CrossMark

Carnapian explication and ameliorative analysis: a systematic comparison

Catarina Dutilh Novaes[1]

Abstract A distinction often drawn is one between *conservative* versus *revisionary* conceptions of philosophical analysis with respect to commonsensical beliefs and intuitions. This paper offers a comparative investigation of two revisionary methods: Carnapian explication and ameliorative analysis as developed by S. Haslanger. It is argued that they have a number of common features, and in particular that they share a crucial *political* dimension: they both have the potential to serve as instrument for social reform. Indeed, they may produce improved versions of key concepts of everyday life, for example those pertaining to social categories such as gender and race (among others), which in turn may lead to social change. The systematic comparison of these two frameworks offered here, where similarities as well as differences are discussed, is likely to provide useful guidance to practitioners of both approaches, as it will highlight important aspects of each of them that tend to remain implicit and under-theorized in existing applications of these methodologies to specific questions.

In recent years, discussions on philosophical methodology have become intensified, arguably in response to the challenges posed by the surge of experimental approaches

✉ Catarina Dutilh Novaes
c.dutilh.novaes@rug.nl

1 Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

in the early 2000s.[1] In particular, debates on the goals of philosophical analysis, especially with respect to commonsensical beliefs and so-called pre-theoretical intuitions, became acutely pressing. When engaging in philosophical analysis, are we trying to faithfully capture commonsensical concepts, which thereby become clarified and systematized but otherwise remain intact? Or do we openly seek to transform and improve on these concepts?[2] The contrast can be described as that between (respectively) *conservative* versus *revisionary* (with respect to commonsensical beliefs and intuitions) conceptions of philosophical analysis. These two conceptions were in fact influential at the very birth of analytic philosophy in the early twentieth century, and can be traced back to two founding figures in this tradition: Moore for conservative analysis, Russell for revisionary (or transformative) analysis[3] (Dutilh Novaes and Geerdink 2017).

While perhaps somewhat simplistic,[4] this contrast highlights interesting similarities and connections between conceptions of philosophical analysis that might seem quite distinct at first sight. In this paper, I focus on two philosophical methods on the revisionary side of the divide, which prima facie may appear to be far apart from each other: Carnap's method of *explication*, and Haslanger's method of *ameliorative analysis*.[5] I argue that they have a number of interesting features in common, and that reflecting on these commonalities deepens our understanding of each of them. Moreover, it will also prove useful to reflect on some of the salient differences between them. The systematic comparison of these two frameworks offered here is likely to provide useful guidance to practitioners of both approaches, as it will highlight important aspects of each of them that tend to remain implicit and under-theorized in existing applications of these methodologies to specific questions.

That there are interesting connections between the two projects was noticed by Sally Haslanger herself, who writes in a footnote after introducing the ameliorative terminology in *Resisting Reality*:

> [In 'Two dogmas'] Quine distinguishes different forms of definition, the third being what he calls (drawing on Carnap) "explicative". In giving explicative definitions, "an activity to which philosophers are given, and scientists also in their more philosophical moments… the purpose is not merely to paraphrase

---

[1] Cappelen et al. (2016) provides a recent overview of these debates.

[2] Some interesting recent discussions of these issues are: Kelly (2008), Rinard (2013) and Jenkins (2014).

[3] 'Transformative' and 'revisionary' are not synonymous, though a given instance of analysis may be both transformative and revisionary. An analysis is transformative when the very conception of what the analysandum is about may change as a result of the process; it is revisionary when it leads to a change in the truth-value attributed to a particular common sense belief. In other words, the transformative component pertains to the meaning of propositions and their parts, whereas the revisionary component to their truth-value.

[4] In chapter 13 of *Resisting Reality*, Haslanger puts pressure on this distinction, while still essentially endorsing it.

[5] A third influential (semi-)revisionary method is that of *reflective equilibrium*, which however (for reasons of space) will not be dealt with extensively here. My general impression is that reflective equilibrium is the least revisionary of these three, and at times in fact more conservative than revisionary. This is also what Haslanger seems to suggest when she mentions reflective equilibrium in connection with what she calls conceptual analysis (see Sect. 2.1 below). See also (Brun 2017) for a comparison between explication and reflective equilibrium.

the definiendum into an outright synonym, but actually to improve upon the definiendum by refining or supplementing its meaning" (Quine 1963, 24–25). "Ameliorative" captures better than "explicative" the sort of project Quine is characterizing as especially philosophical […] It should be understood, however, that on my view, whether or not an analysis is an improvement on existing meanings depends on the purpose of the inquiry. (Haslanger 2012, footnote 1, p. 367)

In both cases, a philosophical analysis takes as its starting point a concept as used in ordinary, everyday contexts (or a scientific concept in earlier stages of development), and then goes on to deliver a different but closely related concept that represents an *improvement* (on a suitable metric—depending on the purpose of the inquiry, a number of different criteria may be relevant, as pointed out by Haslanger)[6] over the original concept. Carnapian explication and ameliorative analysis focus on how we *should* understand a concept, not only on how we currently do (though the latter is the starting point for the analysis). This is the core revisionary commitment that these two conceptions of philosophical analysis, Carnap's and Haslanger's, share.

If a revisionary stance were all that these two approaches shared, a comparison between the two might not be particularly illuminating. However, Carnapian explication (broadly understood) and Haslanger's ameliorative analysis also share a *political* dimension that is worth spelling out. Haslanger is very explicit on the political goals of ameliorative analysis, whereas Carnapian explication is prima facie not an overtly political project. Nevertheless, if Carnapian explication is viewed within the broader context of the Vienna Circle's political engagement (Uebel 2004; Romizi 2012) and Carnap's own commitment to the values of the Enlightenment (Carus 2008), then explication may rightly be seen as an instrument for social reform.

Both approaches belong to left-leaning political traditions, but they occupy different locations within this big tent. Indeed, there is an important difference of background between the two approaches; while Carnapian explication is embedded in Enlightenment ideals (broadly understood as associating knowledge with liberation and emancipation (Pearson 2017)) and the Vienna Circle's 'scientific world conception' (Romizi 2012), Haslanger is influenced by the critical theory tradition of the Frankfurt School (among other influences, in particular feminist thought). As is well known, the Frankfurt School authors were themselves fierce critics of the Vienna Circle's alleged 'scientism' (O'Neill and Uebel 2004), and of Enlightenment ideals more generally (Horkheimer and Adorno 1972). These different stances with respect to science in turn lead to different methodological choices: Carnapian explication often favors methods borrowed from the exact, natural sciences—formalization, mathematical frameworks—whereas critical theorists tend to focus on conceptual frameworks coming from the humanities (e.g. history) and (qualitative) social sciences (sociology, political sci-

---

[6] In his presentation of explication in *Logical Foundations of Probability*, Carnap does not emphasize much the purpose-relative nature of explication. However, in his later 'Replies and systematic expositions' in the *Living Philosophers* volume (Carnap 1963), Carnap adopts a more pragmatic perspective, stressing that different purposes and goals can be adopted for different explicative projects. See Brun (2016) for a defense of a resolutely pragmatic interpretation of the method of explication, where Carnap's 'Replies and systematic expositions' plays a prominent role.

ence). (Haslanger herself is open to contributions that any methodology may make to ameliorative projects).

