Erkenntnis (2023) 88:1509–1531
https://doi.org/10.1007/s10670-021-00413-w

ORIGINAL RESEARCH

Experimental Explications for Conceptual Engineering

Samantha Wakil[1]

Received: 7 March 2020 / Accepted: 23 March 2021 / Published online: 28 June 2021

Abstract
This paper argues for two conclusions: (1) evaluating the success of engineered concepts necessarily involves empirical work; and (2) the Carnapian Explication criterion *precision* ought to be a methodological standard in conceptual engineering. These two conclusions provide a new analysis of the race and gender debate between Sally Haslanger and Jennifer Saul. Specifically, the argument identifies the resources Haslanger needs to respond to Saul's main objections. Lastly, I contrast the methodology advocated here with the so-called "method of cases" and draw out some general implications for how we should think about concepts.

1 Introduction

The conservative idea that it is a philosopher's job to clarify common sense beliefs about ordinary concepts is being weeded out from the population. In its place, a revisionist agenda is being selected for: philosophers should not merely describe but analyze and suggest ways to improve our concepts. This new research program is called "conceptual engineering" (CE) (see Plunkett and Cappelen 2020). Conceptual engineering is predicated on two assumptions. First, concepts are the tools we use to understand ourselves and the world around us: "[O]ur conceptual repertoire determines not only what we can think and say but also, as a result, what we can do and who we can be" (Burgess and Plunkett 2013, p. 1091). Second, our concepts are not immutable. Many of the concepts we currently possess have changed over time and it's possible for us to investigate whether our current concepts are defective in some way. Evaluating these concepts, and suggesting ways to improve any defects, is the ultimate goal.[1]

[1] Given the controversy surrounding what concepts are, Cappelen (2018) suggests conceptual engineering refer to representational devices broadly construed. There's no problem with this suggestion, but following the trend in the literature, I will continue to use the word concept, keeping in mind that it does not presuppose any views about what concepts are.

✉ Samantha Wakil
samantha.wakil@protonmail.com

[1] The University of Nevada, Las Vegas, Las Vegas, NV, USA

Of course, if we expect our concepts to do things (and do those things well), then conceptual engineering needs an explicit and effective methodology. This methodological question has, unsurprisingly, become a focal point in recent literature:

> Obviously some big questions in the vicinity remain entirely unresolved, such as what the proper methodology for conceptual engineering is. (Eklund 2015, 382)

> So far I've been talking as though we are looking for a descriptive account of when conceptual revision has gone well … We should also be interested in how it *should* be done—i.e., what *should* be considered successful. (Cappelen 2018, 119)

> If metaphysics centrally involves normative conceptual work, how ought we to be doing it? What methods and standards should we employ? While the idea that work in metaphysics may involve normative conceptual work has begun to gain some traction, far less has been said about how that work is to be done … Answering the methodological question of how it should be done is important. (Thomasson 2020, 436)

Most conceptual engineers take Carnap's method of *Explication* as the starting point for methodological inquiry.[2] As Carnap (1950, p. 2) describes it,

> The task of explication consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the explicandum, and the exact concept proposed to take the place of the first (or the term proposed for it) the explicatum.

Carnap's most detailed description of Explication is located in the first chapter of the *Logical Foundations of Probability*. Carnap specifies four adequacy conditions that characterize the procedural component of explication *and* provide an evaluative schema for gauging success (1950, p. 7; my emphasis):

[2] There are other varieties of explicative methodology. Carnap's (1950) description was the most developed and thorough at the time (see Carus 2007). Since then, numerous philosophers have contributed to the method of explication through negative critique (Quine 1951; Strawson 1963; Reck 2012) and friendly-amendments and defence (Brun 2016; Hanna 1968; Justus 2012; Koch 2019; Maher 2007; Quine 1960). A distinguishing feature of Carnapian Explication is that there is no true or correct explicatum: "Carnap's distinction between 'external' and 'internal' questions [has] an obvious application to the process of explication in general. The explicatum … belongs to some formalized discourse—some 'framework'. The explicandum … belongs ipso facto to a mode of discourse outside that framework" (Stein 1992, p. 280). Therefore, for any given explicandum there is no correct explicatum any more than there is a "correct" choice of language. Crucially, though, according to Carnap, we can rationally compare different frameworks. This contrasts sharply with Quine, for whom "there is no stepping outside the 'conceptual scheme' in which our 'mother tongue' places us. We complicate and sophisticate that scheme by means of our science, but we can never be in a position to choose our language" (Carus 2007, p. 24). Although other versions of explication may have interesting connections to conceptual engineering, the centrality of Carnap's original conception—and the pluralism at the heart of that conception—aligns much better with the goals of conceptual engineering.

1. 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.
2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an **exact** form, so as to introduce the explicatum into a well-connected system of scientific concepts.[3]
3. The explicatum is to be a **fruitful** concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
4. The explicatum should be as **simple** as possible; this means as simple as the more important requirements (1), (2), and (3) permit.

Curiously, despite Carnap's claim that "[t]he task of explication is of very general importance for the construction of concepts" (Carnap 1950, p. 2), conceptual engineers seem to think Explication does not make good on its promise as a generalizable methodology. Cappelen (2018, p. 11), for instance, claims "Carnap's notion of explication, however, is narrower than the activity I'm interested in." Similarly, Pinder (2020, p. 4) states that "Carnap's proposal is not, as it stands, suitable for this particular task." Assuming these *wholesale* assertions are correct (Sect. 3 argues they are not), it's worth considering whether some of the *specifics* from Carnap's proposal can be methodological pillars for CE. I argue that experimental philosophy's utility in evaluating concepts establishes a demand for the precision criterion in all conceptual engineering projects.

Analyses of Carnapian Explication and experimental philosophy are common topics in the CE literature. But, despite their prevalence, important methodological considerations appear to have been overlooked. The reason is two-fold. First, analyses of Explication have focused almost exclusively on the similarity and fruitfulness criteria, while precision has fallen to the wayside (Lindauer 2019; Koch 2019; Pinder 2017, 2020; Thomasson 2020). Second, arguments about the use of experimental philosophy typically address the 'change of subject' objection (Nado 2019a; Prinzing 2018) or focus on pre-engineering theorizing—similar to what Shepherd and Justus (2015) call "explication preparation" (see also Brun 2016; Nado 2019b; Schupbach 2017).

Section 2 argues that experimental philosophy is not merely useful for the pretheoretical component of conceptual engineering but is actually necessary for ultimately evaluating CE success. Following a useful distinction made by Koch (2019), this argument applies to theoretical and practical success. Section 3 builds off the first argument to establish a second claim: that the third Explication criterion *precision* should be a methodological standard in conceptual engineering. With the exception of Novaes (2018), precision has been problematically ignored

[3] Following Carnap, 'exactness' and 'precision' are used interchangeably.

by conceptual engineers. Section 4 presents the debate between Haslanger (2000, 2006) and Saul (2006) as a paradigmatic case of the pitfalls that accompany neglecting precision. The conclusions established in Sects. 2 and 3 also reveal a new line of response to a set of objections against Haslanger's proposed definitions of gender and race. Finally, Sect. 5 concludes by considering how this analysis relates to the method of cases and responds to anticipated objections.

