

Similarity in conceptual analysis and concept as proper function

Louis Chartrand

February 15, 2019

Abstract

In the last decades, experimental philosophers have introduced the notion that conceptual analysis could use empirical evidence to back some of its claims. This opens up the possibility for the development of a corpus-based conceptual analysis. However, progress in this direction is contingent on the development of a proper account of concepts and corpus-based conceptual analysis itself that can be leveraged on textual data. In this essay, I address this problem through the question of similarity: how do we evaluate similarity between two concepts, as similarity relates to identity? After a survey of prominent conceptual analysis methods, I propose a cursory account of corpus-based conceptual analysis. Then I formulate the question of similarity, and argue for an account that is functionalist in Millikan's (1984) sense. In this process, I propose a new account of concept that bases itself on millikanian teleosemantics in order to account for concepts' contribution in discourse. I then illustrate its fruitfulness by showing how it enables accounts of concept presence detection in textual data, both automatically and by a human judge.

Say a philosopher, named Alice, wants to study a given concept—in particular, she wants a picture of how it is being used. She gathers a very large corpus, large enough that for most concepts, she will have enough instances in the text so that she can observe the full variety in kinds of sentences, narratives, arguments and contexts in which it is used. In other words, her corpus is large enough to assume that it is representative of the kinds of discourses that run within the context where it was collected. As a very competent reader, Alice can intuitively pick up concepts when they are used.

However, she can hardly translate this “picking up concept” into a set of procedures¹. Indeed, this “picking up concepts” should not be assimilated with, say,

¹A reviewer brought to my attention an argument that can be brought up against any empirically based conceptual analysis that resorts to studying folks' understanding of a concept. Say we want to study ordinary people's knowledge of a concept, then we have to grant that the participants of our study understand the concept in question. But then, if it is folk knowledge, admittedly it shared by all the community, and unless the researcher is from a different culture, then she should also understand the concept as well as her subject. Why, then, doesn't she

picking up words that stand for it: concepts are often implicit, and they may exert their influence on a text’s content and structure without there being a word or set of words that make reference to it. When Alice picks up a concept’s presence in the text, she is not merely recognizing material shapes, but recognizing the concept’s role in the discourse’s structure, at least as she understands it. This kind of operation is opaque even from Alice’s point of view: while we can learn to better read and understand from others, we cannot tell exactly which operations take us from ink shapes to a certain concept.

In this scenario, if Alice has read the corpus, she probably has strong grounds for trusting herself with the various judgments that she makes as part of her interpretative activity. So long as she trusts her competence as a reader, she could go through the texts, identify the sentences which mobilize the concept she is interested in, and make an inventory of what the corpus tells us about it. The problem is that she probably lacks the time and resources to read the whole corpus by herself. As she needs to outsource parts of her reading and interpretative process in order to be able to treat massive amounts of data, she might not be able to trust the results of this outsourcing, even if she knows which operations are being performed. Given the opacity of her competence as a reader, even a simple operation such as detecting a concept in the text becomes a challenge.

An algorithm that detects the presence of concepts in textual data, such as the ones developed by researchers of the LANCI in the last decade (e.g. Chartrand et al. 2016; Pulizzotto et al. 2016; Chartrand, Cheung, and Bouguessa 2017)², might go a long way towards helping Alice. Indeed, given the importance of concepts in philosophical practice, we have speculated that the lack of computational tools to detect concepts in textual data is one of the reasons why philosophy is lagging behind other disciplines of social science and of the humanities with regards to the penetration of natural language processing and text mining in the research practice (Chartrand et al. 2016). While there is some opacity in the way these algorithms make their interpretative decisions, computer scientists will usually lean on our faith in human judgment in order to validate their algorithms: they engineer and evaluate them by comparing them to what humans would do when they perform the same operation. For example, Chartrand, Cheung, and

simply reflect on her concept, and produce an account of it from her armchair?

An obvious answer is that the researcher might be more interested in about the concept than simply her personal account of it, which might be skewed by her social positioning and personal history. However, there is more to it, as this argument fails if we distinguish, as we have done here, between the capacity to use a concept in everyday uses, such as discourse structuring or comprehension or for producing statements about the world, and the capacity to represent that concept as an object for discussion (or, in other words, to make it explicit). We might call the first kind of capacity “operative knowledge” of the concept being inquired, and the second kind might be called the “theoretical knowledge” of the concept. At the beginning of the study, we might grant the researcher an operative knowledge of the concept she wishes to study, but what she is looking for is a theoretical knowledge of it—she is certainly not looking for information she already has.

Cf. also footnote 10 on page 11.

²See also Chartrand (2019).

Bouguessa (2017) had participants annotate text segments for the presence of a concept, and evaluated their method against this metric.

One might argue that this strategy of relying on the trust we have on human competence merely displaces the problem. Indeed, in practice, even getting annotators to make the right calls require that the annotation protocol be well thought through—short of which they might be fulfilling a different task. This, in turn, requires that we have a good understanding of what it is to detect concept presence in text.

In post-war analytic philosophy, the association of conceptual analysis with *a priori* (non-empirical) knowledge (e.g. Jackson 1998) has meant that questions pertaining to topics such as the observability of concepts in empirical data have remained underdeveloped. While recent discussions around experimental philosophy and its methods have led to some progress on this topic, the focus on adapting methods from cognitive and social psychology on one side and on the role of intuitions on the other has meant that little has been developed to characterize the role of concepts in natural language. On the other hand, concepts have been discussed as instantiated in language, such as in the notion of “lexical concept”, but it is often in a very limited role, where the concept is viewed as attached to a particular expression or lexical pattern, which typically brings up the concept in question by referring to it (e.g. Fodor 1998; Evans 2006). But concepts in discourse are often implicit; they may have an important role in structuring narratives or discourses without being attached to specific expressions. Conceptual analysis would be incomplete if it failed to account for the roles a concept plays when it is not directly expressed through reference.

Therefore, while it is probably true that the lack of algorithmic tools is an obstacle to the development of corpus-based conceptual analysis, it also seems that it is in need of a proper account of concepts (as it plays a role both in formulating a question in conceptual analysis and in concept detection) and corpus-based conceptual analysis itself³.

In this essay, I address this problem through the question of similarity: how do we evaluate similarity between two concepts, as similarity relates to identity? Concepts are public entities, and they achieve their roles by being repeated from an instance to another. However, individual humans likely don’t internalize concepts exactly the same way, which is to say that we likely have slightly different accounts of the same concepts. Furthermore, as we keep learning and updating these accounts, it is likely that those also change across time—I probably don’t have the same account of the concept of CAT (the animal) as when I was five years old. Still, we say of my current concept of CAT that

³ Not to be confused with, for example, Meunier *et al.*’s (2005) *Computer-Assisted Conceptual Analysis of Text*. Meunier *et al.*’s aim is to unearth associations of a concept (as it is explicitly employed in text) in order to contribute new knowledge to an interpretation. Corpus-based conceptual analysis, in contrast, shares a similar objective with experimental philosophy as it is employed for the sake of conceptual analysis: namely, the idea is to give an account of a concept as it is employed in relevant linguistic behaviour.

it is the same (in the relevant sense) as the concept of CAT that I had as a five-years-old. How do we judge this identity?

As we shall see, this question poses itself slightly differently in the context of conceptual analysis and in the context of concept detection. However, I will argue that there can be a single answer to these two varieties of the similarity problem.