But these dissimilarities are also the reason why each of the two approaches may benefit from engaging with the other. In particular, ameliorative projects may well incorporate the formal techniques that are typically deployed in instances of Carnapian explication, despite some suspicion that might still hover regarding the risk of 'scientism' (to be clear, this is not a suspicion that Haslanger herself seems to share). In turn, the comparison with ameliorative projects will suggest that Carnapian explication may also rely on tools borrowed from the humanities and social sciences, especially in view of broader social and historical contextualizations that are sometimes missing in explicative analyses (for example, by incorporating genealogical elements, as suggested by Haslanger in chapter 13 of *Resisting Reality*). Finally, the more explicit political articulation of ameliorative projects can serve as a model for explications of more politically laden concepts.[7]

I proceed in the following way. In Sect. 1, I start with a brief presentation of Carnapian explication, focusing in particular on the fruitfulness requirement for a successful explication. Closing Sect. 1, I discuss an example of a Carnapian explication of the concept of intersectionality (Bright et al. 2016). In Sect. 2, I present the outlines of Haslanger's ameliorative method as described and deployed in Haslanger (2012), including a brief discussion of how she applies the method to gender and race. In Sect. 3, I compare the two methods on two aspects: criteria of adequacy/success conditions; and with respect to an objection that can be leveled against both, namely the 'change of subject' objection: when explicating/analyzing a concept so as to improve it, what we in fact do is to introduce a new concept that bears little to no resemblance to the initial subject of analysis. I conclude with some general remarks on the significance of the present analysis for questions pertaining to philosophical methodology in general.

1 Carnapian explication

1.1 The basics[8]

Throughout his career, Carnap developed a number of constructed formal systems, as early as in *Der logische Aufbau der Welt* (1928) and *Logische Syntax der Sprache* (1934). Indeed, the notion of a *constructed language* is perhaps the main theme running through all of Carnap's large oeuvre. In his later writings, in particular in *Meaning and Necessity* (1947) and *Logical Foundations of Probability* (1950), the key methodolog-

---

[7] An anonymous referee correctly points out that the idea of borrowing tools from each other may mean that, in the limit, Carnapian explication and ameliorative analysis might in fact collapse into each other and become one and the same method. It seems to me however that important differences would still remain. For example, Carnapian explication may veer towards either ameliorative or descriptive analysis (in Haslanger's terminology). (More on what distinguishes the two methods in Sect. 3).

[8] This section draws from the more detailed exposition of Carnapian explication in Dutilh Novaes and Reck (2017).

ical concept became that of *explication*,[9] which is not exclusively tied to constructed languages and formalization,[10] but often does take the form of applications of artificially constructed languages. Explication is a process whereby a vague, informal concept, either from everyday life or from more regimented contexts such as scientific contexts (but in earlier stages of development), is given a more exact, often formalized formulation.[11] As such, the approach is a successor of some related earlier methodologies used by Frege, Russell, and others (Beaney 2013). Here is one of Carnap's earlier formulations of the idea of explication, from *Meaning and Necessity*:

> The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. We call this the task of explicating, or of giving an *explication* for, the earlier concept; this earlier concept, or sometimes the term used for it, is called the *explicandum*; and the new concept, or its term, is called an *explicatum* of the old one. (Carnap 1947, pp. 7–8)

The most worked-out exposition of the concept of explication is to be found in Chapter 1 of *Logical Foundations of Probability*. It is there that Carnap offers the now famous Fish-Piscis example to illustrate the general idea:

> When we compare the explicandum Fish with the explicatum Piscis, we see that they do not even approximately coincide […]. What was [the zoologists'] motive for […] artificially constructing the new concept Piscis far remote from any concept in the prescientific language? The reason was that [they] realized the fact that the concept Piscis promised to be much more fruitful than any concept more similar to Fish. A scientific concept is the more fruitful the more it can be brought into connection with other concepts on the basis of observed facts; in other words, the more it can be used for the formulation of laws. (Carnap 1950, p. 6).

It is thus the *fruitfulness* of the concept Piscis that motivates the adoption of Piscis in certain contexts where Fish had so far been used, in particular *scientific* contexts. But fruitfulness is not all that it takes for an explication to be successful (though it is arguably the main criterion, as will be argued below). Carnap in fact presents four main criteria of adequacy for an explication:

---

[9] There has been renewed interest in Carnapian explication in recent years, as attested by a number of recent articles such as (Maher 2007; Justus 2012; Leitgeb 2013; Brun 2016) an edited volume dedicated to explication (Wagner 2012) as well as the influential monograph by Carus (2008).

[10] "[The theorist engaging in explication] may merely state a few simple rules, or he may prefer a more or less elaborate procedure, and for this he may or may not use an artificial language." (Carnap 1963; p. 936).

[11] However, Carnap's own insistence that the gist of explication is to reduce vagueness has a number of problematic consequences, as argued in Brun (2016). Instead, Brun suggests that decreasing vagueness is merely one of the goals related to exactness, alongside goals such as giving explicit rules for using the explicatum, eliminating ambiguity, and avoiding inconsistency and paradoxes.

> A concept must fulfill the following requirements in order to be an adequate explicatum [...]: (1) similarity to the explicandum; (2) exactness; (3) fruitfulness; (4) simplicity. (Carnap 1950, p. 5)

Indeed, while there is considerable room for maneuver when engaging in explication (a theme related to Carnap's famous principle of tolerance: "In logic, there are no morals"), this does not mean that 'anything goes'. There are standards that an explication must satisfy to be deemed adequate, and some explications will be more successful than others (though it is possible that one fares better on one criterion while a different one fares better on another, for one and the same explicandum). Let us briefly look into three of the requirements, which on the list above are presented in a somewhat unsystematic order. (I leave simplicity aside, as it does not have much to add to the present discussion).

(1) Similarity is a rather weak criterion for Carnap; as we saw with the Fish-Piscis example, the two concepts "do not even approximately coincide". However, to ensure that an explicatum is an explicatum *for a given explicandum*, some degree of similarity must be in place; otherwise, the two concepts might end up being about completely different things. (This then leaves room for the 'change of topic' objection, more on which below.) Carnap resorts to the notion of *use* in order to ground the relation of similarity:

> 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, p. 7)

In the case of Fish-Piscis, the idea is that zoologists would henceforth use Piscis in most of the situations where they would have used Fish prior to the introduction of the concept Piscis.

(2) Exactness might prima facie appear to be the main requirement for a successful explanation; indeed, Carnap says that the explicatum being more exact than the explicandum is essential for an explication.[12] Critics of Carnapian explication (e.g. Strawson—see Sect. 3.2 below) often object to what they perceive as an excessive preoccupation with criteria and methods imported from the exact sciences that are (arguably) out of place in philosophical contexts—in other words, they object to the perceived scientism of the general approach espoused by Carnap and other members of the Vienna Circle.

The scientism charge has been amply discussed in recent reappraisals of the Vienna Circle movement, and has been shown to be at the very least questionable (Wartofsky 1982; Uebel 2004; Romizi 2012). Importantly, within the specific interpretation of Enlightenment ideals by some members of the Vienna Circle, in particular Neurath and Carnap (Carus 2008) (more on which shortly), formalism and mathematization were in fact viewed as means for intellectual emancipation, "an instrument against obscurantism" (Wartofsky 1982, p. 89). Setting aside the criticism of Enlightenment

---

[12] "The only essential requirement is that the explicatum be more precise than the explicandum; it is unimportant to which part of the language it belongs." (Carnap 1963, p. 935).

ideals that have been put forward by a wide range of authors (with varying degrees of success), what matters for us here is the idea that, for Carnap, exactness is not an end *an sich*; it is an *instrument*, a means to an end, or to a variety of specific ends as the case may be. This brings us directly to the third, arguably most important requirement for a Carnapian explication: it must be fruitful.