2 The Argument: Part 1

This section argues for the necessity of empirical methods in conceptual engineering. The term necessary should be understood in an instrumental sense, not a metaphysical or an epistemological one. This is reminiscent of Quine's famous engineering analogy about naturalizing epistemology: "For me, normative epistemology is a branch of engineering … The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed" (Quine 1986, pp. 664–665). Similarly, the necessity claim in my argument is merely derivative of CE's explicitly stated goals. With this qualification out of the way, here is the argument:

(P1) The goal of conceptual engineering is to evaluate and improve concepts.
(P2) If conceptual improvement is the goal, then the metric for success will be the degree to which an engineered concept makes progress towards achieving its function compared to the original.[4]
(P3) Evaluating that success requires information about the functionality of the original concept *and* the engineered concept.
(P4) Acquiring that information requires empirical investigation.

(C1) Conceptual engineering necessarily involves empirical work.

Let's proceed in order. Premise one is uncontroversial. A quick glance at the CE literature immediately reveals the goals of the discipline. Here are just three examples:

> 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. (Nado 2019a, 2)

[4] Function talk is also rampant in the CE literature. As Nado (2019a) rightly notes, however, the notion of function can be suitably deflated. It need not invoke any commitments to an essentialist notion of telos, or to a telosemantic theory of reference: "it implies no more than the banal fact that we use concepts to do things" (2019a, p. 11). Accordingly, I will continue to talk about concepts having functions.

> Holding in place some function or functions the concept is to serve, we may aim to redesign old concepts to serve that function better, or to engineer new concepts that can serve a function that was done imperfectly or not at all by our prior conceptual scheme. (Thomasson 2020, 440)

> The terms or concepts which we use to talk and think about a particular subject matter can be defective and can be improved to address these defects … If our representational devices can be defective [ ] then we should be engaged in two kinds of activities: (i) investigating whether these concepts are defective and (ii) if defects are found, then ameliorating the defective concepts. (Cappelen 2018, 39–40)

Premise two extends from a straightforward understanding of improvement, which I take to mean something like "making better". Of course, there are numerous ways an engineered concept could be considered "better." Perhaps a concept C_x is better than C_y because it's unifying, it's more explanatory, it generates testable predictions, it's simple, etc. Because of this plurality, and because CE is predicated on our concepts *doing* something (i.e., performing some kind of function), I suggest that improvement, more specifically, be conceptualized in terms of progress as defined by Alexander Bird:

> **Definition of progress**: "If the aim of X is Y, then X makes progress when X achieves Y or promotes the achievement of Y." (Bird 2007, 83)

Thus, when Y = [a concept's function] and X = [an engineered concept], X is an improvement over the original concept in so far as it makes progress by achieving or promoting the achievement of Y. Notice this definition of progress is neutral regarding what the function of a concept is. Whether the respective concept's function is the advancement of science (as for Carnap), social and political amelioration (as for Haslanger 2006, 2012), epistemic theorizing (as for Eklund 2002; Scharp 2013), or explaining moral phenomena (as for Railton 2003), the notion of improvement and therefore the general metric for success is the same: whether and to what degree the engineered concept makes progress (as defined above) over the original.

Premise three is based on the comparative structure of the success standard from premise two. Interestingly, contrastiveness also features prominently in Sober's (1999) account of testability. As Sober (1999, pp. 57–58) candidly remarks, when it comes to scientific hypotheses,

> Don't ask whether a hypothesis can or cannot explain an outcome. And don't ask whether the hypothesis says that the outcome was probable or improbable. The relevant question is whether the outcome is more probable according to one hypothesis than it is *according to another*… We can judge which hypotheses do better and which do worse in their competition, but that is all. (My emphasis)

Similarly, we can only judge which concepts do better and which do worse when they're competing. Therefore, in order to assess whether an engineered concept is progressive at least two sources of information are obviously required. First, we

need information about the functionality of the original concept. Second, we need information about the functionality of the engineered concept to make the relevant comparison.[5]

Premise four is the most controversial, especially for those not naturalistically inclined. There are several reasons to accept premise four, starting with another analogy between CE and the Philosophy of Science. Besides the claim that hypothesis testing is contrastive, Sober also asserts that

> When biologists test evolutionary hypotheses…The goal is not to determine whether an evolutionary explanation can be invented that is consistent with the theory's delineation of the possible causes of evolution, but to determine whether there is evidence that discriminates between different evolutionary hypotheses concerning what actually occurred. (Sober 1990, 60)

Perhaps surprisingly, evaluating engineered concepts appears strikingly similar to testing evolutionary hypotheses. The goal is not to devise new terms from a "blank slate" that are merely consistent with a concept's possible functions (Thomasson 2020, p. 447). Rather, the aim is to acquire evidence that discriminates between different explicata regarding which *actually* achieves or promotes the achievement of the concept's function. The best source of that evidence is found through experimental philosophy.

Experimental philosophy uses some of the empirical methods of social psychology and cognitive science to acquire data relevant to addressing traditional philosophical questions and debates (see Plakias 2015 for a useful overview). It is typically divided into two programs. The negative program appeals to empirical results to cast deep skepticism about intuitions being the primary source of philosophical evidence. Although the debate about their epistemic significance has a long and contentious history, a new cluster of worries about the reliance on intuition arguably began with the experimental studies by Weinberg et al. (2001), Knobe (2003), and Alexander and Weinberg (2007). The so-called "Knobe effect", for example, revealed an asymmetry in participants' attribution of intentionality that had implications for the philosophy of action. (More on this below.) Additionally, there are disparate intuitions among philosophers regarding scenarios involving concepts of interest like free will, knowledge ascriptions, moral blame, twin earth cases, etc.

[5] This contrastive element also arguably undergirds Carnap's caution against conflating explicanda with explicata. As his diagnosis of the problem of probability indicates, "When we look at the formulations which the authors themselves offer in order to make clear which meanings of 'probability' they intend to take as their explicanda, we find phrases as different as 'degree of belief,' 'degree of reasonable expectation,' 'degree of possibility,' 'degree of proximity to certainty,' 'degree of partial truth,' 'relative frequency,' and many others. This multiplicity of phrases shows that any assumption of a unique explicandum common to all authors is untenable … But most investigators in the field of probability apparently believe that all the various theories of probability are intended to solve the same problem and hence that any two theories which differ fundamentally from one another are incompatible…. These mutual rejections are often formulated in rather strong terms. This whole controversy seems to me futile and unnecessary. The two sides start from different explicanda" (Carnap 1945, pp. 517–518). Thanks to Jack Justus for highlighting this additional connection.

And it appears this intuitional diversity is found among non-philosophers (Weinberg et al. 2001).

Philosophers in the negative program also argue that intuitions are sensitive to philosophically irrelevant factors like a person's gender and cultural and socioeconomic background, the way an outcome is described, or the order in which the cases are presented. For example, Swain et al. (2008) found that attributions of knowledge were subject to order effects. When participants were presented a clear case of knowledge before a true temp case, they were less likely to attribute knowledge to true temp than participants who received the true temp case first.[6] Yet, when participants received a clear non-knowledge case first they were more likely to attribute knowledge to the true temp case than participants who received true temp first. In sum, experimental philosophy's negative program places an aggressive assault against intuition-based theorizing. The experimental results from the studies just cited (and many others) demonstrate the unreliable, relative, and highly sensitive nature of intuitions, making them a poor evidential source for philosophical theorizing, and thus warranting a significant change to our philosophical practices.[7]

Against the naysayers, the positive program uses experimental data to gain information about folk concepts and intuitions in order to *support* traditional answers to philosophical questions. Philosophers in the positive program are "motivated to explore intuitions experimentally because they think that by doing so they can do a better job of conceptual analysis" (Stich and Tobia 2016). Neta (2012, p. 333), for instance, argues there is an intuitively compelling rationale that predicts the order effect found in Swain et al. (2008):

> It is intuitive that, for a psychologically salient property X (e.g., how substantial a defeater is), when a borderline case of X is compared against clear cases of X, it seems less X-ish, and when a borderline case of X is compared against clear cases of non-X, it seems more X-ish.