This essay is divided in two broad sections. In section , I compare various ways of understanding conceptual analysis: the method of cases (and Machery’s (2017) understanding of it in particular), Haslanger’s (2012) three types of conceptual analysis, and Carnap’s (1950) explication. I propose that these accounts are mostly complementary, and offer a synthesis. In section , I formulate and address the problem of similarity. I assess three ways of understanding Carnap’s similarity criterion: intension, extension and function; and I argue that similarity by function is superior to its alternatives. To replace Carnap’s vague account of function, I offer a millikanian account of it, and I show how it translates into an account of the concept of CONCEPT and into a heuristics to measure similarity between concepts. Finally, in section , I illustrate how the millikanian framework, and in particular similarity as function, plays out in corpus-based analysis and in concept presence detection in particular.

Varieties of conceptual analysis

When talking about conceptual analysis in philosophy, two different ideas come to mind.

Firstly, in the mind of most analytic philosophers, the term “conceptual analysis” conjures a specific type of proposition, with the concept that is being analyzed (the *analysandum*) on one side, its deconstruction into other concepts on the other (the *analysans*), and an operator that asserts some form of identity between the two terms. Usually, this proposition expresses the *analysans* in the form of necessary and sufficient conditions: for instance, “a brother is a male sibling” expresses that the concept BROTHER can be analyzed into the concepts MALE and SIBLING, with both being necessary conditions for BROTHER, and being jointly sufficient. In this sense, a conceptual analysis is a form of representation. It does not tell much about how we can arrive to propositions of this type, but it does tell us about the constraints coming from the form and the properties and paradoxes that come from it (Cf. King 1998; Jackson 2013).

The second sense that is associated with the term “conceptual analysis”, on the other hand, speaks of method rather than form. To some philosophers (Lewis 1970; Jackson 1998; Chalmers and Jackson 2001), it evokes a method to produce a proposition that would be a conceptual analysis in the first sense. Traditionally, conceptual analysis has been mostly about unravelling “our” concept of something, which a philosopher can often study through her own account of this

concept, in a *a priori* manner—viz. without inquiring outside of the realm of her own mind. But it need not be that way, and indeed, many (e.g. Haslanger 2012; Machery 2017) use this term to refer to explicitly empirical methods.

This section’s aim is to make a short review of current accounts of empirically informed conceptual analysis. The motivation for this boils down to this: conceptual analysis is the context within which we shall understand both concept similarity and concepts themselves. In other words, our accounts of concepts and concept similarity will be those that serve the account of conceptual analysis that we shall adopt. Therefore, this section can be thought of as a clarification of the main concerns of this article, those of concept similarity (how is it measured?) and concept detection (how can it be theorized for operationalization?). In a first subsection, I propose a historical perspective on the roots of conceptual analysis in the naturalist/rationalist debate, with an eye for the *a priori/a posteriori* debate, which has been polarizing the way we understand conceptual analysis and engineering in philosophy, especially in the second half of last century. Then I go on to describe the main frameworks through which the relation between conceptual analysis and empirical data have been theorized in the last few years – the method of cases, haslangerian analysis, and carnapien explication⁴. Finally, I show how those different accounts fit together in the context of corpus-based conceptual analysis.

Historical roots of conceptual analysis

While analysis has been a prominent part of the philosopher’s toolset for millennia, we often trace back contemporary analysis to Kant and his analytic/synthetic distinction. Kant is interested in statements as subject-predicate pairs, and calls “analytic” those in which the predicate is contained in the subject and “synthetic” for which it is not the case. For instance, the idea of having three sides is present within the concept of TRIANGLE, thus “All triangles have three sides” is an analytic statement. This dichotomy is closely associated to another, which deals with the means of acquiring truth values for a statement: if we need experience of the world to determine such truth values, then a statement is *a posteriori*, if it can be determined without experiencing the world, then it is *a priori*. Kant thought that there were no such things as analytic *a posteriori* statements, and his successors mostly rejected the possibility of synthetic *a priori* statements,

⁴The reader could probably point out to other frameworks that could fit the bill. In particular, one might argue that debates around natural kinds, for instance, should be addressed. However, these accounts address a very limited subset of concepts: those whose main goals are to refer to natural phenomena in ways that enable descriptions of the world that are as accurate as possible. Not all language is scientific language, and for good reasons; most concepts are adapted to other activities and fulfill other objectives which are not less commendable (Cf. Haslanger 2012; Carus 2008). Furthermore, I have not addressed other historically significant accounts of conceptual analysis, as I felt I should prioritize on current accounts of conceptual analysis. Machery (2017) has addressed the same three frameworks, admittedly for similar reasons.

such that, for most purposes, these two dichotomies are usually addressed as a single one, with the analytic *a priori* being opposed to synthetic *a posteriori*.

For the better part of the 20th century, analysis has thus been presented as a polar opposite to empirical inquiry. This said, the place occupied by this dichotomy in philosophy of science goes beyond the mere separation of analytic from empirical truths, as a defining research question has been to determine how analytic truths are to be integrated in the body of scientific knowledge (Rey 2018). For instance, at least for the early Carnap of the *Logische Aufbau der Welt* (Rudolf 1928), in order to have content, the theoretical terms with which scientific theories and claims are formulated ought to be reducible to observation terms. Analysis, or “rational reconstruction”, is thus the production of a form of definition, whereby a scientific term is related through rules to observation terms. Such definitions, however, have different conditions of adequacy than sentences about empirical terms: whereas the latter gets a truth value when confronted with observation, definitions are adequate if they reflect a convention (Rey 2018). As such, one must distinguish between the language in which empirical statements are produced and the language of reconstruction, with the former reflecting observation and experimentation, and the latter reflecting convention. To a degree, one can thus see the project of the *Aufbau* as attempting to draw a line between analysis and empirical inquiry and between the corresponding languages and epistemologies, and attributing them their roles in the production of scientific knowledge.

With Quine’s “Epistemology Naturalized” (Quine 1971), the debate eventually becomes polarized between reformers of the project of the *Aufbau*—soon joined by defenders of *a priori* methods of analysis—and defenders of naturalized epistemology—often called respectively “rationalists” and “naturalists.” Rather than reconstructing the meaning of empirical concepts through analysis, Quine’s suggestion is that we study how the construction of these terms actually proceeds. Knowledge is thus seen as a natural phenomenon, and the project of epistemology should be to study where and how it emerges. The same goes for empirical concepts, whose meanings are not to be determined by an elaborate definition leading us all the way to a primordial empirical language, but by a function of the processes of categorization they enable.

One of the central friction points is around the question of *a priori* statements. A naturalized epistemology would seek to replace *a priori* analysis of empirical terms with scientific accounts of those terms as they are reliably employed (Rysiew 2017). On the surface, it might seem like it is just another, perhaps more scientific, way of determining what our concepts are. However, rationalists would argue that naturalists who think that they are turning their back on *a priori* intuitions are in fact presuming or assuming them (e.g. Bealer and Strawson 1992). For example, as Bealer suggests, we need to use intuitions to determine what counts as empirical evidence rather than, say, *a priori* intuitions, imaginations or memory. Alternatively, if our starting-point intuition about mundane concepts were wildly unreliable, we might not be able to bootstrap

them to acceptable concepts.

Prominent responses to this challenge often choose to concede Bealer’s point, to the extent that they concede that empirical inquiries need a starting point. But those starting-point judgments need not be interpreted as *a priori*. To Kornblith and others (2002, 13), “the extent to which naive investigators agree in their classifications is not evidence that these judgments somehow bypass background empirical belief, but rather that background theory may be widely shared.” Even judgments which seem to rely on information that we share from the moment we are born are likely informed by lessons learned through our species’ evolution. From Kornblith’s perspective, *a priori* judgments, or at least the judgments that are referred to with this expression, exist and are relevant in epistemology, but they are best explained as natural abilities that draw from experience, including the experience of our ancestors.