(3) Fruitfulness is the crucial requirement for a successful explication (as argued in Carus 2008; Dutilh Novaes and Reck 2017), and yet Carnap says surprisingly little on what he means by fruitfulness. Here is one of his few relevant remarks:

> The explicatum is to be a fruitful concept […] 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). (Carnap 1950, p. 7)

The idea of an explicatum leading to many universal statements may be understood in terms of predictive power and testability: if it allows for the formulation of universal statements, then it will generate predictions that can be tested. In the case of non-empirical theories, allowing for the derivation of many theorems might amount to the general idea of the fruitfulness of concepts and definitions [as discussed by Frege (Tappenden 1995)]. Moreover, as suggested in the passage quoted above on the Fish-Piscis example, an explicatum's fruitfulness is related to its ability to connect with other concepts on the basis of observed facts.[13]

However, an account of the fruitfulness of an explicatum solely in terms of its capacity to generate universal statements and predictions, and to be connected to other observed facts, remains somewhat meager. Interpreters of Carnap have taken upon themselves to flesh out what fruitfulness might mean beyond Carnap's own inchoate remarks. Carus (2008) developed a resolutely pragmatic interpretation of fruitfulness; elsewhere (Dutilh Novaes and Reck 2017) I have offered an account of fruitfulness in terms of the cognitive boost afforded by explication and formalization. In the next section, I further argue that fruitfulness may also be understood in *political* terms, i.e. in terms of promoting much-needed social reform and liberation from obscurantism (thus agreeing with the main lines of Carus' interpretation).

1.2 Fruifulness and political engagement

In Carnap's own writings, little textual evidence can be found to ground a fully-fledged political articulation of the fruitfulness criterion for explication specifically. Formulations of more explicit political commitments, and in particular of the idea that philosophy was of critical relevance to the project of socio-historical changes towards socialism, are mainly to be found in Carnap's earlier writings, as one of the members of the 'Left Vienna Circle' (Uebel 2004).[14] The concept of explication, in turn, was to be fully articulated only in his later writings. But granting at least some degree of

---

[13] Later on, in Carnap (1963, p. 936), Carnap suggests that a number of different aims and purposes may be well served by explication, presumably beyond these two that are explicitly mentioned in Carnap (1950).

[14] See Bright (2017) for an analysis of the progressive stance on race of some of the Vienna Circle members, including Carnap.

methodological continuity between Carnap's earlier work and his later work (given the centrality of formal systems and constructed languages throughout his career), as well as continuity in his political commitments, [15] a case can be made for the relevance of these political commitments also with respect to his later work, and explication in particular.

For Carnap, explication embodies the value of intellectual liberation from the shackles of paralyzing traditions. A passage by one of his students, Richard Jeffrey, illustrates this connection particularly well:

> Philosophically, Carnap was a social democrat; his ideals were those of the enlightenment. His persistent, central idea was: "It's high time we took charge of our own mental lives"—time to engineer our own conceptual scheme (language, theories) as best we can to serve our own purposes; time to take it back from tradition, time to dismiss Descartes's God as a distracting myth, time to accept the fact that there's nobody out there but us, *to choose our purposes and concepts to serve those purposes*, if indeed we are to choose those things and not simply suffer them. […] For Carnap, deliberate choice of the syntax and semantics of our language was more than a possibility—it was a duty we owe ourselves as a corollary of freedom. (Jeffrey 1992, p. 28)[16]

In other words, explication exemplifies the ideal of *epistemic autonomy*, most famously captured by Kant's Enlightenment motto: 'Sapere Aude!'[17]

The general political background for Carnap's thought has been amply documented and discussed by a number of authors (Carus 2008; Romizi 2012; Uebel 2004), so we need not rehearse the details here. So far however, while setting up the stage, these brief considerations remain much too general to ground an interpretation specifically of Carnap's criterion of fruitfulness as related to political engagement. Fortunately, there is at least one fairly clear connection between his brief remarks on fruitfulness mentioned above and political ideals, through the concept of *prediction*. Recall that he emphasized the role of explication as leading to the formulation of universal statements, which in turn allow for testable predictions to be made regarding the relevant subject matter. One might think that prediction thus understood is not a particularly politicized concept; the idea of predictions which can be tested and thus can refute or corroborate theories is a core component of classical conceptions of science. But predictions also allow for *interventions* to be brought about, thus bringing us closer to political ideals of social reform. The political dimension of predictions is well described in the following passage by Wartofsky, commenting on Neurath:

---

[15] In his autobiography in the *Living Philosophers* volume, Carnap explicitly noted that he continued to subscribe to the political views first articulated in Vienna. Moreover, he remained engaged in political activism after moving to the USA (Reisch 2005, pp. 47–53).

[16] I owe this reference to L.K. Bright.

[17] However, it is worth noting that Carnap's version of the Enlightenment ideals is more self-reflective than a number of other versions, in particular because he was an astute reader of Nietzsche (Sachs 2011). Indeed, he seemed to recognize the potential for authoritarianism at the heart of the Enlightenment ideals. For example, his rejection of the idea of morals in logic embodied in the principle of tolerance can be read as an upshot of Nietzsche's critique of morality in general (Mormann 2012, p. 74).

Further, one of the hallmarks of the scientific conception of the world, and certainly central to Neurath's view, was the importance of prediction in the sciences. The social value of prediction was seen by Neurath as a means of applying rational controls over hypothetical plans of action—whether in research in the sciences, involving the use of experiment, or in social planning (or in Neurath's own term, "social engineering"). (Wartofsky 1982, p. 89)

(Wartofsky then goes on to comment on how this project of 'social engineering' came to be seen as inherently authoritarian and oppressive by critics such as the Frankfurt School philosophers.) Given that on these points Carnap and Neurath were very much in agreement, it is not too much of a stretch to connect the idea of the fruitfulness of an explication as generating universal statements to the social value of prediction within projects of social engineering.[18]

At this point, it might seem that a fruitful explication may well have the potential to serve the purposes of social engineering, but that this potential is rarely or perhaps never truly realized. To alleviate this worry, I now discuss a concrete example of an explication of a socially relevant concept, which leads to precise predictions and thus can serve as an instrument for social intervention and social reform: the explication of the concept of *intersectionality* offered in Bright et al. (2016).

Bright et al. indeed explicitly present their analysis as an instance of explication. The explicandum, in this case, is the concept of intersectionality, which is a theoretical concept but might be said to be "in an earlier stage of scientific or logical development" (a phrase from the passage from *Meaning and Necessity* quoted above). It was introduced in the 1990s by legal scholar K. Crenshaw to capture the idea that different dimensions of oppression, targeting different disadvantaged groups, may interact (intersect) in non-trivial ways. For example, the challenges encountered by black women in the United States are different from the challenges encountered by either black men or white women, because gender and race intersect in ways that give rise to different kinds of challenges. But while intersectionality thus described is a powerful and useful concept, it has also been the target of fierce criticism, in particular with respect to its ability to generate predictions about the phenomena in question. (As described by Bright et al., some proponents in fact reject the idea that the concept of intersectionality should lead to testable predictions at all.)