Of course, one example does not vindicate the reliability of intuitions. The claim that philosophers can, in general, reliably predict what ordinary subjects will and will not find intuitive is itself an empirical question, and one the positive program can address. Furthermore, some recent studies have not replicated the original findings of the negative program (Seyedsayamdost 2015; Adleberg et al. 2014) and others have found that there is some cross-cultural uniformity in knowledge attributions in Gettier cases (Machery et al. 2015).

The positive program also emphasizes how experimental philosophy can uncover previously unconsidered features of our concepts. As Plakias (2015, p. 8; my emphasis) summarizes,

> Strohminger and Nichols (2014) presented subjects with the hypothetical case of someone named Jim, who had been in a car crash and suffered psychologi-

[6] True temp cases involve a person who unknowingly gets a device surgically implanted in their brain that reliably measures the temperature of wherever the individual is.

[7] See also Colaço et al. (2014), De Cruz (2017), Machery et al. (2004), May et al. (2010), Nichols and Knobe (2007), Schwitzgebel and Cushman (2015), Cushman et al. (2008) and Tobia et al. (2013).

> cal damage. They then described various changes to the person's personality or memory and asked subjects whether he remained the same person or whether his identity had changed. What they found was that a change in Jim's moral attitude was the most influential factor in determining that his identity had changed—more influential than memories, desires, or other personality traits. *This finding raises the novel possibility that the ordinary concept of the self includes a moral dimension*, such that one's moral views are an essential part of one's personal identity.

Despite their differences, taken together, the negative and positive programs demonstrate how "[f]or every philosophical use of intuition, there is a role for experimental philosophy" (Plakias 2015, p. 7). Experimental methods don't merely provide information about our concepts; crucially, they "reveal the contours of concepts" (ibid., 6). Thus, for the purposes of CE the distinction between the negative and positive program is unimportant. Whether the results of a study are taken as confirming or disconfirming the reliability of a specific intuition, the pressing point remains: experimental philosophy provides a testable, scientifically-grounded, and data-driven means for acquiring the evidence needed for evaluating the functionality of a concept.

Two objections need to be addressed before moving on. First, it might be tempting to reject the conclusion on the grounds that CE is best construed as a form of ideal-theorizing. Meta-semantic considerations, and semantic externalism in particular, have raised concerns about actually implementing suggested conceptual revisions. If one thinks, as the semantic externalist does, that people have very little control over the meanings of our terms, then the feasibility of changing a concept according to an engineer's recommendation will seem highly improbable, if not impossible (Plunkett and Cappelen 2020). Concerns about feasibility have motivated some suggestions that CE not be concerned with implementation at all. Perhaps, as Burgess and Plunkett (2013, p. 1096) claim, "Whether we ought to use a given concept could be completely independent of how hard or easy it would be to do so." Cappelen similarly argues that conceptual engineers should be more like political theorists who

> are comfortable reflecting on and proposing theories of justice, for example, without having a recipe for how they can be implemented. More generally, we can make judgments about what ought to be the case, without knowing how to make the world that way (or even having a plausible strategy in mind). Those are just different projects: figuring out what ought to be the case and making the world that way. (Cappelen 2018, 83–84)

This objection misconstrues the scope of my argument. The argument is not meant to convince semantic externalists that conceptual revision is possible (nor does it assume that it is!). The argument should be viewed conditionally. If the goal is conceptual improvement, then the argument specifies what is required for evaluating that improvement. If improvement isn't valued, or if meaning change really is untenable, the conclusion will seem inconsequential. That said, semantic externalists should welcome this kind of argument even with their doubts about feasibility. The

empirical work required for determining the success of CE projects will have implications for confirming or disconfirming semantic externalism. If the majority of CE projects are unsuccessful, this could be evidence semantic externalists appeal to in arguing that CE be understood as a form of ideal-theorizing. Analogously, CE as non-ideal theorizing could be vindicated by data showing recommended conceptual changes are, at least in some cases, successfully implemented. I have no hunches about which way it will go; we have to let the evidence speak. In order for that to happen, conceptual engineers need to take steps in acquiring the relevant data.

Another objection might be made in light of a distinction between practical success and theoretical success (Koch 2019). For Koch, theoretical success is conditional: "if people were to use it [the engineered concept], they would make epistemic progress" (ibid., 709). Koch also seems to support the purely normative inquiry that Burgess, Plunkett, and Cappelen advocate for: "In general, which concepts we do use is often influenced by factors which have nothing to do with whether we should use them" (ibid., 709). In contrast, practical success concerns "whether people are actually ready to use the concepts they should use" (ibid., 709). Koch departs from Burgess, Plunkett, and Cappelen in thinking the normative question is the only one relevant for CE: "If one's goal is to make actual progress in a given domain, both success conditions matter" (ibid., 709). However, Koch seems to believe the success conditions for the theoretical and the practical are different and should not be conflated:

> Even if you believe that in order to do conceptual engineering well, you need to pay attention to both practical and theoretical standards, you are well-advise to keep them separate. This is because the kind of factors which determine whether a concept meets the theoretical success condition is usually very different from the kind of factors which determine whether it meets the practical one. What matters for theoretical success are broadly epistemic factors: whether the concept in question allows us to make useful and novel distinctions, whether it allows us to secure the progress we already have made, whether it allows us to view a matter in a new and interesting light, etc. What matters for practical success, on the other hand, are psychological and sociological factors: whether the concept triggers positive reactions, whether it 'sticks' to people's minds, whether the people advocating it has enough influence or political power to make it spread, etc. Given that these factors are typically very different, investigations into them take different forms. Evaluating theoretical success is often an issue for philosophers of science or for the philosophically-minded subgroup of the theoretical community in which the concept is used. Evaluating practical success, on the other hand, will typically be a matter of social psychology and sociology. (Koch 2019, 709–710)

The potential objection then is that my argument is limited to practical success. Perhaps Koch is correct that the empirical work done in social science is only for evaluating practical success not theoretical. Even if the argument was limited to practical success, that is still considerable progress. As Koch correctly states, both kinds of success matter; so an argument that elucidates one of the two components for successful CE projects is a valuable contribution. Nonetheless, the claim that empirical

methods are not germane to evaluating theoretical success is dubious. Whether or not a concept has certain theoretical virtues is in many cases (though perhaps not all) an empirical question.