It is unclear that this response really addresses the qualms of traditional epistemologists, as explaining intuitions as natural empirically-informed abilities relies ultimately on intuitions, and this explanation isn’t available to the epistemic agent at its starting point. On the other hand, Kornblith suggests that we might not be more justified in trusting *a priori* intuitions whose legitimacy seems somewhat supernatural. Thus the debate over the *a priori* takes a sort of chicken-or-the-egg flavour: it seems to depend on which perspective—e.g. the natural or the phenomenal—one is starting from.

The distance between naturalists and rationalists should not be overstated. On the one hand, of course, Kornblith’s arguments does not target the practice of using *a priori* intuitions, but rather suggests that the source of their legitimacy do not lie where rationalists think it is. On the other hand, rationalists are not necessarily opposed to the project of revising our account of knowledge in light of discoveries in cognitive science (BonJour 2006), and neither do they take *a priori* intuitions to be unrevisable in light of empirical knowledge (Bealer and Strawson 1992).

Furthermore, the middle way between a “pure” naturalism and a “pure” traditional aprioristic epistemology is actually well-travelled. For instance, Goldman (1986, 2005) has argued consistently that intuition-based conceptual investigation must be the starting point of epistemological inquiries (Rysiew 2017). On his account, intuitions can be interpreted as a window to our internal concepts, and methods to elicit them can be seen as ways to gather evidence for conceptual analysis. On the rationalist side of things, Canberra planners have gone so far as to reclaim the “naturalist” label, in part because of their general commitment to physicalism, and their lack of commitment to the primacy of the *a priori* over the *a posteriori* (Braddon-Mitchell 2009). Moreover, the rationalist’s armchair often looks suspiciously susceptible to empirical inquiry: for instance, the Canberra planners’ “two-step” method for conceptual analysis begins by collecting all the platitudes about this concept⁵ (Nolan 2009).

⁵There is some controversy around what should count as a platitude. Generally speaking,

It would also be a mistake to associate empirical inquiry with the naturalists to the exclusion of the rationalists on account of their positions with regards to the *a priori/a posteriori* dichotomy. After all, the initial positivist project, as it is developed in the *Aufbau*, far from developing a discipline disconnected from empirical inquiry, portrays philosophy as “the handmaiden of science” (Braddon-Mitchell 2009). Furthermore, this separation did not necessarily imply that the analysis should stick to the armchair. It is explicitly in this spirit that Arne Naess pioneered experimental methods strikingly similar to modern experimental philosophy during his years attending the Vienna Circle, and while the project has not been well received by all of the Vienna Circle regulars, Carnap himself saw this as a positive development (Naess 1938; Murphy 2014).

More recent attempts at informing philosophers’ accounts of concepts are also hard to split along the rationalist/naturalist lines, but a generalization can perhaps be made: while naturalists analyze concepts to ensure that they capture the right phenomena or objects, rationalists put more emphasis on capturing *our* concept of something. This is not unexpected, as the naturalist project is more about building concepts from observation and experimentation⁶, whereas the rationalist project begins with an assessment of the concepts we have before diving into data. A typical naturalist project would be, for instance, to determine whether the physical extension of the concept MIND should be limited to the brain or diffuse into a creature’s environment (cf. Clark and Chalmers 1998; Hurley 1998; Clark 2008; Rupert 2009), and would draw heavily on research in psychology, anthropology, neuroscience, etc. to argue for its case. Conversely, more typical of a rationalist project would be to probe laypeople’s intuitions about a concept in order to determine how they understand it.

Therefore, experimental philosophy, which will provide part of the framework for contextualizing corpus-based concept analysis, is probably more rooted in the rationalist tradition, and might be thought as the rationalist response to the naturalists’ use of cognitive science research for their own projects. This is evidenced by its focus on thought experiments and its methodological reliance on intuitions. However, it is worth noting that not all of this focus is an endorsement: in fact, while it is far from forming the bulk of the research in experimental philosophy (cf. Knobe 2016), much of it is devoted to what has been dubbed *the negative program*, viz. a critique of the reliance on intuitions in philosophy. As a result, it is probably best to think of experimental philosophy having its roots in both traditions.

these would be claims that reflect commonplace uses of the concept.

⁶While both are ways of capturing empirical data, experimental studies and observational studies differ in the degree of control being exerted by the researcher. In experiments, the phenomenon being studied is provoked, typically in controlled conditions, such that causes and effects can be isolated. In observational studies, the researcher has no control over the phenomenon she is observing, she might make for more realistic environments, but makes it more difficult to ascertain causality and to control for unwanted interactions, among other things. While most of experimental philosophy has indeed been experimental, corpus analysis would rather qualify as observation.

The method of cases

The method of cases is, at its core, a sort of narrative that goes as follows. We have a concept which we suspect to have a certain attribute. For instance, we might imagine that in a categorization task—when judging whether a certain limit-case object is a representative of the said concept, or not—we think that having a certain feature is important in determining where it belongs. So we think up cases or scenarios where the said feature can be isolated, and test our judgment on it to see where it leads us. For example, Knobe (2003) suspects that whether a side-effect is positive or negative can have an impact on whether the person who brought it about is responsible for it or not. So he concocts this scenario:

The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.’

The chairman of the board answered, ‘I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.’

They started the new program. Sure enough, the environment was harmed.

He tours around Central Park submitting either this case, or a similar one where every instance of the verb “harm” is replaced by “help”, and asks whether the chairman is responsible. As people are twice as likely to say he is in the “harm” cases, Knobe concludes that the valence of the side-effect (its being good or bad) is important in the folk concept of RESPONSIBILITY⁷.

Alternatively, one can do this kind of experiment esoterically, between the author and its readers. The Gettier cases (Gettier 1963) are often understood like this: Gettier thinks that there is more to knowledge than its common analysis—according to which knowledge is justified and true belief. Therefore, he proposes this case (p.122):

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

- (d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith’s evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones’s pocket ten minutes ago. Proposition (d) entails:

⁷At least, it was the case in 2003. Since then, Knobe has adopted the view that this effect is probably more of a widespread cognitive effect than a feature of the concept of RESPONSIBILITY (Pettit and Knobe 2009).

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not know that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

Gettier thus concludes that the “justified true belief” account of knowledge is inadequate.

As Sosa (2007) points out, it is not necessary to associate the method of cases to conceptual analysis. For instance, he suggests, cases can be used to argue for or against any philosophical theory, including those that are not about concepts.

For Sosa and others (among which Bealer 1998; Gopnik and Schwitzgebel 1998; Ludwig 2007; Goldman 2007; Chalmers 2014), intuitions are what drives us to make the relevant judgments on the cases. Their value comes from our competence in making judgments—for instance, when making judgments about concepts, these judgments would derive from our competence in using those concepts. We might assume, in turn, that we would have acquired this competence from living in a society that uses those concepts, or from our experience in using this concept. As a result, the method of case can be seen as a way to highlight those intuitions and make them explicit.

However, the proponents of an intuition-based method of cases have struggled, over the years, to establish intuitions as sources of evidence or other epistemic guarantees for the method of cases. Formulations of the concept INTUITION (understood in the context of philosophical method) are numerous, although they rarely have clear boundaries, and all the most prominent formulations have been the target of numerous critiques. Variety in accounts of a concept is not in itself a problem—we lack consensual accounts of many important concepts, such as COGNITION, LIFE and DEATH, and this does not count as a failing for the theories that rely on those concepts. But intuition-mongers have had to defend against charges that intuitions are too volatile to fulfill their epistemic role (Alexander and Weinberg 2007; Machery et al. 2004; Swain, Alexander, and Weinberg 2008) on one side and arguments that they are not central in the practice or logical structure of philosophical argumentation (Williamson 2008;

Cappelen 2012; Deutsch 2015; cf. also Nado 2016; Pohlhaus 2015) on the other⁸.

