions that are useful for the purposes of prediction and explanation. Thus, the concept 'water' might simply have the function of picking out a certain natural kind.[2] And indeed, Carnap's 'fruitfulness' desideratum is reflective of the idea that a successful explication in science will fulfil a function – for fruitfulness, in Carnap's eyes, is a matter of enabling the concept to serve in the formulation of laws and generalizations.

The function of a concept is simply what it is used for – and concern with said sort of function seems utterly appropriate to a project of 'engineering'. Indeed, if we take the 'engineering' label dead seriously for a moment, it's hard to escape the idea that functional similarity is the most obvious candidate for maintaining continuity between 'pre-engineered' and the 'post-engineered' concepts. Consider the fact that within what we might call 'physical' engineering, there are numerous instances where the goal is to improve upon something that is already in use; to make a tool more precise, for instance, or to make a structural component more durable. We can view such cases as being (at least very broadly) analogous to the conceptual engineer's goal when she undertakes to propose a replacement for, say, the folk concept of free will. Now, it is obvious that *something* like a similarity desideratum holds for cases where a physical engineer aims for improvement – one does not improve the wheel by inventing the toaster, for instance. But the required similarity in such a case is very obviously similarity of function. One improves on, say, the Wright brothers' airplane wing by designing something that performs the function of an airplane wing (that is, generating lift for an aircraft),

[2]Here, I think, is another point where my account departs from Fisher's. Fisher would class the project of delineating natural kinds as one of 'naturalized philosophy', which he portrays as separate from pragmatic conceptual analysis. I would hold that an engineer concerned with well-functioning concepts can, and in fact in a great many cases will, concern herself with identifying natural kinds for the proposed concept to express.

in a way superior to the original. This sort of similarity is compatible with dramatic differences in things like appearance, material composition, and so forth. Consider, for instance, the significant difference in appearance between the original Wright plane and a modern jet – or, even more dramatically, between the first vacuum-tube driven computers and a modern laptop. So long as the proposed successor fills the function of the original in a superior way, anything goes.

We can go further still. The examples just mentioned are cases where an improvement in design results in the 'same thing' – that is, where the thing-to-be-improved and the post-improvement invention fall under the same category, or are called by the same name or term. For instance, an engineer begins with a plane wing and produces a (better) plane wing. But not all cases of improvement via engineering are like this. Consider the development of the telephone from the electric telegraph, or the development of the arch as an improvement over post and lintel architecture. Here the result of engineering is a wholesale replacement, rather than a mere improvement – but continuity of function remains. In other cases, improvement will lead to the replacement of a single design with multiple, more specialized designs, each taking over one of the functions of the original – as in Carnap's example of replacing a pocket knife with a variety of more specialized tools (perhaps a saw, an axe, and a razor, each specialized for different cutting tasks). Note that, in each of these cases, engineering a 'different thing' to fulfil a function of the original design prompts no worry analogous to 'changing the subject'. So why should conceptual engineering be any different?

A few more points to note. First, continuity of function need not entail *exact* continuity of function – it is open to us to decide that a certain function of the original design is no longer needed in its successor. An analogous instance in physical engineering would be the removal of running boards from most modern cars. Since modern cars are lower to the ground than earlier models, we no longer require a design element that fulfils the function of the running board (that is, facilitating entry into the car). Second, certain functions may be filled by sets of concepts, rather than isolated individual concepts – the required continuity of function might then be maintained at the level of the set, rather than the individual. An engineering project, then, might be one-to-many (replacing a single explicandum with multiple successor concepts), one-to-one, many-to-many, or even potentially many-to-one.

Finally, it is of course true that in many cases maintaining continuity in the function of a concept will generate continuity in meaning as a

byproduct – particularly if our current concepts are already fulfilling their functions effectively. It is unlikely, for instance, that we will be able to generate an improvement on the concept 'water' which somehow fails to substantially overlap in extension with the current concept. But this is no real objection to the current proposal. On the contrary, it can perhaps partially explain why philosophers have put so much weight on similarity of meaning – because typically, large departures in meaning indicate that the proposed successor concept does not fulfil the intended function of the original.