Take, for example, the presumed theoretical virtue of parsimony. Parsimony can be defined in a variety of ways against a variety of objectives. There are at least three in the philosophy of science literature: a quantitative concept that refers only to the number of processes, entities, causes, or basic principles postulated; a qualitative concept that refers to the number of types of processes, entities, causes, or basic principles postulated; and a Bayesian concept based on Reichenbach's principle of common causes (Sober 2015). Whether a particular explanation or theory is *simple* according to any of these definitions is not a matter of mere philosophical reflection. Each of these concepts gives very specific criteria by which to gauge simplicity and such evaluations require "looking out" to determine if those criteria are met.[8] Thus, it's not surprising that parsimony claims in evolutionary biology are often challenged or overturned in light of new phylogenetic analyses, ecological data, and other relevant experimental findings.[9]

Consider the Knobe-effect mentioned before as another example. Knobe (2003) asked participants to consider a scenario where the boss of a company decides to start a new program to increase profits knowing there will be certain foreseen side effects in implementing the program. Depending on whether the known side effect either helped or harmed the environment, participants gave opposing answers to the question of whether the boss *intentionally* brought about the side effect. Recall Koch's first condition for theoretical success was that "the concept in question allows us to make useful and novel distinctions" (Koch 2019, p. 709). One thing Knobe's **experimental** results did was reveal a previously unknown feature about the intentionality concept that was central in the philosophy of action and ethics at the time. Before Knobe's studies, the prevailing view was that an agent must have intended to perform an action in order for that person to be held morally responsible. What the data surprisingly revealed, however, was that the moral valence of a consequence influenced participants' intentionality judgments. Most participants believe an act is intentional if the consequence is harmful, and most believe an act is *not* intentional if the consequence is helpful.

Relatedly, Koch's second criteria for theoretical success, "whether it allows us to secure the progress we have already made," is also amenable to empirical investigation. It is important to ask exactly how much theoretical progress we have made and to gauge that progress with the relevant data. Sober explains this point well with regards to the theoretical virtue unification: "We must avoid the mistake of selective attention; we need to count the failures as well as the successes … we need to watch out for a selection bias; the apparent track record to date overestimates

[8] Machery (2017) makes a similar point related to what he calls "prescriptive conceptual analysis." Machery argues that given some specified set for normative constraints, experimental philosophy can help determine which concept proposals meet those constraints.

[9] A great illustration of the connection between empirical work and concept determination in ecology can be found in Justus (2012).

the success rate of unification and it is hard to say by how much" (Sober 2015, pp. 46–47). Again, Knobe's study compels us to question the presumed progressiveness of 'intentionality' in the philosophy of action. Moreover, ameliorative or revisionary CE projects start from the presupposition that our current concepts are defective insofar as they have impeded certain forms of progress. Empirical work is used to support such an assumption. Research on gender and racial bias reinforce this starting point for Haslanger's (2000) analysis and suggested revisions of gender and race concepts, for instance. So, it's possible to maintain the importance of Koch's distinction while recognizing that evaluating theoretical progress, and determining whether our engineered concepts have contributed to that progress, can (and should) be informed by empirical investigation.

3 The Argument: Part 2

The argument in the previous section sought to establish the necessary role of empirical methods for evaluating the success of CE projects. The conclusion that CE necessarily involves empirical work suggests a related methodological consideration. As mentioned in the introduction, Carnapian Explication is given extensive homage in the CE literature. However, Explication is typically characterized as a subtopic of CE with its application being wrongly perceived as limited to scientific and logical concepts:

> Carnap aimed to replace certain everyday concepts with 'exact and fruitful' concepts for use in the formal or empirical sciences … Carnap was interested in conceptual engineering, primarily in the sense of devising new, technical languages. (Thomasson 2020, 6)

> Carnap's notion of explication, however, is narrower than the activity I'm interested in. (Cappelen 2018, 11)

> For some purposes, formulating either empirical laws or logical theorems is not what really matters—just think of typical philosophical inquiries, e.g. into the nature of justice or knowledge. (Koch 2019, 702)

> Carnap's proposal is not, as it stands, suitable for this particular task. (Pinder 2020, 4)

> 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. (Nado 2019a, 3)

Although Carnap focused on logical and empirical explications throughout his career, it's a mistake to conclude that Explication, as Carnap described, is only applicable for those concepts or that formal explications could only be deemed successful. Carnap did not see Explication as carving out a joint between

philosophical and scientific concepts: "I see here no sharp boundary line but a continuous transition" (Carnap 1963, p. 934). Further, he explicitly acknowledged the limits of formal methods for such purposes. This is clear from his reply to Strawson:

> The use of symbolic logic and of a constructed language system with explicit syntactical and semantical rules is the most elaborate and most efficient method. For philosophical explications the use of this method is advisable only in special cases, but not generally… Again misled by his sharp distinction between scientific language and ordinary language, Strawson seems to misunderstand this point. He believes that the explicatum is meant to serve a scientific purpose, in distinction to the explicandum which serves a pre-scientific purpose. (Carnap 1963, 935)

Recall that premise two of the first argument was neutral with respect to the function of our target concepts. Thus, the scope of the argument extends across all varieties of CE projects (scientific progress, political amelioration, etc.). This wide range coupled with the textual evidence that Carnap saw Explication as a generalizable methodology motivated my suspicion that conceptual engineers have missed something important. In particular, discussion of the third Explication criterion *precision* is strikingly absent from the CE literature. At one point Carnap even claims that "[t]he only essential requirement is that the explicatum be more precise than the explicandum" (1963, p. 936). Perhaps conceptual engineers have mistakenly assumed that precision was included for its importance in explicating scientific and logical concepts, and thus wouldn't be relevant for non-scientific CE tasks. Understanding why Carnap included precision as an Explication criterion clarifies why, and demonstrates how, it is pertinent to all CE projects.

Carnap's claim that precision is the "only essential requirement" seems to suggest that he viewed it as the most important condition for an Explication. However, several scholars have made clear that Carnap saw fruitfulness as the ultimate success metric and thus gave it the most evaluative weight (see Brun 2016; Carus 2007; Justus 2012; Novaes 2018; Schupbach 2017). Precision is included and given significant attention because of its relationship to fruitfulness. As Justus (2012, p. 168) explains, "precision is paramount because it *usually* enhances fruitfulness." Although Carnap denied a sharp philosophy and science divide, he took the unparalleled epistemic success of science as a model to inform philosophical methodology. Carnap's great insight was that vague concepts are rarely components of well-confirmed generalizations. Increasing precision, on the other hand,

> often facilitates in developing and discovering well-confirmed generalizations… Without sufficiently precise concepts, it is difficult if not impossible to derive predictions from statements containing them. Without such predictions, in turn, statements cannot be confirmed or disconfirmed. Testing predictions of hypotheses and theories against data is a staple of scientific methodology that has clearly proved to be an epistemically reliable basis for inference. Precision facilitates this methodology and its success therefore grounds the exactness [precision] criterion. (Justus 2012, 169)

How exactly does precision reliably increase fruitfulness? The history of scientific progress suggests that increasing precision "usually enhances mathematical rigor, measurability, testability, [and] theoretical unification..." (Shepherd and Justus 2015, p. 338). Although this list is targeted at empirical concepts, there are two items that are crucial for any conceptual engineering project—measurability and testability. As premise two of the previous argument made clear, if we want to evaluate conceptual improvement, then regardless of the kind of target concept (scientific, social, political, metaphysical, etc.), we must have some way to measure/test the degree to which the engineered concept functions better or worse than the original.[10] Thus, the likelihood that more precise concepts will increase measurability and testability provides a strong motivation for making precision an essential component of CE methodology. Here is the whole argument:

(P1) The goal of conceptual engineering is to evaluate and improve concepts.
(P2) If conceptual improvement is the goal, then the metric for success will be the degree to which an engineered concept makes progress towards achieving its function compared to the original.
(P3) Evaluating that success requires information about the functionality of the original concept *and* the engineered concept.
(P4) Acquiring that information requires empirical investigation.
(C1) Conceptual engineering necessarily involves empirical work.
(P5) Given the empirical necessity for evaluating the success of engineered concepts, we ought to engineer concepts that are empirically measurable/testable.
(P6) Increasing precision *typically* facilitates the evaluation of a concept's functionality by increasing experimental measurability/testability.
(C2) Precision ought to be a methodological standard in conceptual engineering.