Following Machery (2017), defenders of intuitions can be broadly divided into two broad camps⁹. On the one hand, there are those who would have intuitions be a special kind of mental state or competence, whom Machery calls “particularists” and “exceptionalists”: for them, intuitions are not just any opinion or felt state, they are distinct in virtue of things like type of content (e.g. abstract or modal), psychological or phenomenological properties (automaticity, speed, “being drawn to”, etc.), etiology (e.g. came from experience, accepted competence), epistemic status (e.g. reliable opinion), etc. The reason why we would want intuitions to be particular or exceptional lies primarily in the mechanics of the subject/object dichotomy as it functions in the dispatch of epistemic work: we expect one side of the dichotomy to do one part of the epistemic work (provide evidence) and the other to do the other part (evaluate the information, synthesize it, draw conclusions, etc.). Mixing those responsibilities can yield paradoxes and fallacies (Cf. Williamson 2008, 2013; Ichikawa 2009). For instance, say I grant evidential status to intuitions, but I think of intuitions simply as being no different to other judgments. Then using my intuition of P as evidence for my judgment that P would translate into justifying my judgment that P as evidence for my judgment that P . Even if we manage to work around this paradox, there is a legitimate concern that we could be tainting our intuitions with our opinions and motivations¹⁰. Shielding intuitions from the rest of the mental lore by affirming their distinctness serves to avoid this kind of difficulties.

On the other hand, there are the minimalists (among whom Machery 2017; Williamson 2008; Ichikawa 2009) who hold that the method of cases does not require that there be a special epistemic status for intuitions. For the minimalist, intuitions have no special phenomenology, they have no special epistemic or semantic status, they are neither necessarily analytic or justified *a priori*, and

⁸There has been others charges against intuitions. For example, Machery (2017) argues that they are a bit of a nomological dangler and Pohlhaus (2015) argues (among other things) that, in the way they are formulated and employed, intuitions rely on an assumption of universality which is epistemically noxious.

⁹Machery talks about three camps, but one delineation is more important than the other.

¹⁰This argument also offers an answer to the question posed by a reviewer (cf. also footnote 3): “Why doesn’t the researcher simply reflect on her concept, and produce an account of it from her armchair?” From this point of view, armchair reflection runs the risk of contaminating the data. As Machery (2017: 234-5) argues, probing others’ intuitions provides protection against this risk.

A partisan of armchair methods might counter by arguing that armchair methods are rarely confined to a single armchair, but that such conceptual analyses are actually developed in the interaction with colleagues and graduate students. It is not completely clear that this would solve the problem, as a researcher’s colleagues might be as motivated as her towards a conclusion. However, even if this difficulty were circumvented, Machery argues that it might not be reasonable to expect that the general population will share the intuitions of a small group of philosophers, as it has often turned out not to be the case (234-5). This is why experimental philosophers typically avoid probing philosophers’ intuitions when they want to know what is the “ordinary” account of a concept, or the account of a wider population. This is also why, in the following chapters of this thesis, we avoid using a corpus authored by philosophers.

they have no special etiology. They certainly don't come from a different faculty, they have no special psychological status (they can be fast and automatic, but also slow and deliberate), and they don't need to be obvious or conscious.

The motivation for this view largely comes from the perceived failure to pin down specific properties that stand up to scrutiny and successfully manage to map onto what and only what we would want to consider as intuitions. For instance, Williamson (2008) argues that judgments obtained after some reflection should not be less eligible to the status of intuition than spontaneous "seemings". Indeed, it seems to conform with what philosophers do: we don't need a lot of deliberation to agree that the Searle in Searle's Chinese room does not understand the Chinese characters he's manipulating, but people will take a pause before answering a trolley problem or the violinist dilemma. In a different kind of argument, Williamson shows that restricting what we call intuitions by invoking particularities can lead to undergeneration. His example is that we should count it as intuitive that there are mountains, but these restrictions often lead to categorize this statement as not intuitive. Intuitions are thus behaviourally and phenomenologically diverse, and, as Nado (2016) notes, this almost certainly means that we are also facing psychologically diverse phenomena. To account for intuitions, it would seem, we need to be liberal, and accept any judgment or opinion.

But then, minimalism (or liberalism, as Ichikawa 2009 calls it) could be facing the same problem that particularists and exceptionalists were trying to avoid in the first place. Opinions are exactly the kind of things that an argument is meant to sway, so using them as evidence and as ground for said argument seems like begging the question. Machery's defence is to embrace what he calls "sociological psychologism" and consider intuitions as indicators for judgments that happen to be widely shared, as opposed to being the ultimate support of philosophical arguments. So, while opinions from a reader or from the author might indeed have been corrupted, we can survey the opinions of those outside of the ivory tower, and get a good idea of what the widely shared judgment is from participants untainted by philosophical debates.

Thus, by expanding the domain of what counts as intuition, Machery significantly expands the domain of what can count as evidence for a certain account of a certain concept. As he notes, the judgments provoked by common thought experiments such as Kripke's Gödel case and Foot's trolley case lack the features that mainstream particularist theories of intuition deem they should have, hence the significance of minimalism for case-based experimental philosophy.

Given the importance that intuitions have historically had in conceptual analysis as its source of evidence, what counts as intuition is relevant for both case-based experimental philosophy and corpus-based conceptual analysis. However, if it turned out that particularists were right, the method of case would simply have to restrict the judgments it elicits, and perhaps philosophers would try to rewrite the Gödel and trolley cases so as to elicit truly intuitive judgments. As such, particularists do not pose much of a threat to experimental philosophy.

On the other hand, it is not impossible that particularism about intuitions could restrict our ability to produce data from textual corpora. Indeed, it is not clear that, given our current level of understanding of writing, we could discriminate or control for the linguistic behaviours observed in a text that are not the product of intuition, understood the way a successful particularist theory would understand it. Because minimalism advocates for an inclusive account of INTUITION, it dispenses us with the need to determine which linguistic behaviour reflects intuitive judgments, and which behaviour reflects non-intuitive judgments. Minimalism makes it possible to look at all the linguistic behaviour involving a concept that can be found in a corpus.

Haslangerian analysis

While the method of cases gives us some tools to formulate the nature and role of empirical data in philosophical analysis, it isn't sufficient for analyzing a concept. What makes a proper portrait of a concept depends on what we are trying to do with it, therefore we need tools to formulate our philosophical research projects and derive research and evaluation strategies.

When philosophers are tasked to study thinking tools, they can be involved in three broad types of projects. Firstly, they can study the thinking tool as it is used. For example, they might try to give an adequate portrayal of a concept as they believe it is used from their own understanding of the concept. They might also try to study a corpus (as Von Eckardt 1995 did when studying the concept of COGNITIVE). Or, as experimental philosophers have been doing, they might design and run experiments in order to understand through behaviour how participants conceive and employ these thinking tools.

Secondly, philosophers can study thinking tools as they work within their own systems. This might take the form of studying a concept within a formal system. For instance, a philosopher might study how removing the law of the excluded middle gives rise to different regimes of logic and logical thinking. This work has often been characterized as being *a priori*, but one might also think that it is analog to building reduced models of a new plane in order to study its aerodynamics: the idea is to play with the object to understand how it behaves in various conditions that can be expected to arise.

Thirdly, philosophers might study a thinking tool with the express objective to improve it. They might appreciate how the tool works in a certain context, and wish to adapt it to another or they may believe it has a certain failing, and wish to modify it in order to correct it. They might even think that current tools are not getting the job done in a fundamental way, and try to build up new thinking tools from the ground up in order to replace them. Typically, philosophers who engage in such works have a definite idea of the function of the thinking tools they target, and they will take measures to ensure this function is properly fulfilled.