Bright et al. thus give the concept of intersectionality a causal interpretation in terms of the formalism of graphical causal modeling, which in the last decades has become an influential approach in a number of areas (Spirtes et al. 1993; Pearl 2000). As seen in Fig. 1 (taken from Bright et al. 2016, p. 67), directed acyclic graphs (DAGs) allow for the representation of causal relations between different phenomena. In this graph, Parental Income and Education are represented both as direct causes of Wealth, and additionally, Parental Income is an indirect cause of Wealth via its influence on Education. In turn, Education is independent of Gender given Race and Parental Income (no arrow between these two nods), which roughly means that, in the contexts that the graph represents/describes, boys and girls have the same access

---

[18] Carnap himself also wrote on the role of prediction in informing practical decisions in Carnap (1934) (I owe this reference to G. Brun.).
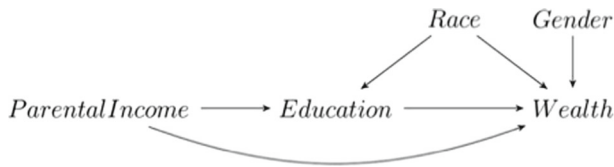
Fig. 1 Directed acyclic graph (DAG)

to education holding these two parameters fixed (say, within the same family).[19] But Gender does affect Wealth, as typically men are wealthier than women even holding the other parameters fixed.

Equipped with this formalized causal account of intersectionality, researchers are better prepared to understand the underlying causal relations between different aspects of the phenomena in question (i.e. being a member of different social groups and being thereby privileged or disadvantaged in certain respects), make relevant predictions, and potentially formulate interventions that may address the social inequalities in question. In Fig. 1, for example, the fact that Education is independent of Gender given Race and Parental Income suggests that the kind of intervention needed to bridge the pay gap between men and women that is still observed even in so-called modern, industrialized societies is not that of investing in early education for girls. Instead, the phenomenon seems to be caused by factors within the dynamics of the labor market itself and the distribution of domestic labor at home. These observations indicate the kind of intervention needed to address this inequality, such as: policies for gender pay equity in the workspace; policies to increase male involvement in childrearing, such as paternal leave.[20]

And thus, the explication of intersectionality in terms of graphical causal modeling by Bright et al. is an excellent example of the potential for social and political relevance of the explication methodology, and indeed an illustration of the links between predictions and potential for social engineering. In effect, this explication scores high on the fruitfulness criterion as described by Carnap because it leads to a number of general statements which can then both serve to test the empirical adequacy of the framework, and to guide targeted social interventions at a later stage.[21] While this is but one example of how Carnapian explication, including its focus on formalization, can be enlisted for politically and socially relevant projects, there is no reason to think that this should be an isolated case. Concrete uses of explication for such projects may

---

[19] This is of course not true of many regions in the world, where girls systematically receive fewer educational opportunities than boys. This is the case though in many industrialized countries in Western Europe and North America, which is the kind of situation represented by this graph.

[20] The formal apparatus in Bright et al. (2016) is considerably more complex and sophisticated than the example from Fig. 1 alone suggests, but the example should at least give the reader an idea of how formalism and social interventions can be related.

[21] An important debate in the intersectionality literature is whether the concept should be applied at the group level or the individual level (Collins 2003). Bright et al. make a case for the group level perspective with their causal analysis, which in turn allows for more general statements to be generated.

still be somewhat rare, but this simply means that there is much potential for future work to be done in this direction.[22]

2 Haslanger's ameliorative analysis

2.1 The ameliorative method

Sally Haslanger has been deploying the ameliorative method in her work on race and gender since the late 1990s (Haslanger 1999, 2000), but the term 'ameliorative' is not used in her earlier work. To my knowledge, its first occurrence is in (Haslanger 2006), and its most worked-out exposition is to be found in her 2012 book *Resisting Reality* (Haslanger 2012), which will be the main source for the present discussion.[23] Haslanger summarizes the gist of her approach in the following way:

> In developing constructionist accounts of race and gender, I've maintained that my goal is not to capture the ordinary meanings of 'race' or 'man' or 'woman', nor is it to capture our ordinary race and gender concepts. 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. (p. 366)

Haslanger thus clearly sides with what was described above as a revisionary/transformative conception of philosophical analysis—in fact she uses the term 'revisionary' herself. These revisionary commitments are well served by the constructionist approach, given the potential for interventions on concepts of kinds such as race and gender once they are understood as social constructions rather than as pertaining to immutable, biological essences. In particular, she explicitly discusses the idea of 'debunking' intuitions about social kinds through ameliorative analysis—the result of which will be unsurprisingly counterintuitive, since this is precisely the goal of the analysis.

> In debates over the existence and nature of social kinds such as 'race' and 'gender', philosophers often rely heavily on our intuitions about the nature of the kind. Following this strategy, philosophers often reject social constructionist analyses, suggesting that they change rather than capture the meaning of the kind terms.

---

[22] Another recent example is (Pearson 2017) on the concept of marriage. Pearson also offers an illuminating critical discussion of Carus' interpretation of Carnap's ideal of explication as a model for liberal political deliberation.

[23] The book comprises re-worked versions of a number of her earlier articles, and for this reason we will focus on the book exclusively here, since it contains all the relevant material. (All page references are to the book.) Moreover, since the aim of the book is to make each chapter self-contained, there is some overlap between the chapters. In particular, chapters 13 and 14 both contain extensive discussions of the ameliorative method. Indeed, Haslanger explicitly addresses methodological questions at length, and thus it is often sufficient to let her speak for herself without much need for further textual interpretation.

> However, given that social constructionists are often trying to debunk our ordinary (and ideology-ridden?) understandings of social kinds, it is not surprising that their analyses are counterintuitive. (p. 381)

Ameliorative projects are contrasted with two other ways of engaging in philosophical analysis: conceptual and descriptive approaches.

> For example, consider the question 'What is knowledge?' Following a *conceptual* approach, one is asking: What is *our* concept of knowledge? and looking to a priori methods such as introspection for an answer. Taking into account intuitions about cases and principles, one hopes eventually to reach a reflective equilibrium. On a *descriptive* approach, one is concerned with what objective types (if any) our epistemic vocabulary tracks. The task is to develop potentially more accurate concepts through careful consideration of the phenomena, usually relying on empirical or quasi-empirical methods. (p. 367)

The conceptual approach thus described is clearly conservative, in the sense introduced at the beginning of the paper: pre-theoretical intuitions can be systematized and scrutinized, but never outright rejected or substantially modified. The descriptive approach is prima facie neutral with respect to common sense and intuitions; it may well be that "careful consideration of the phenomena" will confirm commonsensical beliefs about the relevant phenomena, but often empirical or quasi-empirical investigation will lead to revisions of these beliefs. Ameliorative projects, however, are from the beginning motivated by practical purposes, and thus start off with revisionary proclivities.

> Ameliorative projects, in contrast, begin by asking: What is the point of having the concept in question—for example, why do we have a concept of knowledge or a concept of belief? What concept (if any) would do the work best? In the limit case, a theoretical concept is introduced by stipulating the meaning of a new term, and its content is determined entirely by the role it plays in the theory. If we allow that our everyday vocabularies serve both cognitive and practical purposes that might be well-served by our theorizing, then those pursuing an ameliorative approach might reasonably represent themselves as providing an account of our concept—or perhaps the concept we are reaching for—by enhancing our conceptual resources to serve our (critically examined) purposes. (p. 367/8)

Ameliorative projects thus start by asking functionalist/pragmatic questions: given a certain concept, what function or functions does it have in the constellation of our practices and discourses?[24] Once this question is at least partially answered, we can ascertain whether the version of the concept that we currently entertain is adequate for the job(s) we expect it to do. If not, then a revisionary, ameliorative process can be set in motion, hopefully leading to a more suitable version of the concept in question, given the functions/goals that have been established in the first step of the investigation.