The rationale behind premise five is straightforward: if evaluating the success of engineered concepts matters, then engineering concepts with features that aid in that evaluation should be a priority. Again, if you don't think feasibility is important for CE then you won't find this compelling. But most conceptual engineers do care about actual success, especially those interested in scientific and sociopolitical progress. The justification for premise six was given above by clarifying Carnap's rationale for including precision as an Explication criterion. Two clarifications about the argument are noteworthy.

[10] Notably, in his work on Interpretation and Preciseness (1950), Arne Naess endorses the importance of precision in testing and modifying conceptual frameworks. Naess also undertakes his own experimental philosophy studies on folk uses of various semantic concepts like 'synonymity.' The results of these studies leads Neass to a similar conclusion of the first argument presented here: "acceptance of intuitions reported by the philosophically sophisticated about the verbal and conceptual habits of others leads to confusion and error… empirical procedures should be applied to empirical questions. When philosophers offer conflicting answers to questions that have empirical components, empirical research is needed." Thanks are owed to an anonymous reviewer for bringing Naess's experimental work and Carnap's (1955) praise of that work to my attention.

First, a worry might be that the argument does not provide any reason to think it's easier to assess function-achievement of precise concepts than imprecise ones. Rather, what the argument shows is that conceptual engineers need to be more precise about 'concept function' or 'function achievement' in order to investigate concept functionality. The notions of concept function or concept achievement are, of course, amenable to more precification, but without the *concept* itself being sufficiently precise it will be difficult to formulate predictions from statements, descriptions, hypothesis, or explanations containing them. Without such predictions, the concept's functionality cannot be confirmed or disconfirmed.

Second, notice the preceding remark said *sufficiently* precise. This is important because it would be a mistake to think the argument requires all concepts to be maximally precise. Carnap characterized non-logical concepts in order of increasing precision: classificatory, comparative, and quantitative (Carnap 1950, p. 12). Even though increasing precision along these dimensions typically facilitates evaluating concept functionality, it is a truism this is not always the case. The argument allows for decreases of precision when warranted. Again, for Carnap,"Precision for precision's sake is not the agenda. Rather, enhancing precision usually enhances fruitfulness, which is the agenda" (Shepherd and Justus 2015, 388). In cases where the added precision does not result in increases in fruitfulness, or even decreases fruitfulness, then the focus should be on finding fruitful comparative or classificatory explicata (Carnap 1950, p. 14). Nonetheless, it remains necessary for appraising conceptual improvement that functionality can be measured. In cases where quantified methods seem ill-suited, or are unavailable, qualitative measures can be used. Given that some concepts of interest to conceptual engineers might not be improved by quantification, this flexibility is a feature and not a bug of the argument. To buttress this conclusion, the next section will examine a well-known debate in the literature that demonstrates how overlooking precision impedes progress in CE.

4 Carnap Versus Strawson, *Again*.

The title of this section naturally suggests I'll be discussing the unduly famous 'change of subject' objection Strawson (1963) levels against Carnap. But that's not the case. I want to analyze the modern rendition of this dispute that occurs in the feminist literature on race and gender. In this production, Jennifer Saul and Sally Haslanger are our contemporary equivalents of Peter Strawson and Rudolf Carnap. As illustrated by a brief comparison of the two parties, Strawson and Saul are conservatives; in Carnapian terms they prioritize similarity:

> The actual use of linguistic expressions remains his [a philosopher's] sole and essential point of contact with reality; for this is the only point from which the actual mode of operation of concepts can be observed. (Strawson 1963, 508)

> It seems to me that communication is difficult enough as it is, and that we should instead try to use ordinary terms in as ordinary a way as possible. (Saul 2006, 141)

Strawson and Saul also appear critical of formal or technical explications in general:

> 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 text-book on physiology to someone who says (with a sigh) that he wished he understood the workings of the human heart. (Strawson 1963, 505)

> Haslanger's definitions are, as she concedes, at odds with intuitions. For most of us, Haslanger's definitions are just not the sort of thing we are likely to come up with if asked about the nature of race and gender, and when presented with them we will very likely reject them. (Saul 2006, 122)

> We may well have a need for technical terms that work as Haslanger's do, but it seems to me a mistake to risk confusion by using ordinary vocabulary to do this work. (Saul 2006, 141)

Carnap and Haslanger, on the other hand, are revisionaries. What they value most is fruitfulness:

> When we compare the explicandum Fish with the explicatum Piscis, we see that they do not even approximately coincide… The change which zoologists brought about in this point was not a correction in the field of factual knowledge but a change in the rules of the language… they realized the fact that the concept Piscis promised to be much more fruitful than any concept more similar to Fish. (Carnap 1950, 6)

> I've cast my inquiry as an analytical—or what I here call an ameliorative—project that seeks to identify what legitimate purposes we might have (if any) in categorizing people on the basis of race or gender, and to develop concepts that would help us achieve these ends. I believe that we should adopt a constructionist account not because it provides an analysis of our ordinary discourse, but because it offers numerous political and theoretical advantages. (Haslanger 2012, 366)

Although Haslanger and Carnap obviously have different subject areas in mind (politics for Haslanger and science for Carnap), they clearly share a commitment to improving our concepts to achieve various goals. I am not going to focus on the change of subject objection or, more generally, whether similarity or fruitfulness should be given the most weight. There are already plenty of arguments in the literature regarding that particular trade-off. What is interesting about the debate between Saul and Haslanger is that closely examining the dialectic reveals why precision is paramount for CE.

First, a very general remark: if we invoke only two criteria for assessing the success of an engineered concept, then any time the two criteria diverge from each other we are left at a standstill. The Saul–Haslanger and Strawson–Carnap debates are presented as involving a false dichotomy: we either go with a fruitful explicata *or* go with a similar explicata. But if we have more than two criteria, then those other standards will be brought to bare on our conceptual decision-making. For example, if one explicata ranks high in fruitfulness and another high in similarity, we should also look at how they measure against simplicity and precision. Of course, using four criteria as opposed to two does not guarantee there will never be conflict; simplicity and precision can often pull apart the way fruitfulness and similarity do. Trade-offs between the four criteria will always have to be made, and on a case-by-case basis. The point is that rather than making our conceptual decisions based on a false dichotomy between fruitfulness and similarity, including all the explication criteria gives us much more evaluative capability.[11]

Second, besides the problem of a false dichotomy, it appears failing to account for precision left Haslanger's analysis of gender and race open to a set of objections concerning measurability.[12] Saul (2006, p. 138), for instance, claims that when it comes to the supposed benefits of Haslanger's proposed definitions, "there are many uncertain empirical matters." Curiously, it seems lacking empirical tractability is one of Saul's main points of contention:

> How, then, could we ever hope to study our actual usage of race concepts, as opposed to colour concepts? … It strikes me as unlikely that we would have much evidence enabling us to attribute such a distinction in a non-question-begging manner, collecting together one set of usages of 'Black' as telling us about a race concept and another as telling us about a colour concept. (ibid., 132)

> In order to support this view, we would need to find behavioural evidence that speakers do in fact classify people in this way. It seems to me that such evidence is unlikely to be forthcoming, given both the muddled linguistic usage and its tendency to go unnoticed. (ibid., 131)

> To learn about the operative concept, what we really need to do is find some females who are not subordinated on the basis of real or perceived sex characteristics and check to see whether people apply the term 'woman' to them. But this is far from straightforward… (ibid., 129)

[11] See Olsson (2015) for a compelling case study using the four explication criteria to adjudicate theories in epistemology.