Concepts are a kind of thinking tool. When we take it to be a constituent of propositions, a concept enables the representation of a set of thoughts that mobilize it. When we take it as holding knowledge about regularities in the world, it enables inferences about its objects (Millikan 1998). As a result, philosophers who study concepts can also adopt three kinds of strategies, which mirror the three ways of approaching thinking tools in general. They are broadly those described by Sally Haslanger (2012, chp. 13). The first one is the *conceptualist* or *internalist* approach: it corresponds to the question, “what is *our* concept of X?”—“our” corresponding to “our community”, as in the community that is involved in using and perpetuating a concept, be it philosophers, members of a Western society, etc. Here one might argue that once a philosopher goes about the world asking others if they share their intuitions about a concept, “*internalist*” becomes a bit of a misnomer. Nevertheless, the gist of the question is the same: it is about achieving an understanding of concepts as they are actually used as human thinking tools.

The second is the *descriptive* approach to analysis. Haslanger describes it as being involved in understanding “what objective types (if any) our epistemic vocabulary tracks” (p. 386). For example, a descriptive analyst might wonder if our concept of DOLPHIN actually corresponds to what dolphins really are. Or a descriptive analysis might inquire if what falls under the concept of DOLPHIN should really fall under this concept. Realizing that there is no principled reason to lump together oceanic dolphins and river dolphins, a philosopher involved in such a project might propose that this concept should only refer to oceanic dolphins. Thus, there is a normative aspect to this approach: if the concept being studied is not as efficient as we might like in tracking its objective kind, then the analyst will suggest adjustments. However, the suggestion of changing the concept DOLPHIN here is not driven by any concern for our understanding of dolphins or our interactions with them, but rather on the apparent disjointedness of the category. Nor is it concerned with whether current use of the concept DOLPHIN would require this revision. It is concerned with the workings of the concept within its conceptual system—in other words, it studies concepts as thinking tools as they work within their own system.

With this in mind, we might want to think that the descriptive approach should apply not only for the referential function of concepts, but also for other functions. Indeed, concepts don’t always refer to something, they only do when it is their role to do so. For instance, when concepts are expressed in verbs like “apologize” or operators like “and”, they often function to position the participants of a conversation, or to determine how other concepts fit together in the context of a proposition. Some concepts might even work within a formal system that lacks semantics, and then, studying what the concept does—and whether it does it well—isn’t about object tracking but about whether it performs the relevant operations in the relevant contexts. As a result, perhaps “*functionalist*” is a better term to describe this approach than “*descriptive*”¹¹.

¹¹Haslanger’s views are strongly realist, in that, for her, concepts like KNOWLEDGE and

Finally, projects of *ameliorative analysis*¹² put this function in question: “What is the point of having the concept in question [...]? What concept (if any) would do the work best?” (p.386) In such works, we might find a philosopher introducing a new fauna of concepts in order to achieve a theoretical goal (e.g. Millikan 1984: chp. 1-3), or arguing for a redefinition of a concept in order for it to fulfill a new, or a modified, role.

These three approaches are connected in various ways. For instance, there is a sense in which all projects are ameliorative: whether someone is making explicit how we use a concept, or what its role and operation within a system is, this person is always making a new concept to represent what she found. We do not encode the concept of APPLE in the same way when we use it to recognize apples in visual stimuli or when we use it to talk and think about them. Likewise, it is different to encode a concept simply to perform its role in a formal system or to describe to a colleague the role it has been given in it. Inevitably, there is change, which involves gain in the fact that the concept can be used in new contexts (e.g. explanation, reasoning, etc.) and often some losses. In any case, any successful conceptual analysis, whether conceptualist, functionalist or ameliorative in its approach, will yield a new concept that is identical to its initial target in a way, but different in that it is tailored for new contexts. There is also a sense in which conceptual analysis always involves a descriptive or functionalist analysis: one could not forego a careful examination of a concept’s function and functioning before proposing adjustments, and without attending to function, a conceptualist analysis would be a mere description of use cases. And finally, any functionalist or ameliorative analysis ought to have a grounding in actual use, which involves a conceptualist analysis.

This is not to say that conceptualist, functionalist and descriptive approaches are all the same. The difference lies in the purported contribution: a conceptualist analysis’ contribution lies in a better understanding of actual use, a functionalist analysis must give us a better understanding of the function and operation of the concept and an ameliorative analysis should improve on the function itself. However, no single analysis stands by itself: a good understanding of a concept involves a good understanding of its use, of how it operates and of its role.

This being said, to Haslanger, differences in types of analyses actually boil down to different projects with different objects. Among them, she makes a distinction between *manifest* and *operative* concepts. The former is typically the one that transpires from an explicit description, like a law or a rule (e.g. “a kilogram shall be defined as whatever weighs as much as the standard in Paris”), while the

JUSTICE have a referent that is an objective kind—not a natural kind, but a social kind. As a result, from her point of view, all concepts that may be of interest for conceptual analysis have referents. I’m not willing to commit to such a view, and I feel this work can be of use to those who are also disinclined to adopt it, hence my redefinition of descriptive analysis to one that is more tolerant of non-realist views.

¹²Haslanger (2012) also uses the term “analytical approach” (p. 352) in reference to a tradition in contemporary feminist theory. But authors who reference her work mostly use “ameliorative analysis”.

latter is more like the often implicit, but effective characterization that transpires from use. The object of a conceptualist analysis is typically the manifest concept, while that of descriptive or functionalist analysis usually is the operative concept. Meanwhile, ameliorative analysis' object is the concept that it tries to create: the target concept.

The problem here is that Haslanger does not provide us with a clear-cut way to distinguish between manifest and operative concepts. While she does suggest some criteria, they are sometimes in tension with the examples she produces. For instance, on page 388:

Consider again my requests to Zina (my daughter) that she lower the volume of her music. Suppose I don't want to listen to music with misogynistic lyrics. I have a concept of misogynistic lyrics and I also have a rough-and-ready responsiveness to what she is listening to. When Zina complains about my interventions into her listening, I may come to find that my responses are not tracking misogynistic lyrics after all, even though that's the concept I was attempting to use to guide my interventions. Let's call the concept I thought I was guided by and saw myself as attempting to apply, the *manifest concept*.

In this passage, the manifest concept seems to be the conscious one, the one that she perceives to be applied, while the operative concept is the one that actually reflects her interventions. Indeed, as she explains a little later (p. 398):

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.

However, in another example, she speaks of the concept of PARENT, as in the institution of parent-teacher conferences. A parent for a human being is usually understood to be one of the two persons that are the immediate progenitors of a person. However, when the school invites parents to a parent-teacher conference, they actually mean to invite the primary caregivers of the children attending classes. Here, there is no disconnect between the concept the school authorities take themselves to be applying and the one that is actually applied: if asked who counts as a parent, they would describe the caregiver, not the progenitor. Therefore, in this example, Haslanger describes the manifest concept as "the concept that speakers generally associate with the term", and the operative concept is "the concept that captures how the term works in practice"¹³ (p. 390). In a further example (p. 368-370), she speaks of how her son and his school have different definitions for the tardiness: for the school, following the official rulebook, a student is tardy if she arrives in class past 8:25, but in practice, teachers have different policies concerning tardiness, so a student can arrive past 9 on Wednesday and still be on time, because the Wednesday teacher will not

¹³By this, Haslanger probably has in mind something along the lines of "social and institutional practice" rather than "linguistic practice".

mark her down as tardy. Once again, both of these conceptions of tardiness are conscious and explicit (perhaps the second one is a little less so, as it would not do to make it explicit in some contexts, but students can certainly do so), so the difference here rather seems to be that the manifest concept is the one that is associated with an authority, while the operative concept is not.