---

[24] It is interesting to notice that this functionalist perspective is shared with a number of genealogical projects such as (Williams 2002) and (Craig 2007). For more on genealogical analysis, see (Dutilh Novaes 2015). Haslanger herself develops the concept of 'ameliorative genealogy': "An ameliorative genealogy that undertakes to evaluate the point of having a concept or structure of concepts (along with related practices) and proposes improved resources to fulfill them." (p. 372).

Importantly, in the limit case this can be achieved by *stipulation* and the introduction of a new term. Recall that in Carnap's Fish-Piscis example, zoologists decided at some point to use Piscis instead of Fish because the category Piscis was better suited for scientific investigation than the category Fish, which was based on superficial morphological resemblances between animals who are otherwise very different (in terms of evolutionary lineage, DNA etc.).[25] This seems to be the kind of stipulation that Haslanger is alluding to.[26]

There is another aspect in which the comparison with Carnap may be instructive, but this time it is a point of disanalogy between the two authors. Carnap (1950) does not explicitly say that different purposes and goals may be addressed in an explication; the pragmatic component is simply alluded to with the umbrella term 'fruitfulness'. He seems to imply that fruitfulness is an absolute concept rather than one relative to different purposes and goals one may have. [Later, in Carnap (1963), he is more explicit on the plurality of goals and purposes that may be addressed by an explication. Perhaps tellingly, he no longer speaks of fruitfulness in this later text.] In contrast, Haslanger very clearly states that ameliorative analysis is a two-tiered process that starts with a thorough, critical examination of one's purposes, and then proceeds to reformulate a concept in view of the purposes previously established.[27]

> Ameliorative analyses elucidate "our" legitimate purposes and what concept of F-ness (if any) would serve them best (the target concept). Normative input is needed. (p. 376)

2.2 The method in practice: gender and race

As mentioned above, it is primarily within her work on race and gender that Haslanger has been developing and deploying the ameliorative method, and thus it is useful for our purposes here to observe the method in action (albeit briefly), just as we had an example of Carnapian explication with intersectionality. The first step in the ameliorative method, as described above, is a critical examination of the purposes to be achieved. Rather than seeking to systematize commonsensical notions of race and gender (conceptual analysis), or to investigate (purported) deeper metaphysical and empirical properties of the phenomena in question (descriptive analysis), and in keeping with the spirit of critical theory broadly understood (Bohman 2005), Haslanger's (critically examined) chief purpose when providing an analysis of race and gender is that of fighting inequality, domination, and oppression in social structures, thus

---

[25] See also Carnap (1963, p. 936) on the issue of introducing a new term for the newly formed concept (explicatum), or sticking with the 'old' term that was used in connection with the explicandum.

[26] Notice though that the Fish-Piscis example, although revisionary, is not necessarily ameliorative. Instead, it might be a case of what Haslanger calls descriptive analysis (more on descriptive instances of Carnapian explication below). (I owe this point to an anonymous referee.)

[27] This may suggest a strictly linear conception of the process, whereas a more plausible view would be that there may well be feedback loops between these two components of the process. Brun (2016, 2017) argues for a non-linear conception of Carnapian explication along these lines, and it seems that ameliorative analysis could well be conceived of in a non-linear way as well.

promoting social justice. One of the four main concerns (perhaps even the main one) guiding the project is formulated in the following terms:

> The need to identify and explain persistent inequalities between females and males, and between people of different "colors"; this includes the concern to identify how social forces, often under the guise of biological forces, work to perpetuate such inequalities. (p. 226/7)
>
> More specifically, the goal of the project is to consider what work the concepts of gender and race might do for us in a critical—specifically feminist and antiracist—social theory… (p. 226)

Another aspect where Haslanger borrows from the critical theory tradition is the idea of *ideology critique*. As stated in a passage quoted above, her starting point is that our ordinary understandings of social kinds tend to be ideology-ridden, in such a way that they in fact perpetuate inequalities and hierarchies among men and women, white people and people of color. This is another reason why the approach required, given her stated purposes, cannot be either conceptual analysis, which leaves these ordinary understandings essentially intact, or descriptive analysis, which purportedly addresses the phenomena directly without critically examining the ideological components of our ordinary understandings of social kinds. Ameliorative (revisionary) analysis is what is required.

With the relevant political/pragmatic goals in place, Haslanger then goes on to investigate what a suitable analysis of gender and race might be, one that would be instrumental in promoting these goals. To this end, she draws from materialist feminist accounts of gender, which view gender categories in terms of women's subordinate position in systems of male dominance—that is, a social and indeed political understanding of gender, as opposed to the folk-biological notion of sex/gender which places the relevant differences at the level of biological reproductive functions. A similar approach is adopted for race, which is then to be conceived of in terms of dominance by white people, and the corresponding subordination of people of color. Tellingly, in chapter 14 of *Resisting Reality*, Haslanger criticizes error-theoretical approaches such as the one defended by Appiah with respect to race, according to which "there are no races: there is nothing in the world that can do all we ask race to do for us" (as cited by Haslanger). On her view, Appiah's conceptual analysis leading to the conclusion that there is nothing that satisfies all the different properties that we associate with race (thus being an eliminativist proposal) fails to have the desirable effect of promoting social justice, and thus fails the political goals described above.

Haslanger thus arrives at the following definitions of woman and man:

> S *is a woman* iff$_{df}$ 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. (p. 230)
>
> S *is a man* iff$_{df}$ S is systematically privileged 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 male's biological role in reproduction. (p. 230)

By emphasizing the subordination of women by men, and by identifying (observed or imagined) bodily features related to biological reproduction as the purported but mistaken locus for gender differences, these definitions, precisely because they are revisionary, are suitable to serve the political purposes that motivated the analysis in the first place. (She later offers more precise definitions that accommodate certain objections, but for our purposes we need not get into these details.)

Race is then defined in similar terms:

> A group is *racialized* iff$_{df}$ 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. (p. 236)

Prima facie, these definitions seem to satisfy the political desideratum of promoting values of social justice. It has been argued however that Haslanger's definition of woman fails to promote social justice with respect to (some) trans women (Jenkins 2015). Jenkins argues that some trans women do not count as women according to Haslanger's definition; in other words, the definition undergenerates with respect to individuals that we do want to count as women on the basis of other considerations (in particular self-identification). And since trans women are among the most marginalized and oppressed groups in many societies, not counting some of them as women entails failing the political goal of promoting social justice. (Whether something similar might be said of Haslanger's definition of race, including the analogies and disanalogies between transgender people and the phenomenon of race fluidity, is an important question which will however not be addressed here.)

The objection by Jenkins highlights an important element of the framework, namely that ameliorative analysis is best seen as an open-ended project; proposed reconceptualizations may then be further revised if it turns out that additional improvement can be made. This is a feature that ameliorative analysis shares with Carnapian explication (Carus 2008; Dutilh Novaes and Reck 2017), which is also an open-ended process that can be iterated: the explicatum (the result) of one explication can then become the explicandum (the starting point) for a different explication.