[12] Haslanger's proposed definition of 'woman' is the following: "S is a woman if S is systematically subordinated along some dimension (economic, political, legal, social, etc.) and S is 'marked' as a target for this treatment by observed or imagined bodily features presumed to be evidence of a female's biological role in reproduction" (2000, p. 39). Similarly, for race, "A group is racialized if its members are socially positioned as subordinate or privileged along some dimension (economic, political, legal, social, etc.) and the group is 'marked' as a target for this treatment by observed or imagined bodily features presumed to be evidence of ancestral links to a certain geographical region" (ibid., 44).

> Haslanger recognizes that the proponent of a revisionary analysis faces an additional issue: when does an analysis cross the boundary from being revisionary to being an analysis of something entirely different?… [S]he suggests two criteria for deciding this, a semantic and a political one … The semantic condition, she suggests, is far more straightforward: 'the proposed shift in meaning of the term would seem semantically warranted if central functions of the term remain the same, e.g., if it helps to organize or explain a core set of phenomena that the ordinary terms are used to identify or describe.' Haslanger seems to think that her analyses of race and gender terms will pass these tests, although she does not argue this in detail… *it is far from obvious how the phenomenon in question should be specified*. (ibid., 134; my emphasis)

It is crucial to situate these criticisms in the broader context of the Saul–Haslanger debate to appreciate their bite. Saul's critique is primarily referring to what Haslanger calls the "descriptive project," which aims to "identify and explain persistent inequalities" (Haslanger 2000, p. 36). This is contrasted with the "ameliorative project," which aims at providing conceptual "tools in the fight against injustice" (ibid., 36). Despite their distinct goals, the two are not unrelated. As Haslanger (2006, p. 96) rightly admits these "projects cannot, of course, be kept entirely distinct" and that "we might want to know the 'folk theory' of race (and other categories) in order to engage in hermeneutical deliberation about what we have meant and should mean" (2010, p. 180). Not only are the projects intertwined, but Haslanger claims that continuity between the concepts coming from different projects is conceivably the final goal:

> [W]e might hope that through reflection and discussion we could come to the point where (a) the concept we take ourselves to be employing, (b) the concept that best captures the type we are concerned with, and (c) the type we ought to be concerned with coincide. In such cases the conceptual, descriptive, and ameliorative projects yield the same concept. (Haslanger 2006, p. 96)

Accordingly, the success of Haslanger's definitions (with regards to the descriptive *and* ameliorative aims) requires that the folk have concepts of race and gender that extend beyond mere color and sex concepts. Saul's critique challenges this possibility (more on this below). Notably, Herman Cappelen endorses Saul's criticisms but takes it one step further by suggesting those difficulties are insurmountable:

> What Haslanger's account needs, but doesn't provide, is a procedure for identifying 'the phenomena' that are being organized and explained. Haslanger has never responded to Saul's objection and my bold conjecture is this: *there simply isn't a good way to identify 'the phenomenon'* except disquotationally. (Cappelen 2018, 184; my emphasis)

Saul is not as pessimistic as Cappelen, but strangely concludes there is no more work for the philosopher to do:

> It may well be that Haslanger's terminology is so useful that this usefulness outweighs the ways in which I have worried that they might be counterpro-

> ductive. But we need to know more in order to be able to evaluate this…. So, in the end, it seems to me that what really needs to be done is some careful empirical work on the most effective vocabulary for combating racism and sexism. And this takes us outside the purview of philosophy. (Saul 2006, 142)

Saul and Cappelen are both mistaken. First, there is a good way to identify the phenomena that are the target of CE projects—by doing experimental philosophy! Recall that the positive program is aimed precisely at that goal; those philosophers are "motivated to explore intuitions experimentally because they think that by doing so they can do a better job of conceptual analysis" (Stich and Tobia 2016). And as mentioned in the introduction, the use of experimental philosophy in methods of explication has already been advocated for by Shepard and Justus (2015), Pinder (2017), and Schupbach (2017). Second, the experimental methods needed to further the analysis of Haslanger's proposed definitions fall *inside* the purview of philosophy, and especially of CE. Outside the argument given in this paper, one can simply look at the activities of other philosophers. For example, Paul Griffiths's empirical work on genetic essentialist thinking, Josh Knobe's studies on the factors influencing our moral judgments, Walter Sinnott-Armstrong's studies on moral cognition and decision-making, and Edouard Machery's investigations of knowledge attribution across cultures effectively rebuts this dogmatic philosophy and science divide.

Crucially, as explained in Sect. 3, increasing precision usually increases measurability and testability. Perhaps if Haslanger had included precision as a guide at the start of her engineering project, the empirical difficulties highlighted by Saul could have been avoided or accounted for.[13] Fortunately for Haslanger, there is a sizable literature on race and gender concepts that provides the resources to rebut Saul's criticisms. One example, in particular, ought to convey the importance and relevance of that research. As mentioned above, Saul casts serious doubt on whether Haslanger's ameliorative concepts represent the operative race and gender concepts of ordinary speakers. Again, this requires that the folk distinguish color from race concepts and sex from gender concepts. But Saul boldly conjectures the folk do not have anything resembling the kind of concepts Haslanger's ameliorative definitions presume:

> It seems to me that there is no good reason to even attribute gender or race concepts to ordinary speakers. Ordinary speakers do, of course, have concepts associated with the terms 'woman' and 'Black', but there is no reason to take these to be gender and race concepts—that is, to take them to be social correlates of sex and colour concepts. Moreover, it does not seem as though ordinary speakers have a tendency to organize the world (linguistically or otherwise) according to categories that really look like gender and race categories—rather than sex and colour categories, or some muddled amalgams. Why should we, then, attribute race and gender concepts to ordinary speakers? I'd

[13] Despite Saul's criticisms, it's noteworthy that the exactness of Haslanger's definitions reflect an increase in precision that, as Brigandt and Rosario (2020, p. 104) note, defeated "prior scepticism about the possibility of putting forward a coherent concept of gender."