But only typically. It is certainly clear that in some cases, proposed successor concepts may well substantially diverge in meaning from the originals. Consider an analogy from science. Pre-Lavoisier, the dominant view of combustion claimed that it involved the release of a substance called 'phlogiston' – this release was held to account for the fact that flammable objects such as wood and other organic material become lighter after burning. There was a puzzle, however: some metals, after burning, gain mass. Thus, there was an apparently pressing question to be resolved: what is the mass of phlogiston, such that its release can lead to both a gain and loss in mass in different objects? To address this question, it was even proposed that one type of phlogiston might have negative mass. As we now know, of course, this was simply the wrong way of looking at the problem; current theories of combustion invoke oxygen rather than phlogiston. The oxygen concept, then, came to fill the function that the phlogiston concept was introduced for – it facilitated explanation the phenomenon of combustion, in a more effective way than phlogiston did.[3] Must we worry about the unanswered, original question regarding the mass of phlogiston? Must we ask whether 'oxygen' is similar enough in meaning to 'phlogiston' to answer this question, rather than 'changing the subject'? Of course not. And if we do, we'll be led to conclude that the subject was most certainly changed – the standard view within the philosophy of language is that phlogiston is an empty concept, referring to nothing.

Return, finally, to the worry that permitting changes of subject will enable stipulative solutions to genuine philosophical problems. Of course, engineering a new concept will likely involve an element of stipulation – but the stipulation will not be arbitrary. To be successful, a conceptual engineer must convince us that her proposed successor concept(s) fulfils the functions of the original concept in a more effective manner (or, alternately, demonstrate that certain functions of the original are

[3]Of course, the oxygen concept has functions beyond the explanation of combustion as well.

not needed in the successor). Since we still impose a form of similarity desideratum, conceptual engineering is not a free-for-all – good philosophy will still require hard work.

3. Conceptual Engineering via X-phi

Let's turn, at long last, to experimental philosophy and its potential role in this enterprise. Suppose I've convinced you that the only 'similarity desideratum' for conceptual engineering is continuity of function. Where, then, does that leave x-phi? One might have the following worry: neither 'positive' nor 'negative' experimental philosophy will really be of much use for a functionalist conceptual engineer.

Consider positive x-phi first. On the functionalist model of conceptual engineering, we don't *need* to uncover much of anything about the meaning of a pre-theoretic concept before engineering its replacement. We certainly don't need to know the sort of subtle meaning details that might be uncovered by surveying folk responses to baroque thought experiments. This might seem to indicate that positive x-phi will be fairly useless – no need to figure out exactly what folk concepts entail about Gettier cases, or trolley problems, or free will in a determinist universe. The sort of 'meaning-clarification' role suggested by Schupbach, and by Shepherd and Justus, is not obviously needed.

What about negative x-phi? Negative experimentalists have primarily concerned themselves with exposing the epistemic flaws of intuition. But on a conceptual engineering view, the use of intuitions is plausibly minimized. The standard picture of current philosophical methodology, with its focus on the method of cases and use of intuitive counterexamples, doesn't present a very promising way to go about engineering – to improve a concept, we must depart from intuition. So proponents of 'negative' x-phi are in one sense vindicated – they were right to critique slavish reliance on intuitions in philosophy. But in another sense negative x-phi might turn out to be a self-limiting project: once everyone is convinced that conceptual engineering is the right way to do philosophy, what's left for the experimentalist to do?

To attempt to feel out an answer to these questions, let's once again use physical engineering as an analogue. Specifically, let's consider what's called 'biomimetics': the practice of devising solutions to engineering problems by mimicking nature. The idea behind such a practice is that Mother Nature knows her stuff – evolution has already solved many of the problems engineers puzzle over, and thus we'd do well

to look to her solutions as a guide to our own. A biomimeticist might, for example, base an airplane design on the principles of bird flight – and, in fact, both Leonardo da Vinci and the Wright brothers did exactly this.