If the boundary between the manifest and the operative seems to be changing, it might be because rather than being a problem, it is a feature of the dichotomy that it is dependant on the context. In *Resisting Reality*, this dichotomy first comes when she addresses theories of the concept RACE¹⁴. These theories take conceptual analysis to be mainly descriptivist: concepts (at least the referential kind) must map onto essences. Concerning races (as applied to humans), there is no identifiable essence. Therefore, there are no races, and the concept RACE is superfluous—that is, if we take the descriptivist analysis to be the only legitimate one. Distinguishing between manifest and operative concepts enables Haslanger to argue that there is more than one aspect of the concept, which in turn suggests that there should be more than one possible type of analysis. This opens up the possibility for a complementary account of RACE, that is based on the operative concept: in such an account, race might be seen as a concept that plays a variety of social roles, branding those who are identified with a race for discrimination and special treatments.

Given the context in which it is introduced, it seems that the manifest/operative distinction serves to open door rather than circumscribe the concept in a dichotomy. It is meant to distinguish her projects from other types of conceptual analyses which have wrongly assumed to be the only game in town. Branding their object as the manifest concept enables Haslanger to simultaneously frame their contribution to a larger research endeavour, while opening up the space for new types of conceptual analyses. But once those projects lose their claim to monopoly or higher authority, it isn't clear at all that the types of conceptual variations they are interested in is systematically different to the ones they are not interested in.

Therefore, we may read Haslanger's contribution as being mostly critical: conceptual analysis cannot be understood as transparent representation of an object (the concept) into a descriptive language. Rather it is dependent on the kind of purposes that we have for this description, and on how we have carved the object (which contexts are to be included in the study). Furthermore, how concepts function, be it in scientific or naturalistic project or in a social context, is fundamental: conceptualist and descriptive or functionalist projects will ultimately aim to faithfully represent how the manifest or operative concept functions in the contexts they wish to represent, while ameliorative concepts will try to reforge their object into a target concept that performs the function that we want them to perform.

But there is perhaps one fundamental dimension where Haslanger's three types of

¹⁴This is Haslanger's example (2012: 383-385).

concept actually differ, and it is in what we might call their empirical grounding (what Haslanger calls “subject matter”). While a conceptualist analysis might produce a concept that fails to capture a natural kind and that fails to have any noticeable effect within the discourse community for which it was crafted, it will be judged much more harshly if it fails to capture what our concept actually is, because it will have failed on its own terms. One might say that, by virtue of how Haslanger’s framework divides conceptual analysis labour, conceptualist analyses (even if they are done from the armchair) will have their empirical grounds in phenomena that indicate how we conceive of a concept (e.g. linguistic or categorization behaviour) and descriptivist analyses will have their empirical grounds in phenomena that indicate something about the objects themselves. Even ameliorative analyses will be grounded in an understanding of the context in which concepts are meant to play as well as the role they play in it (in particular, their purposes, the constraints that act on them, the mechanism in which they participate, etc.). Haslanger, of course, frames this as knowledge of the why: “why do we have a concept or belief?” And it is from the answer to this question that the analyst can move to the central question of ameliorative analysis: “what concept (if any) would do the work best?” It should be obvious that without knowledge on the workings of a concept in a linguistic context and/or in a community, any speculation on this matter would be moot.

One takeaway from this is that, firstly, all conceptual analyses have an empirical ground, and therefore, all conceptual analyses have the potential to profit from empirical data¹⁵. This is why even the rationalist tradition, which has been symbolically associated with armchair speculation, has seen its proponents attracted to various attempts to incorporate empirical data in the debate, as we have seen in section . Secondly, analyses of every type actually have some stake in the empirical grounds of other types of analyses, because despite the division of labour, a successful concept has to be successful in working with the concepts that we actually have (as opposed to the ones a philosopher might think we have), in fulfilling its function and in being suited for the context in which it is meant to play. This is why, for example, in the context of Carnapian explication, experimental philosophy has been proposed to play a role in informing us on those three aspects of the concept (Shepherd and Justus 2015; Pinder 2017; Koch 2019). Thirdly, there is nevertheless value in distinguishing among these different types of conceptual analysis, if only to apply the right standards of evaluation.

¹⁵Furthermore, because experimental philosophy works by provoking linguistic behaviour, this statement also applies in principle to textual data found in corpora. Indeed, if in a study experimental philosophers elicited linguistic behaviour that could not be found in any possible corpus constructed from speech and writing taking from natural settings, then this would suggest that the experiment is not ecologically valid.

Carnapian explication

Carnap’s project lies somewhere between scientific ambitions of the conceptualist and descriptive projects and the revisionary ambitions of the ameliorative projects. In *The Logical Foundations of Probability* (Carnap 1950), for example, Carnap is involved in clarifying concepts that scientists already commonly use: DEGREE OF CONFIRMATION, INDUCTION, PROBABILITY. These concepts are “usually sufficiently well understood for simple, practical purposes” (2), but Carnap gives himself the task of reaching a more precise understanding of them through the method of *explication*. Put this way, his project does not seem revisionary: we expect those concepts to keep functioning the same way, at least in the simple and practical purposes for which they are already commonly employed. However, Carnap recognizes that clarifying a concept necessarily produces a new concept: clarifying ambiguity involves determining features of a concept, and thus modifying it. Hence, explication has both a conservative and a revisionary aspect.

Carnapian explication describes a process by which we form a new, more precise concept (the *explicatum*) from a typically less precise and relatively unscientific concept (the *explicandum*). Much like Haslanger’s ameliorative analysis, the process is described, at least by Carnap himself (1950: chp. 1), as a two-step process¹⁶. Since the *explicandum* is relatively imprecise, the problem that an individual explication is meant to solve is, at the outset, never very clear. This is why we ought to clarify the *explicandum*. Carnap suggests that this can be achieved by giving examples of contexts where we use the concept we wish to explicate, and examples of contexts where we might think it is being used, but where the concept mobilized is actually a distinct concept. For example: “I mean by the explicandum ‘salt’, not its sense which it has in chemistry but its narrow sense in which it is used in the household language” (Carnap 1950: 4-5). Once we have clarified the *explicandum* as such, we can go on to provide an explication, which “may be given, for instance, by the compound expression ‘sodium chloride’ or the synonymous symbol ‘NaCl’ of the language of chemistry” (*idem*).

Furthermore, explication is usually about taking a concept from a conceptual paradigm (or “system of concepts”, as Brun 2016 calls it) and making it available to another. Explicating table salt with a chemical formula enables us to insert this concept into chemical discourse. For scientific purposes, explication can thus serve as a bridge between various models (Meunier 2017). For instance, if we want to know whether crows are more intelligent than finches, we might need to explicate the folk concept of intelligence into a relevant ethological framework. From this, we might want to formalize this ethological concept of intelligence, so that crow behaviour and finch behaviour can be made comparable. Then,

¹⁶Brun (2016) further analyses explication into a four-step process, which he obtains by making the role of evaluation in explication more explicit. Thus, after clarifying the *explanandum*, one should clarify how the various criteria should be interpreted in the context of the explication, and the fourth and final step is simply a critical appraisal using those criteria.

this formalized concept can be explicated into a physical/experimental concept that can be used to construct experimental protocols and applied to observed behaviour. In this string of operations, the explicator must have guarantees that, at each transition, what has been learned about a concept from the *explicata* can permeate back to the *explicanda*. For example, if crows do better than finches in a set of experiments, it is through a well-constructed structure of explications that we can convert statements about behaviours into general statements like “Crows are more intelligent than finches” in everyday settings.