> like to tentatively suggest that we should not do so, although properly establishing this would require more argument than I am in a position to make here. If this is right, then it makes no sense to even ask whether Haslanger's analyses of race and gender concepts are revisionary with respect to ordinary ones: there are no ordinary race and gender concepts. (Saul 2006, 133)

Saul argues that ordinary speakers do not distinguish between color/race and sex/gender in the way that tracks Haslanger's proposed race and gender concepts. But this claim is demonstrably false. Numerous studies on race and gender categorization, stereotype bias, self-concepts, and identity make obvious the fact that ordinary people have race and gender concepts that are more robust than mere associations with sex and color categories. In fact, several studies have actually investigated the content of racial concepts among the folk (see Martin and Parker 1995; Condit et al. 2003, 2004; Dubriwny et al. 2004; Jayaratne et al. 2006; Glasgow et al. 2009). Taken together, the results indicate that, in general, race is conceived of as a complex combination of ancestral, phenotypic, and social factors. In their research on focus groups, for example, Dubriwny et al. (2004, p. 187) explicitly asked participants, "What do you think is generally meant when people use the term 'race'?" That question was followed by others to discern whether participants were defining race by culture, geography, heredity or genetics, color, and religion. The results demonstrated that participants thought of race as multifactorial, that is, containing a mix of genetic, geographic, cultural, and socially constructed features. Similarly,

> Shulman and Glasgow (in press) administered a questionnaire that directly asked adults whether they think race is real or merely imagined, and then asked the realists to select whether people's races are determined by the way they look (classified as 'biological'), their social ties ('social'), their personality or abilities ('psychological'), or some combination thereof. Only about half of realists chose biology alone, and another 23% chose biology in combination with one or both of the other two determinants. Twenty-one percent chose a purely social conception of race. These data, along with the results of Dubriwny et al., challenge both the claim that race-thinking is univocally biological (or univocally social) and the conventional wisdom that race-thinking uniformly subscribes to essentialism and the one-drop rule.[14] (Glasgow et al. 2009, 18)

The takeaway is that empirical work on race and gender is crucial to philosophical investigations of these topics. In fact, in a recent comparison between Carnapian Explication and Ameliorative Analysis, Novaes (2018, p. 1027) notes that the precision criterion "has no immediate counterpart in Haslanger's framework" and suggests that "[t]his is, in turn, where practitioners of the ameliorative method may benefit from engagement with the techniques employed by Carnapian explicators."

[14] The 'one-drop rule' refers to the idea that race is typically thought of in terms of biological essences, such that if a person has even one Black ancestor, he or she is Black, for example.

Exactly. My analysis of the debate between Saul and Haslanger demonstrates the benefit of including precision in ameliorative projects. Given the epistemic rewards that typically come from increases in precision, it would behoove conceptual engineers to make precision a methodological standard in guiding future CE projects.[15]

5 Conclusion

I first argued that given the stated goals of conceptual engineers, empirical work is required in order to evaluate the success of CE projects. Second, I argued that precision's utility in generating concepts amenable to those empirical methods warrants making it a methodological standard in CE. Finally, I demonstrated how ignoring precision can interfere with philosophical progress by highlighting where lack of precision left Haslanger exposed to Saul's objections. The upshot, however, was that this also revealed Haslanger has the resources to rebut those criticisms. To conclude, I will briefly mention how the methodological analysis presented here relates to a different methodology and what implications follow.

The so-called 'method of cases' has become prominent in discussions of philosophical methodology, especially in light of what has been dubbed the "experimenters challenge" (Nado 2015). The method of cases refers to the traditional practice of presenting cases (thought experiments) which describe actual—but more often hypothetical—situations and take our intuitions about the respective cases as evidence for what facts hold in the situation. Those intuition-inferred facts are then brought to bear on philosophical controversies (Machery 2017). The experimenter's challenge refers to the large body of empirical evidence (some of which was summarized in this paper) that raises serious worries about the cognitive value of intuitions and chips away at the method of case's epistemic foundation. The argument presented here is not directly aimed at the method of cases debate, but it is obviously sympathetic to the experimenter's challenge. That said, the scope of my argument pertains to CE specifically. The argument should not be taken as purporting to establish an empirical methodology for philosophical inquiry *tout court*. This analysis leaves open the possibility that other methodologies may be fruitful for non-CE philosophical projects, a conclusion which is in line with Carnap's cooperative attitude:

> We all agree that it is important that good analytic work on philosophical problems be performed. Everyone may do this according to the method which seems to be the most promising to him. The future will show which of the two methods, or which of the many varieties of each, or which combinations of both, furnishes the best results. (Carnap 1963, 939)

Lastly, the argument has implications for the controversy surrounding the metaphysics of concepts. Given that our concepts need to be empirically tractable in order

[15] Haslanger has recently endorsed the interdisciplinary approach being argued for here. Haslanger (2020, 7) asserts that "the social critic embraces the normative dimension of philosophical theorizing, and also relies crucially on empirical research".

to gauge their success, a naturalized account of concepts would be best suited for the purposes of CE. One option from Machery (2009) holds that concepts are psychological entities. Another is the causal theory of conceptual content from Prinz (2002). A third might be the theory-theory view, which claims that concepts are relational in the same way the content of scientific terms are specified by the role they play within a particular scientific theory. Analyzing these possibilities is outside the purview of this paper. I mention them only to highlight the interesting implications the methodological argument here has for ontology.

References

Adleberg, T., Thompson, M., & Nahmias, E. (2014). Do men and women have different philosophical intuitions? Further Data. *Philosophical Psychology*, *28*, 1–27.

Alexander, J., & Weinberg, J. M. (2007). Analytic epistemology and experimental philosophy. *Philosophy Compass*, *2*(1), 56–80.

Bird, A. (2007). What is scientific progress? *Noûs*, *41*(1), 64–89.

Brigandt, I., & Rosario, E. (2020). Strategic conceptual engineering for epistemic and social aims. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 100–124). Oxford University Press.

Brun, G. (2016). Explication as a method of conceptual re-engineering. *Erkenntnis*, *81*(6), 1211–1241.

Burgess, A., & Plunkett, D. (2013). Conceptual ethics I. *Philosophy Compass*, *8*(12), 1091–1101.

Cappelen, H. (2018). *Fixing language: An essay on conceptual engineering*. Oxford University Press.

Carnap, R. C. (1945). The two concepts of probability: The problem of probability. *Philosophy and Phenomenological Research*, *5*(4), 513–532.

Carnap, R. C. (1950). *Logical foundations of probability*. University of Chicago Press.

Carnap, R. C. (1955). Meaning and synonymy in natural languages. *Philosophical Studies*, *6*(3), 33–47.

Carnap, R. C. (1963). Replies and systematic expositions. In A. S. Paul (Ed.), *Iteschilpp* (pp. 859–1013). Prc. Open Court.

Carus, A. W. (2007). *Carnap and twentieth-century thought: Explication as enlightenment*. Cambridge University Press.

Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). Epistemic intuitions in fake-barn thought experiments. *Episteme*, *11*(2), 199–212.

Condit, C., Templeton, A., Bates, B. R., Bevan, J. L., & Harris, T. M. (2003). Attitudinal barriers to delivery of race-targeted pharmacogenomics among informed lay persons. *Genetics in Medicine*, *5*, 385–392.

Condit, C. M., Parrott, R. L., Harris, T. M., Lynch, J., & Dubriwny, T. (2004). The role of 'genetics' in popular understandings of race in the United States. *Public Understanding of Science*, *13*, 249–272.

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, *108*(2), 353–380.

De Cruz, H. (2017). Religious disagreement: An empirical study among academic philosophers. *Episteme*, *14*(1), 71–87.

Dubriwny, T. N., Bates, B. R., & Bevan, J. L. (2004). Lay understandings of race: Cultural and genetic definitions. *Community Genetics*, *7*, 185–195.