Consider a particular area where a biomimetic strategy is frequently applied – robotics. In many cases, robots are designed to perform tasks that humans already perform. There are robots that walk, robots that grasp items, robots that carry on conversations, and so forth. And in many cases, the designers of such robots have devoted considerable time and energy to examining how humans accomplish such tasks, in hopes of applying such knowledge in service of improving the analogous robotic skill. A thorough understanding of the mechanics of the human hand, for instance, might be employed to engineer a robot similarly capable of fine motor control; one might give the robot an opposable thumb to facilitate increased manual dexterity.

Of course, in some of these cases the goal is mimicry for the sake of mimicry – one might design a walking robot because one is interested in the study of human locomotion. But one might also just want a robot that is able to effectively locomote while, say, maintaining the free use of hands – and bipedal walking serves this goal. In the latter case, the robot mimics human walking because it is an effective means of fulfilling the functions the designer intends. The study of human locomotion thus serves to uncover potential solutions for engineering challenges – but, of course, the engineer is also permitted to make alterations and improvements to nature's designs where desired. Mother Nature is smart, but she's not perfect.

Couldn't experimental philosophy be employed in a similar way – as a sort of 'biomimetics' for conceptual engineering? I think it clearly could. On this approach, we would study our pre-theoretic concepts not because we need to ensure semantic similarity in our final products, but because we recognize that our current concepts are likely already fulfilling their intended functions to a reasonably good degree. Studying our current concepts provides, in essence, potential engineering solutions. It prevents us from having to design our concepts 'from scratch'. In fact, I think much of the work currently done under the banner of experimental philosophy is *already* more or less the sort of work that a 'biomimetic' approach to conceptual engineering ought to concern itself with. By contrast, I'd argue that traditional conceptual analysis via the method of cases does not necessarily produce the needed information.

Joshua Knobe has recently pointed out that x-phi, even in its 'positive' variety, is not merely empirically informed conceptual analysis.

> Just try picking out an experimental philosophy paper at random and taking a look at what it says. Almost certainly, you won't find that it makes any attempt at all to develop an analysis of a concept. Instead, you will find something quite different. Most typically, what you will find is an attempt to identify and explore a specific *effect*. (Knobe 2016, 42)

When an experimental philosopher studies, say, the concept of intentional action, she does not typically aim to churn out a set of necessary and sufficient conditions for the concept – instead, she aims for a characterization of the psychological mechanisms underlying the target class of judgments. She wants to know not only *whether* such-and-so factor is a necessary condition on intentional action, but *why* it is taken to be so.

The distinction is crucial. Mere discovery of the necessary-and-sufficient conditions for F-hood (should there be any) does not necessarily reveal *why* we have an F concept, or what the F concept is *used for*, or any other information about the potential functions of the concept or the means of their fulfilment. One might know (e.g.) that reliability is a necessary feature of knowledge, without knowing that we employ a reliability-centered concept due to our psychological need to identify effective informants. For traditional conceptual analysts, the project is complete once the analysis is counterexample-free. But for an experimental philosopher interested in assisting a project of conceptual engineering, the project is not yet complete. Knowing THAT x is knowledge iff x is a,b,c does not tell us WHY x is knowledge iff x is a,b,c.[4] Compare this to e.g. a biomimeticist trying to figure out *why* we have two bones in the lower leg (to improve foot rotation), instead of just stating *that* we do. The experimental philosopher's focus on underlying psychological mechanisms seems to be a promising route (though of course not the only possible route) for discovering the purposes our concepts serve, and the means by which these purposes are achieved – it thus stands to contribute much more to the engineering enterprise than a mere attempt to fit bi-conditionals to intuited data points.