Carnap does not spell out many constraints on what an explication may be, but he gives us the means to guide and evaluate it in the form of four criteria. The first is *similarity*: the *explicatum* must be reasonably similar to the *explicandum*, which is to say that in the contexts that count, the *explicatum* has to be able to do the job of the *explicandum*. Then comes *exactness*: since this is the whole idea of the project of explication, it stands to reason that the *explicatum* has to be more exact, “so as to introduce the explicatum into a well-connected of scientific concepts.” (Carnap 1950: 7). Of particular importance is the criterion of *fruitfulness*, which he describes as usefulness for the formulation of universal laws in *The Logical Foundations of Probability* (Carnap 1950). As Dutilh Novaes and Reck (2017) put it, the *explicatum* ought to be conducive to the production of new knowledge. Finally, the weakest criterion is *simplicity*—a criterion that Carnap presents as a tie-breaker between equally good candidates for an *explicatum*. In other words, an explication that is simpler (more parsimonious) than another is *ceteris paribus* also better.

While the context of scientific discovery is central to Carnap’s project, it might be useful to generalize explication beyond this restricted endeavour. This would mean, for instance, that a concept being fruitful might mean more than just enabling the creation or formulation of new knowledge, but that it also could lead to some social improvements. In fact, even when restricting ourselves to the scientific project, we can see how fruitfulness comes to represent very different things in different cases. For example, Carnap discusses the explication of the everyday concept FISH (which would include such things as whales and cuttlefishes) with the scientific concept PISCES (cold-blooded gill-bearing vertebrates¹⁷). There are things that can be said of pisces that cannot be said of fishes (in the old sense): that they evolved from an amphioxus-like creature, that they are chordates, etc. Because it carves nature at the right joints, this new conceptualization enables new generalizations. Now, compare this with the kind of explication we employ to make experimentation possible: for example, when we explicate CONSCIOUSNESS as “the content of the participants’ experience as she or he is able to convey it”. In such a case, the *explicatum* does not shed any light on the subject matter by itself, but only because it enables the construction of an experimental protocol. It appears that fruitfulness is not tied to any specific way of contributing to knowledge—all that is required is that

¹⁷This is Carnap’s definition, but it is not perfect, as it would include the axolotl, but exclude some lungfishes. Wikipedia’s *Fish* entry defines them as “gill-bearing aquatic craniate animals that lack limbs with digits.”

it enables knowledge to progress further. Therefore, employing explication for other purposes is by no means a big stretch¹⁸.

As a referee of Dutilh Novaes (2018) notes, it is easy at this point to imagine that ameliorative analysis and explication might, to a certain degree, be fruitfully thought of as being approximately the same. Dutilh Novaes initially counters by noting that explication projects can be pursued within any approach of conceptual analysis that Haslanger describes, including conceptualist or descriptivist/functionalist approaches. She also notes that in exactness, explication possesses a criterion that is absent from Haslangerian analysis, and that, in turn, ameliorative analysis employs tools, such as ideology critique¹⁹, that are absent from explication. However, on the one hand, if some descriptivist projects can also be explications, perhaps it is that descriptive analyses could also be described in terms of an ameliorative analysis, with the goal of the amelioration being of a scientific nature (perhaps “carving nature at its joints”, or promote a goal that would lead to new knowledge down the line). On the other hand, Dutilh Novaes argues that explication and Haslangerian ameliorative analysis should take inspiration on each other, which is testament to the fact that their respective virtues are not restricted to the domains of issues from which each type of analysis originates. Perhaps, down the line, we will see explication projects that include ideology critiques, and ameliorative analyses that make attempts at formalization, such that they will become indistinguishable. Furthermore, neither Haslanger nor Carnap mean to define their respective forms of analysis. As we mentioned, Haslanger’s trichotomy seems to be aimed at opening the stage for new types of analysis, rather than closing down the possibilities to three rigid types of projects. Carnap, on his side, formulated his criteria in ways that could afford a variety of interpretations. Therefore, while we can agree with Dutilh Novaes that explication and Haslangerian analysis are still different things, it is not clear that it should remain so.

This said, there is at least one sense in which explication goes farther than haslangerian analysis, and it is in its capacity to link concepts from different systems of concepts in a relation of identity. Haslanger takes Appiah’s descriptivist analysis of RACE and her own ameliorative analysis of it to be about the same

¹⁸Some might argue that the move would affect the exactness criterion. Indeed, the association between exactness and scientificity might seem like a natural one, but it isn’t clear that it actually applies more specifically to scientific discourse. Brun (2016: 1222) argues that exactness is about such various objectives as reducing ambiguity (including reducing the amount of cases where we can doubt whether a concept applies or not), not leading to paradoxes and allowing for finer and more precise descriptions. Unlike Dutilh Novaes (2018), I don’t think we should read the exactness criterion as a call for formalization. Beyond the connection with Enlightenment ideals (Carus 2008), the conceptual hygiene that is evoked through the criterion of exactness does seem like a practical necessity for entertaining the desideratum of fruitfulness: how could we affirm that a new concept is fruitful, if we ignore whether it will lead to paradoxes or if we do not know how it will behave in limit cases? Furthermore, these are all important preoccupations in Haslanger’s politically motivated ameliorative analyses as well.

¹⁹Ideology critique is an analysis that is focused on the thinking tools like concepts and narratives that we employ to navigate the world. Cf. Haslanger (2012: 17-22).

concept. However, there is little substance about this identity, which goes from having its extension in things like genes and phenotypes to things like social representations and dynamics. Through Carnap’s account of similarity, which be unpacked in section section , we can explain why conceptualist, descriptivist and ameliorative analyses can give different but equally valid accounts of the same concept.

Unsurprisingly, explication’s binding powers also come in handy with experimental philosophy and corpus-based conceptual analysis, where it connects the language of the hypothesis to the language of experimentation. As a conceptual analysis project mobilizes a corpus in order to answer its questions, it must go from the concept as it lives in the context that has motivated the conceptual analysis to a concept that lives in its empirical domain. Most importantly, the connection between the origin concept and the target concept must be articulated in such a way that discoveries on the empirical side can translate into insights for conceptual analysis. Articulating an explication means articulating the conditions under which what can be said about the *explicandum* can also be said about the *explicatum*, and vice-versa.

Conceptual analysis for Alice

One interesting recent development in empirically-informed methods of conceptual analysis is a novel interest in articulating methods that have mostly evolved separately. Dutilh Novaes (2018) has investigated the intersection of explication and haslangerian amelioration analysis and has argued for convergence, and Machery (2017: 215-7) draws a similar parallel. Meanwhile, Shepherd and Justus (2015), Pinder (2017) and Koch (2019) have explored the possibility of using experimental philosophy to inform explication—though the main prize might have been to provide experimental philosophy with a method that avoids intuitions and its pitfalls.

The consensus, so far, has been that syncretism is probably a winning strategy for the development of conceptual analysis. After all, in articulating explication and haslangerian analysis, we have seen that we gain from an extension of the available tools on both sides (Dutilh Novaes 2018); we also gain Haslanger’s insight into the division of labour, and carnapien explications capacity to articulate the different accounts of a single concept that are conceived through conceptualist, functionalist and ameliorative analyses. Similarly, experimental philosophy contributes empirical grounding to explication, while explication brings in a way to connect concepts from philosophical discussions into the experimental (or observational) realm, thus foregoing the need for intuitions.

I find no reason to doubt this consensus. As noted in the previous section, haslangerian analysis and explication are formally the same. As such, adopting Dutilh Novaes’ strategy of recuperating insights and tools of inquiry from both sides makes sense. Thus, a syncretic analysis would probably employ the

elaborate methods devised for Carnapian explication (Cf. Brun 2016), but it would also have to position itself with regards to the division of labour, and use this position to leverage the information from other types of analyses of the same concept in order to improve upon itself.