Eklund, M. (2002). Inconsistent languages. *Philosophy and Phenomenological Research*, *64*(2), 251–275.

Eklund, M. (2006). What good are our intuitions? Philosophical analysis and social kinds. *Proceedings of the Aristotelian Society Supplementary*, *80*(1), 89–118.

Eklund, M. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Eklund, M. (2015). Intuitions, conceptual engineering, and conceptual fixed points. In C. Daly (Ed.), *The Palgrave Handbook of Philosophical Methods*.

Glasgow, J., Shulman, J., & Covarrubias, E. (2009). The ordinary conception of race in the United States and its relation to racial attitudes: A new approach. *Journal of Cognition and Culture*, *9*(1–2), 15–38.

Hanna, J. F. (1968). An explication of "explication". *Philosophy of Science*, *35*(1), 28–44.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Nous*, *34*, 31–55.
Haslanger, S. (2006). What good are our intuitions: Philosophical analysis and social kinds. *Aristotelian Society Supplementary*, *80*(1), 89–118.
Haslanger, S. (2010). Language, politics, and "the folk": Looking for "the meaning" of 'race'. *The Monist*, *93*(2), 169–187.
Haslanger, S. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.
Jayaratne, T. E., Ybarra, O., Sheldon, J. P., Brown, T. N., Feldbaum, M., Pfeffer, C. A., et al. (2006). White Americans' genetic lay theories of race differences and sexual orientation: Their relationship with prejudice toward blacks, and gay men and lesbians. *Group Processes and Intergroup Relations*, *9*, 77–94.
Justus, J. (2012). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science*, *2*(2), 161–179.
Knobe, J. (2003). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology*, *16*(2), 309–325.
Koch, S. (2019). Carnapian explications, experimental philosophy, and fruitful concepts. *Inquiry: An Interdisciplinary Journal of Philosophy*, *62*(6), 700–717.
Lindauer, M. (2019). Experimental philosophy and the fruitfulness of nor-mative concepts. *Philosophical Studies, 177*, 2129–2152.
Machery, E. (2009). *Doing without concepts*. Oxford University Press.
Machery, E. (2017). *Philosophy within its proper bounds*. Oxford University Press.
Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics, cross-cultural style. *Cognition*, *92*(3), 1–12.
Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., & Hashimoto, T. (2015). Gettier Across Cultures. Noûs, 645–664.
Maher, P. (2007). Explication defended. *Studia Logica*, *86*(2), 331–341.
Martin, C. L., & Parker, S. (1995). Folk theories about sex and race differences. *Personality and Social Psychology Bulletin*, *21*, 45–57.
May, J., Sinnott-Armstrong, W., Hull, J. G., & Zimmerman, A. (2010). Practical interests, relevant alternatives, and knowledge attributions: An empirical study. *Review of Philosophy and Psychology*, *1*(2), 265–273.
Nado, J. (2015). Philosophical expertise and scientific expertise. *Philosophical Psychology*, *28*(7), 1026–1044.
Nado, J. (2019a). *Conceptual engineering via experimental philosophy*. In *T and f: Inquiry* (pp. 1–21).
Nado, J. (2019b). Conceptual engineering, truth, and efficacy. *Synthese*. https://doi.org/10.1007/s11229-019-02096-x
Naess, A. (1950). Interpretation and preciseness. II. *Journal of Symbolic Logic*, *15*(1), 73–74.
Neta, R. (2012). Knowing from the armchair that our intuitions are reliable. *The Monist*, *95*(2), 329–351.
Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, *41*(4), 663–685.
Novaes, C. D. (2018). Carnapian explication and ameliorative analysis: A systematic comparison. *Synthese*, *197*(3), 1011–1034.
Olsson, E. J. (2015). Gettier and the method of explication: A 60 year old solution to a 50 year old problem. *Philosophical Studies*, *172*(1), 57–72.
Pinder, M. (2017). Does experimental philosophy have a role to play in Carnapian explication? *Ratio*, *30*(4), 443–461.
Pinder, M. (2020). What ought a fruitful explicatum to be? *Erkenntnis*, 1–20.
Plakias, A. (2015). Experimental philosophy. In *Oxford Handbooks Online*. https://doi.org/10.1093/oxfordhb/9780199935314.013.17.
Plunkett, D., & Cappelen, H. (2020). A guided tour of conceptual engineering and conceptual ethics. In H. Cappelen, D. Plunkett, & A. Burgess (Eds.), *Conceptual engineering and conceptual ethics*. Oxford University Press.
Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. MIT Press.
Prinzing, M. (2018). The Revisionist's rubric: Conceptual engineering and the discontinuity objection. *Inquiry: An Interdisciplinary Journal of Philosophy,61*(8), 854–880.
Quine, W. V. O. (1951). Two Dogmas of Empiricism. *Philosophical Review*, *60*(1), 20–43.
Quine, W. V. O. (1960). *Word and object*. MIT Press.
Quine, W. V. O. (1986). Reply to white. In L. Hahn & P. Schilpp (Eds.), *The Philosophy of W.V. Quine* (pp. 663–665). Open Court.

Railton, P. (2003). *Facts, values, and norms: Essays toward a morality of consequence*. Cambridge University Press.
Reck, E. (2012). Carnapian explication: A case study and critique. In P. Wagner (Ed.), *Carnap's ideal of explication and naturalism* (pp. 96–116). Palgrave-Macmillan.
Saul, J. (2006). Gender and race. *Aristotelian Society Supplementary*, *80*(1), 119–143.
Scharp, K. (2013). *Replacing truth*. Oxford University Press.
Schupbach, J. N. (2017). Experimental explication. *Philosophy and Phenomenological Research*, *94*(3), 672–710.
Schwitzgebel, E., & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127–137.
Seyedsayamdost, H. (2015). On normativity and epistemic intuitions: Failure of replication. *Episteme*, *12*(1), 95–116.
Shepherd, J., & Justus, J. (2015). X - phi and carnapian explication. *Erkenntnis*, *80*(2), 381–402.
Sober, E. (1990). Contrastive empiricism. In C. Wade Savage (Ed.), *Scientific Theories* (pp. 392–410). University of Minnesota Press.
Sober, E. (1999). Testability. *Proceedings and Addresses of the American Philosophical Association*, *73*(2), 47–76.
Sober, E. (2015). *Ockham's Razors: A user's manual*. Cambridge University Press.
Stein, H. (1992). Was Carnap entirely wrong, after all? *Synthese*, *93*(1–2), 275–295.
Stich, S., & Tobia, K. (2016). Experimental philosophy and the philosophical tradition. In J. Sytsma & W. Buckwalter (Eds.), *Blackwell companion to experimental philosophy*. Wiley-Blackwell.
Strawson, P. F. (1963). Carnap's views on conceptual systems versus natural languages in analytic philosophy. In A. S. Paul (Ed.), *The Philosophy of Rudolf Carnap* (pp. 503–518). Open Court.
Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171.
Swain, S., Alexander, J., & Weinberg, J. M. (2008). The instability of philosophical intuitions: Running hot and cold on truetemp. *Philosophy and Phenomenological Research*, *76*(1), 138–155.
Thomasson, A. (2020). A pragmatic method for normative conceptual work. In *Conceptual Engineering and Conceptual Ethics* (pp. 435–458).
Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology*, *26*(5), 629–638.
Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics*, *29*(1–2), 429–460.