The critiques of intuition pressed by negative x-phi also have a clear role to play in a 'biomimetic' approach to conceptual engineering. By uncovering weaknesses and deficiencies in our current concepts, the experimentalist learns how to improve the re-engineered versions of those concepts. To continue the robot analogy: study of the human foot might suggest that many of the smaller bones are vestigial holdovers from our tree-climbing ancestors (who needed greater flexibility for

[4]This is not to claim that conceptual analysts never *in practice* attempt to explain why certain necessary or sufficient conditions hold. But it is not, strictly speaking, a requirement of the conceptual analysis enterprise as traditionally understood (that is, understood as isolating a counterexample-free bi-conditional).

grasping), and that their persistence does little in modern humans but increase the incidence of sprained ankles and tendonitis. Armed with this information, the roboticist might then with confidence plan a foot-like structure with fewer moving parts. Had she merely detailed knowledge of the *form* of the foot but not the *function*, she might have designed her robot's foot as a mere copy of our own – bringing the flaws of human foot architecture along for the ride. Note, though, that the proper reaction to a discovered flaw in human anatomy is to *correct* it in the robotic design – not to conclude that the study of human anatomy has nothing to contribute to the field of robotics. Similarly, the biases uncovered in intuition do not justify a complete rejection of intuitions' philosophical value.

Let's look at a few examples of how these potential applications of x-phi might work out in practice. One of the most well-studied phenomena in positive experimental philosophy concerns asymmetries in judgment on cases with different moral valence. The original finding, first reported in Knobe (2003), revealed that subjects are more inclined to judge side-effects of an action to be intentionally brought about when the side-effect is considered morally wrong. Subsequently, parallel effects were found for a slew of other mental state attributions, including desiring (Tannenbaum, Ditto, and Pizarro 2007), deciding (Pettit and Knobe 2009), knowing (Beebe and Buckwalter 2010), and believing (Beebe 2013); moral asymmetries even seem to influence our attributions of causation (Hitchcock and Knobe 2009).

The subsequent literature on this pattern of phenomena quite frequently aims to explain the effects by appeal to the purposes or functions the affected concepts might serve. Knobe's original, competence-based account of his findings, for instance, bears stunning similarities to the 'tool' analogy Carnap deploys in his reply to Strawson:

> it appears that people's concept of intentional action should be understood as something like a multi-purpose tool. If we want to understand why the concept works the way it does, it is not enough to examine its use in the tasks of prediction, control and explanation. Many important facts about the concept can only be correctly understood when we see that it also plays an important role in the process by which people determine how much praise or blame an agent deserves for his or her behaviour. (Knobe 2006, 227)

By attempting to tease out the psychological processes that lead to the observed pattern of judgments, Knobe aims to provide insight into the

functions served by a concept and the means by which those functions are fulfilled.

Applying 'negative' experimental results in service of conceptual engineering is equally natural on the functionalist approach. As a fun recent example, De Brigard (2010) provides experimental evidence that intuitions regarding the desirability of life in an 'experience machine' seem to reflect more of a bias towards maintaining the status quo than a genuine preference for reality over simulation. Such results might lead a conceptual engineer to, for instance, explore the possibility of designing her concept of 'well-being' in a way that willingly embraces simulated hedonism. The squeamishness most people have towards 'plugging in', she might argue, is no more than a bias, and we'd be better served by concepts that reject said bias.

More generally, insofar as experimental results indicate that a given intuitive response is largely a reflection of the operation of irrelevant factors such as culture, gender, framing, and so forth, an engineer can take this as plausible evidence that retaining the intuition isn't necessary to ensuring the relevant concept's effective fulfilment of its function. You don't intuit that x is F because x's being F is essential to a well-functioning F concept, an engineer might argue – you intuit that x is F because of the contingencies of your background or situation. Therefore, I have no reason to think that a revised F concept that excludes x will thereby neglect some important purpose. The approach is, in fact, essentially the same as suggested above: if an engineer knows *why* people intuit that x is F (rather than merely *that* they do), this can allow her to determine whether the inclusion of x in a revised F concept will improve, degrade, or be irrelevant to the efficacy of the post-revision concept.