3 Comparison

With these brief but, I hope, sufficiently instructive presentations of Carnapian explication and Haslanger's ameliorative analysis in place, we can now engage in a more systematic comparison between the two. Two specific aspects will be singled out: what counts as adequacy criteria and success conditions for each of the two approaches; and an important objection that can be (and has been) leveled against both approaches, namely the 'change of subject' objection.

3.1 Criteria of adequacy/success conditions

In Sect. 1.1 I briefly discussed the criteria of adequacy for an explication as stated by Carnap. The main ones are: similarity, exactness, and fruitfulness. (Simplicity is rather a general theoretical virtue.) For convenience, let me re-state the similarity criterion:

> 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; p. 7).

As for exactness, as discussed above, it is of course closely related to the idea (characteristic of the Vienna Circle) that the methods in the exact and empirical sciences should be emulated elsewhere, in philosophical analysis in particular.[28] But importantly, the left Vienna Circle philosophers such as Carnap and Neurath in fact viewed formalization and exactness through political lenses—as embodying the Enlightenment ideals of intellectual emancipation, not (as later critics would claim) as an objectionable form of scientism. Finally, fruitfulness remains rather underdeveloped in Carnap's own words, but above I've offered an interpretation of fruitfulness as also related to political ideals of social reform through the notions of prediction and social engineering.

Haslanger also addresses the issue of success conditions (criteria of adequacy) for ameliorative analyses. In particular, she distinguishes two kinds of conditions: semantic conditions and political conditions.

> Are there principles that determine when it is legitimate to appropriate the terms of ordinary discourse for theoretical purposes?[29] An answer, it seems to me, should include both a semantic and a political condition (though in some cases the politics of the appropriation will be uncontroversial). The semantic condition is not surprising: 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 organize or explain a core set of phenomena that the ordinary terms are used to identify and describe. Framing a political condition in general terms is much more difficult, however, for the politics of such appropriation will depend on the acceptability of the goals being served, the intended and unintended effects of the change, the politics of the speech context, and whatever the underlying values are justified. (p. 225)

---

[28] However, Carnap later explicitly denies that such methods should be mandatory for philosophical analysis: "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." (Carnap 1963; p. 935).

[29] This is often the gist of ameliorative analyses, i.e. to introduce new, improved meanings for terms in ordinary discourse, though at times an ameliorative analysis may introduce a new term altogether to replace a term until then used in ordinary discourse (just as in Carnap's Fish-Piscis example). There are in fact four main options: "In many such cases, we face two questions: what policy do we want to promote (or what objective type do we want to track), and what do we want to do with the bit of language we have been using? Do we want to change our policy and keep the same term, change it and introduce a new term, keep the policy and change the term, or keep the policy with the old term?" (Haslanger 2012, p. 379) (More on this issue in the next section.).

It is quite natural to view Haslanger's semantic condition, based on the idea of the central functions of a term remaining the same, as a counterpart of Carnap's similarity criterion, which is also based on linguistic use. The political condition, in turn, can be viewed as a counterpart of Carnap's fruitfulness criterion, at least under the political interpretation of fruitfulness that I've articulated in Sect. 1.2. Indeed, it is interesting to notice that Haslanger recognizes the difficulties with formulating the political condition in general terms, given significant contextual variations and the varying acceptability of the goals being served. We may hypothesize that it was for similar reasons that Carnap could not go much beyond a rather vague and underdeveloped characterization of the notion of fruitfulness—precisely because framing it in general terms proves to be very challenging, given the multiplicity of uses and contexts for the method.

Given that framing the political condition in general terms is most likely not possible (in view of the inevitable situatedness of the theorist and her goals), Haslanger focuses on the political dimension of ameliorative analysis for individual cases. Indeed, we have seen that critically examining goals and purposes, which thereby become more explicitly formulated, is the first stage of analysis within the ameliorative method. In this respect, practitioners of Carnapian explication have something to gain from emulating the careful discussion of goals and functions that is to be found in applications of the ameliorative method, rather than adopting an absolute sense of fruitfulness. (Carnap himself seems to have moved in this direction in his later work, as attested by Carnap 1963—see footnote 6 and Brun 2016).

Obviously, there is a third criterion of adequacy on Carnap's list that has no immediate counterpart in Haslanger's framework: the exactness criterion.[30] This is indeed a significant difference between the two frameworks: while Carnapian explication is in theory not exclusively tied to formalization and mathematization, in practice the most conspicuous examples of explication tend to be instances of a formal language or formal framework being applied to a particular informal question or concept. In contrast, the ameliorative approach does not place any special emphasis on formalization, though there is no a priori reason why the former would be inimical to the latter. This is, in turn, where practitioners of the ameliorative method may benefit from engagement with the techniques employed by Carnapian explicators, especially if exactness is not understood merely as decreasing vagueness, as proposed in Brun (2016).

However, even if explicators incorporate more sustained attention to political dimensions from the ameliorative method, and ameliorativists begin to use formal tools systematically, it does not mean that the two methods will become identical. In particular, while the ameliorative method as described by Haslanger is always tied to political goals, Carnapian explication can be but need not be exclusively used for social engineering. Indeed, explication continues to be a fruitful methodology for the analysis of scientifically relevant concepts, whereas ameliorative analysis is primarily

---

[30] Obviously, this does not mean that Haslanger does not value exactness and precision in philosophical analysis (after all, she was the recipient of top-notch training as an analytic philosopher!). The point is rather that formalization and mathematization are not (until now at least) among the tools typically employed within ameliorative projects.

geared towards socially and politically charged concepts. In particular, the negative component of *ideology critique*, which is crucial in the ameliorative method, does not seem to have a natural counterpart in Carnapian explication, at least not in its classical formulation.

Moreover, some instances of Carnapian explication may in fact belong to Haslanger's category of *descriptive* analysis: they may be revisionary, but not necessarily ameliorative (such as the Fish-Piscis example, which seems to lead to a descriptively (scientifically) more accurate categorization of the organisms in question). Indeed, while the ameliorative method is essentially *functional*—the goal is to engineer concepts that will better perform certain (social) functions—certain instances of Carnapian explication will not necessarily be functional in the sense that the ameliorative method is; some will be largely descriptive.[31] However, if the scientific goal of leading to predictions and general statements, or more generally to better explanations of the phenomena in question, counts as a relevant 'function', then Carnapian explication can be seen as essentially functional as well, at least in a weaker sense (based on a loosely instrumentalist, anti-realist conception of scientific theories).

And thus, the conclusion seems to be that Carnapian explication and the ameliorative method, even when borrowing from each other, will remain overlapping but not identical methodologies, both extensionally (in terms of the questions and concepts to which they primarily apply) and intensionally (in terms of their criteria of adequacy, theoretical tools, and goals). For example, the analysis of intersectionality in Bright et al. (2016) is clearly an instance of explication, but arguably *not* an instance of ameliorative analysis, despite having the potential to lead to social interventions.[32]

3.2 The 'change of subject' objection

One of the main criticisms that have been voiced against Carnapian explication is the 'change of subject' objection, perhaps most famously formulated by Strawson in his piece for the Schilpp *Living Philosophers* volume dedicated to Carnap. At first sight, Strawson's critique appears to be a variation of the scientism objection:

> For however much or little the constructionist technique [explication] is the right means of getting an idea into shape for use in the formal or empirical sciences, it seems prima facie evident that 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

---

[31] I owe the distinction between functional versus descriptive analysis to an anonymous referee. However, given Carnap's own rejection of heavy metaphysical commitments, 'descriptive' here should not be understood in terms of describing freestanding essences in reality or such like.