To sum up – we have a set of 'natural' pre-engineering concepts that have developed over millennia of physical and cultural evolution. It stands to reason that they function pretty well. But it also stands to reason that they are not perfect. This suggests that we should study these pre-engineering concepts in order to 'get a leg up' on designing concepts which successfully serve the functions we want. We do so by identifying (a) the purposes our current concepts serve, (b) the elements of those concepts that help them fulfil those purposes, and (c) the elements of those concepts that are more like the philosophical analogue of an appendix, a set of wisdom teeth, or a pair of male nipples. Current survey techniques employed by experimental philosophers provide one way to pursue these investigations, but we might imagine others –

study of linguistic corpus data, neuroscientific investigation, and anthropological work, to name just a few.

As a final note, however: study of our current concepts, via x-phi or other means, is not *strictly speaking* necessary for a project of conceptual engineering. It would be *possible* – though admittedly much more difficult – to engineer a walking robot without studying the workings of human anatomy. Similarly, it is possible that a particularly brilliant conceptual engineer might invent an effective concept without bothering to carefully study our current concepts. That said, studying our current concepts clearly has the potential to make the job of the conceptual engineer much, much easier – just as study of human anatomy makes the roboticists' job easier. But the role such study plays is not to place restrictions on post-engineering meaning or extension. It is to provide possible ways to fulfil the functions we want our concepts to perform.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Beebe, J. 2013. "A Knobe Effect for Belief Ascriptions." *Review of Philosophy and Psychology* 4 (2): 235–258.

Beebe, J., and W. Buckwalter. 2010. "The Epistemic Side-Effect Effect." *Mind and Language* 25 (4): 474–498.

Brigandt, I. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177 (1): 19–40.

Carnap, R. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Carnap, R. 1963. "Replies and Systematic Expositions." In *The Philosophy of Rudolf Carnap*, edited by P. A. Schilpp, 859–1013. La Salle: Open Court.

Craig, E. 1990. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Clarendon Press.

De Brigard, F. 2010. "If You Like It, Does It Matter If It's Real?" *Philosophical Psychology* 23: 43–57.

Fisher, J. 2015. "Pragmatic Experimental Philosophy." *Philosophical Psychology* 28: 412–433.

Goldman, A. I. 1993. "Epistemic Folkways and Scientific Epistemology." *Philosophical Issues* 3: 271–285.

Haslanger, S. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34 (1): 31–55.

Hitchcock, C., and J. Knobe. 2009. "Cause and Norm." *The Journal of Philosophy* 106 (11): 587–612.

Jackson, F. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.

Knobe, J. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63: 190–194.

Knobe, J. 2016. "Experimental Philosophy is Cognitive Science." In *A Companion to Experimental Philosophy*, edited by J. Sytsma, and W. Buckwalter, 37–52. Hoboken: Wiley.

Machery, E. 2017. *Philosophy Within its Proper Bounds*. Oxford: Oxford University Press.

Pettit, D., and J. Knobe. 2009. "The Pervasive Impact of Moral Judgment." *Mind and Language* 24 (5): 586–604.

Prinzing, M. 2018. "The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Objection." *Inquiry* 61 (8): 854–880.

Schupbach, J. 2015. "Experimental Explication." *Philosophy and Phenomenological Research* 94 (3): 672–710.

Shepherd, J., and J. Justus. 2015. "X-Phi and Carnapian Explication." *Erkenntnis* 80 (2): 381–402.

Strawson, P. 1963. "Carnap's Views on Conceptual Systems Versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, edited by P. Schlipp, 503–518. La Salle: Open Court.

Tannenbaum, D., P. H. Ditto, and D. A. Pizarro. 2007. Different Moral Values Produce Different Judgments of Intentional Action." University of California-Irvine, Unpublished manuscript.

Thomasson, A. L. 2017. "What Can We Do, When We Do Metaphysics?" In *Cambridge Companion to Philosophical Methodology*, edited by G. D'Oro and S. Overgaard, 101–121. Cambridge: Cambridge University Press.

Thomasson, A. L. Forthcoming. "A Pragmatic Method for Conceptual Ethics." In *Conceptual Ethics and Conceptual Engineering*, edited by H. Cappelen, D. Plunkett, and A. Burgess. Oxford University Press.