[32] In turn, I suspect that many instances of ameliorative analysis can, with some flexibility at least, be considered as instances of explication in a loose sense, especially if they involve, as is typically the case, an increase in precision with respect to the initial concept in question. If this is true, then perhaps ameliorative analyses become instances of explication, that is if explication is extended in the ways suggested here, e.g. by integrating more sustained discussion on the different needs an explication may address. However, the component of ideology critique that is often at the core of ameliorative analysis still distinguishes it from typical instances of explication.

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; pp. 504–505)

Strawson's objection thus amounts to reclaiming a specific domain of inquiry—non-scientific discourse and concepts of daily life—as not amenable to this 'scientific' methodology.[33] But it can also be read as a rejection of the possibility of philosophical clarification having a transformative or revisionary import: philosophical clarification ought to be purely descriptive of beliefs and concepts underlying daily life, rather than producing new concepts which should replace the old ones, or revising entrenched commonsensical beliefs.

> 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; that the language of science could not in this way *supplant* the language of the drawing-room, the kitchen, the law courts and the novel. (Strawson 1963, p. 505)

Strawson's rejection of revisionary philosophical analysis such as Carnapian explication is further articulated in the following passage:

> [I]n most cases, either the operation [scientific language replacing non-scientific language for non-scientific purposes] 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.* More of the types of linguistic activity in which we constantly engage would succumb to such an attempt than would survive it… (Strawson 1963; pp. 505–506, emphasis added)

Carnap's response to this criticism in Carnap (1963) is to emphasize the commonalities between the approach of 'naturalists' such as Strawson and his own as a 'constructionist', but also to isolate the main points of disagreement. One of them is what could be described as Strawson's conservative stance towards ordinary language, which Carnap rejects:

> For the naturalists, ordinary language seems to have an essentially fixed character and therefore to be basically indispensable, just like our body with its organs, to which we may add accessories like eyeglasses, hearing aides, and the like, but which we cannot essentially change or replace. However, a natural language is not an unchangeable function of our body, but something we have learned; therefore we can replace it by another language. (Carnap 1963; p. 937)

Another reason to reject an overly deferential attitude towards ordinary language is that its users are themselves often confused with respect to its expressions:

> Strawson believes that philosophical problems are raised by people "who know very well how to use the expressions concerned". I should rather say that these

---

[33] Carnap's sensible and to my mind convincing reply is that there is no such strict separation between scientific and ordinary discourse (Carnap 1963).

people usually believe they know this very well, but often deceive themselves. The first step in helping these people consists in leading them to the insight that something is wrong with their use of certain expressions, that it involves confusions or even inconsistencies. (Carnap 1963; p. 934)

The change of subject objection is similarly a challenge for the ameliorative method. Haslanger engages head-on with it, further developing some of the main points of Carnap's response to Strawson (though probably with no direct inspiration drawn from Carnap). She formulates the objection in the following terms:

[W]e come to inquiry with a conceptual repertoire in terms of which we frame our questions and search for answers: hence, the subject matter of any inquiry would seem to be set from the start. In asking what *race* is, or what *gender* is, our initial questions are expressed in everyday vocabularies of race and gender, so how can we meaningfully answer these questions without owing obedience to the everyday concepts? Or at least to our everyday usage? Revisionary projects are in danger of providing answers to questions that weren't being asked. (pp. 224–225)

[I]n an explicitly revisionary project, it is not at all clear when we are warranted in appropriating existing terminology. Given the difficulty of determining what "our" concept is, it isn't entirely clear when a project crosses over from being explicative to revisionary, or when it is no longer even revisionary *but simply changes the subject*. If our goal is to offer an analysis of "our" concept of X, then the line between what's explication and what's not matters.[34] But if our goal is to identify a concept that serves our broader purposes, then the question of terminology is primarily a pragmatic and sometimes a political one: should we employ the terms of ordinary discourse to refer to our theoretical categories, or instead make up new terms? The issue of terminological appropriation is especially important, and especially sensitive, when the terms in question designate categories of social identity such as 'race' and 'gender.' (p. 225, emphasis added)

Haslanger's full response to the change of subject objection is much too complex to be presented in detail here, but the gist of it can be thus summarized:

[O]rdinary concepts are notoriously vague; individual conceptions and linguistic usage varies widely. […] [P]recisely because our ordinary concepts are vague (or it is vague which concept we are expressing by our everyday use of terms), there is room to stretch, shrink, or refigure what exactly we are talking about in new and sometimes unexpected directions. (p. 225)

In other words, and in line with Carnap's reply to Strawson, the point is that we users of ordinary languages often do not know exactly what it is that we are talking about when using everyday life vocabulary. At the very least, there is considerable leeway and flexibility in usage. This is of course a familiar point (think of Wittgenstein's notion of family resemblance), almost a platitude perhaps. But it does suggest a suitable reply

---

[34] Notice here another possible difference between explication and ameliorative analysis: explication is potentially more vulnerable to the change of subject objection.

to the change of subject objection: more often than not, it is not clear that there is a neatly circumscribed 'subject' for the everyday use(s) of a term in the first place, which philosophical analysis should be deferential to.

Haslanger further develops an account of vagueness and ambiguity in everyday language, incorporating some key topics in philosophy of language such as semantic externalism and meaning holism in her discussion. In this context, she introduces the helpful distinction between *manifest and operative concepts* in everyday usage. (A third category, target concepts, corresponds to the concepts we should be using.[35]) "Roughly, the manifest concept is the more explicit, public, and "intuitive" one; the operative concept is the more implicit, hidden, and yet practiced one."[36] (p. 370) Very often, there is a mismatch between manifest and operative concepts with respect to one and the same 'subject'. Haslanger's example is the concept of 'parent', which in the context of communication between a school and the relevant group of adults responsible for its pupils may in fact be taken to encompass all primary caregivers, regardless of whether they satisfy the biological or legal definition of 'parent'. In this case, the manifest concept of parent does not encompass e.g. grandparents who are primary caregivers, but the operative concept does, and it is the operative concept that takes the upper hand.[37]

Whenever there is a mismatch between manifest, operative, and target concepts, a number of options are available with respect to practices, content of a concept, and linguistic usage: change the practices and keep the manifest concept, or replace it with a newly formulated target concept; change the manifest concept so as to fit the practices; introduce a new term for a new concept, in order to dispel ambiguity in established usage; change the meaning of the term already in use. As discussed in chapter 14 of *Resisting Reality*, there is no one-fits-all answer to the question of how to proceed in cases of mismatch. But the main point is that the manifest concept, which is usually the one that is referred to when philosophers such as Strawson and others speak of 'concepts in everyday language', does not have obvious primacy over the other two. And thus, it is not in any way obvious that not giving a faithful account of a manifest concept by means of philosophical analysis will necessarily constitute an objectionable 'change of subject'. Haslanger thus offers subtle answers to the change

---

[35] "Let's call the *target concept* the concept that, all things considered (my purposes, the facts, etc.), I should be employing." (p. 388).

[36] "The manifest concept is the concept I take myself to be applying or attempting to apply in the cases in question. The operative concept is the concept that best captures the distinction as I draw it in practice. And the target concept is the concept I should, ideally, be employing." (p. 389).

[37] A mismatch between manifest and operative concepts may also suggest problematic practices hiding under apparently innocent manifest concepts, and it is part of the enterprise of ideology critique to expose such problematic practices. Haslanger refers to the work of C. McKinnon on gender, and of C. Mills on the Enlightenment social contract as a racial contract. "Such analyses purport to show that our manifest understandings of crucial political notions are masking how the concepts in question actually operate." (p. 390).

of subject objection, which might also be of use to Carnapian explicators who are confronted with similar objections.[38]

4 Conclusions

In this paper, I presented a comparative analysis of two philosophical methods: Carnapian explication and ameliorative analysis, as introduced by Haslanger. I have argued that there is much similarity between them, but also that there are some significant differences in how each is presented and practiced. More generally, the paper can also be seen as a defense of revisionary philosophical analysis, as I have attempted to further articulate two such revisionary methods based on the writings of their main architects (Carnap and Haslanger), and to defend them against objections. As such, I hope to have made a more general contribution to debates on philosophical methodology, especially with respect to the increasing popularity of projects of conceptual engineering; on my story, such projects are not only philosophically viable and interesting, but also have the potential for real-life impact in terms of promoting social interventions.[39]

References

Beaney, M. (2013). Analytic philosophy and history of philosophy: The development of the idea of rational reconstruction. In E. Reck (Ed.), *The historical turn in analytic philosophy* (pp. 231–269). Basingstoke: Palgrave Macmillan.

Bohman, J. (2005). Critical theory. In Zalta, E. (Ed.), *Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/critical-theory/.

Bright, L. K. (2017). Logical empiricists on race. *Studies in the History and Philosophy of Science: Part C, 65,* 9–18.

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally interpreting intersectionality theory. *Philosophy of Science, 83,* 60–81.

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

Brun, G. (2017). Conceptual re-engineering: From explication to reflective equilibrium. *Synthese*. https://doi.org/10.1007/s11229-017-1596-4.

Cappelen, H., Gendler, T., & Hawthorne, J. (2016). *The Oxford handbook of philosophical methodology*. Oxford: Oxford University Press.

---

[38] As pointed out by an anonymous referee, it is not clear that the distinction between manifest and operative concepts is one that the Carnapian can avail herself of. However, as seen above, Carnap did recognize that speakers often do not know themselves what exactly they are talking about, which suggests that there might be room for such a distinction in a neo-Carnapian conception of explication perhaps.

[39] For example, the ConceptLab project currently being hosted by the University of Oslo (http://www.hf.uio.no/ifikk/english/research/projects/cl/) is now working towards implementing some of this potential.

Carnap, R. (1934). Theoretische Fragen und praktische Entscheidungen. *Natur und Geist, 2,* 257–260.

Carnap, R. (1947). *Meaning and necessity*. Chicago, IL: University of Chicago Press.

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

Carnap, R. (1963). Replies and systematic expositions. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 859–1013). La Salle: Open Court.

Carus, A. (2008). *Carnap and twentieth-century thought: Explication as enlightenment*. Cambridge: Cambridge University Press.

Collins, P. H. (2003). Some group matters: Intersectionality, situated standpoints, and Black feminist thought. In T. L. Lott & J. P. Pittman (Eds.), *A companion to African-American philosophy* (pp. 205–229). London: Blackwell.

Craig, E. (2007). Genealogies and the state of nature. In A. Thomas (Ed.), *Bernard Williams*. Cambridge: Cambridge University Press.

C. Dutilh Novaes (2015). Conceptual genealogy for analytic philosophy. In J. Bell, A. Cutrofello, P. M. Livingston (Eds.), *Beyond the analytic-continental divide: Pluralist philosophy in the twenty-first century* (Routledge Studies in Contemporary Philosophy) (pp. 75–110).

Dutilh Novaes, C., & Geerdink, L. (2017). The dissonant origins of analytic philosophy: Common sense in philosophical methodology. In S. Lapointe & C. Pincock (Eds.), *Innovations in the history of analytic philosophy* (pp. 69–102). London: Palgrave Macmillan.

Dutilh Novaes, C., & Reck, E. (2017). Carnapian explication, formalisms as cognitive tools, and the paradox of adequate formalization. *Synthese, 194,* 95–215.

Haslanger, S. (1999). What knowledge is and what it ought to be: Feminist values and normative epistemology. *Philosophical Perspectives, 13,* 459–480.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs, 34,* 31–55.

Haslanger, S. (2006). What good are our intuitions? In *The Aristotelian Society—Supplementary volume* (Vol. 80, pp. 89–118).

Haslanger, S. (2012). *Resisting reality*. Oxford: Oxford University Press.

Horkheimer, M., & Adorno, T. (1972). *Dialectic of enlightenment*. New York, NY: Seabury.

Jeffrey, R. (1992). *Probability and the art of judgment*. Cambridge: Cambridge University Press.

Jenkins, C. S. I. (2014). Intuition, 'intuition', concepts and the a priori. In A. Booth & D. Rowbottom (Eds.), *Intuitions*. Oxford: Oxford University Press.

Jenkins, K. (2015). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics, 126,* 394–421.

Justus, J. (2012). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science, 2,* 161–179.

Kelly, T. (2008). Common sense as evidence: Against revisionary ontology and skepticism. *Midwest Studies in Philosophy, 32,* 53–78.

Leitgeb, H. (2013). Scientific philosophy, mathematical philosophy, and all that. *Metaphilosophy, 44,* 267–275.

Maher, P. (2007). Explication defended. *Studia Logica, 86,* 331–341.

Mormann, T. (2012). Carnap's boundless ocean of unlimited possibilities. In P. Wagner (Ed.), *Carnap's ideal of explication and naturalism* (pp. 63–78). Basingstoke: Palgrave Mcmillan.

O'Neill, J., & Uebel, T. (2004). Horkheimer and Neurath: Restarting a disrupted debate. *European Journal of Philosophy, 12,* 75–105.

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Pearson, J. (2017). Carnap, explication, and social history. *Social Theory and Practice, 43,* 741–774.

Reisch, G. A. (2005). *How the cold war transformed philosophy of science. To the icy slopes of logic*. Cambridge: Cambridge University Press.

Rinard, S. (2013). Why philosophy can overturn common sense. In T. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 4, pp. 185–213). Oxford: Oxford University Press.

Romizi, D. (2012). The Vienna Circle's "scientific world-conception": Philosophy of science in the political arena. *HOPOS: The Journal of the International Society for the History of Philosophy of Science, 2,* 205–242.

Sachs, C. (2011). What is to be overcome? Nietzsche, Carnap, and modernism as the overcoming of metaphysic. *History of Philosophy Quarterly, 28,* 303–318.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. Berlin: Springer.

Strawson, P. F. (1963). Carnap's views on constructed systems versus natural languages in analytic philosophy. In P. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (pp. 503–518). La Salle: Open Court.

Tappenden, J. (1995). Extending knowledge and 'fruitful concepts': Fregean themes in the foundations of mathematics. *Noûs, 29,* 427–467.

Uebel, T. (2004). Carnap, the left Vienna Circle, and Neopositivist Ant-Metaphysics. In S. Awodey & C. Klein (Eds.), *Carnap brought home. The View from Jena* (pp. 247–277). La Salle: Open Court.

Wagner, P. (2012). *Carnap's ideal of explication and naturalism*. Basingstoke: Palgrave Macmillan.

Wartofsky, M. (1982). Positivism and politics: The vienna circle as a social movement. *Grazer Philosophische Studien, 16*(17), 79–101.

Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton: Princeton University Press.