However, syncretism might be a bit more difficult to achieve when it comes to determining how empirical data can inform analysis. For Shepherd and Justus (2015) and Koch (2019), experimental philosophy (and, we can assume, textual data analysis as well) can inform the explication preparation phase, where one has to get a clearer account of the *explicandum*. For Pinder (2017), on the other hand, the contribution experimental philosophy can make to explication preparation is too small to make it worthwhile, and the way should rather be to use experimental philosophy to probe the conceptual environment to which the *explicatum* is destined in order to predict if it will be successful enough to be adopted by the community. Koch (2019) disagrees, finding uptake to be a poor indicator of success, and judging Pinder’s plan difficult to materialize.

That being said, taking into account the division of labour in conceptual analysis, it seems that the most productive contribution of experimental philosophy for an explication should depend on the kind of explication. If we are in a conceptualist project, then the explication preparation is certainly the most important step, as the goal of the *explicatum* is to enlighten us on the *explicandum*. On the other hand, while we might agree that uptake might not be a good indicator of a concept’s quality, an empirical study of the *explicatum*’s conceptual environment is crucial for an ameliorative project, as we need to predict how the modified concept will play in it. As such, the debate between Shepherd, Justus, Pinder and Koch is probably misguided.

Finally, one might wonder where the method of cases fits in this picture. Machery’s (2017) take is that an overly aprioristic method of case, whereby philosophers only investigate their own intuitions and that of their friends, is empirically underpowered and might beg the question. His solution is to study intuitions on more representative samples, which then enables him to adopt a minimalist view of intuitions. As such, because it enables the full breadth of textual data from corpora to be taken into account, this solution also does a lot towards making corpus-based conceptual analysis practicable.

Thus, we might describe a syncretic method of conceptual analysis as follows.

To Alice, a conceptual analysis might start with an inquiry into the problem she’s facing²⁰. Firstly, she needs to clarify her problem—in particular, as Carnap suggests, she needs to clarify which concept she wishes to analyze (we will call it the *original concept*). She might also want to determine whether her project is ameliorative, conceptual or descriptive, and determine how her conceptual analysis might play in the philosophical debates in which she wishes to engage. From this, she will have an idea of the discursive and pragmatic space that the concept that will be constructed in the analysis (let’s call it the *target concept*)

²⁰Here I take inspiration from Brun’s (2016: §3) “recipe” for explication.

will have to inhabit. Hence, she can determine what purpose the target concept is meant to fulfill. From the knowledge of the target concept's purpose and purported context, she can infer constraints: form of representation, contexts where it needs to play the same function as the original concept, degree of exactness, etc. There will likely also be constraints that are specific to the function the target concept is meant to fulfill. For instance, if the target concept is meant to represent in a theoretical discussion a concept as it is used in a corpus, it will only be adequate if it efficiently reflects the original concept, and if Alice can be justified in thinking so. If, on the other hand, she wants to improve on a concept as it is used in a corpus, she will need to highlight how the original concept performs its function, and where it could be improved.

From this, Alice can plot a path getting to the target concept. In her case, this means that she will first need to determine what kind of corpus she needs to construct. On the one hand, from the way her project is formulated, she will be able to draw conclusions as to which set of assertions are relevant. For instance, she might want to have a corpus that is representative of the linguistic and discursive behaviour of the community that uses the original concept. Then, she needs to determine what kind of contexts are contexts where the original concept is present—that is, performing the discursive functions that are relevant to the question at hand—and which contexts are actually good indicators for the concept being analyzed. While Machery argued that there is no reason to demand that she discriminates according to, say, the mental faculty that is involved in applying the concept, there might be cases where a concept would appear to be associated to another concept only for discursive reasons that fall outside of its function: for example, if our corpus is collected during the 2018 World Cup, the concept STADIUM might seem strongly associated with the concept RUSSIA, but this does not reflect on the function of any of those concepts.

Finally, from observations and experiments on the corpus[], Alice can propose a target concept that fulfills the criteria as previously stated and interpreted, and evaluate how well the new concept fulfills its objectives.

Similarity

Between the vagueness of the terms employed to theorize conceptual analysis and the difficulties that arise in operationalization, theoretical and methodological difficulties abound—which is why we can only rejoice in the increasing interest philosophers have been putting into method and metaphilosophy. In the rest of this article, we will be concerned with a pair of related problems with regards to similarity.

The first problem is about identification between the original concept (in Carnap's words, the *explicandum*) and the target concept (the *explicatum*): how should

we understand this relation? Which requirement comes with it? It might seem that this problem is sometimes treated a bit lightly. For instance, in the *Logical Foundations of Probability*, Carnap (1950) insists that the requirement of similarity should be flexible (although he doesn't give any limits to this flexibility) and in "Replies and systematic expositions" (Carnap 1963), he proposes that if we interpret similarity as synonymy, we should allow at least three different senses of synonymy to be employed, depending on the context. Haslanger, on her part, does not explicitly address it, and Machery sees it as merely embodying a form of conservatism: "Concepts should not be modified without reason, and when they are modified they should be modified as little as possible" (Machery 2017: 215).

Surely, however, similarity is about more than just conservatism. Take cases where the target concept is meant to play a role for the original concept: for example, cases where the target concept is an *explicatum* that enables us to do experiments. Here, the objective is to learn new things about the target concept that will also apply to the original concept. Such a transfer from target to original concept supposes that the two concepts are similar enough that, barring some constraints, properties of one concept can be justifiably applied to the other. Inversely, in order to fulfill their role, the new concepts need some of the information that is contained in the old one. Indeed, it is hard to imagine a case where we don't want to transfer knowledge or functions from an original concept to the target concept. There is more at stake here than, say, prevent costly and unnecessary change: we need to justify the transfers between original and target concepts.

The second problem hits closer to home for philosophers who (like Alice) practice conceptual analysis with empirical data, and perhaps observational data in particular: how exactly are we justified in identifying two instances of the use of a concept as two instances of the use of the same concept? For example, how can we feel secure in thinking that two cases of thought experiments mobilize the same concept, or that the concept they mobilize is the concept we wish to inquire? Alternatively, in corpus-based conceptual analysis, how can we feel justified in thinking that two segments exhibit traces of the presence of the same concept?

These two questions could have demanded two distinct answers. Indeed, the causal threads which link concepts in those two questions might be of different nature: in a deliberate conceptual analysis like an explication or an ameliorative analysis, the target concept is constructed from the original one, whereas the dynamics of concept diffusion, drift and repeated reinterpretation that occurs naturally in a community are much less deliberate and likely are the result of a very different, natural evolution. As a result, we might expect that what unites and distances concepts in those two contexts would turn out to be very different. However, as we shall see, there is a single answer to these two problems.

In the rest of this section, we will assess various propositions for establishing how similar or dissimilar concepts can be: firstly, by similarity of intension,

then by similarity of extension, and then by similarity in function. “Intension” and “extension” are terms that tend to take different definitions depending who you are speaking to. However, generally speaking, intensions correspond to the internal content or the essential properties of a term or a concept. For instance, if the concept is determined by a definition, it would be that definition; if it is a cluster concept²¹, then it might be a list of properties, perhaps along with weights representing the importance of the property for a object instantiating the concept. For our purpose, intension shall be a representation of the properties germane to a concept. Extension is more straightforward: it is the objects that the concept is meant to represent. For our purpose, the extension of a concept is the set of possible objects that would be its instances if they are/were real²².

I will argue that there are major issues with similarity of intension, and extension that make them poor candidates for the similarity criterion. On the other hand, not only is function more apt to account for the similarity in a diversity of contexts, but it comes with a perk: it afford natural cutoff points for judgments of identity.

Similarity by intension

The question of similarity touches on the question of what is fundamental in a concept. When we say that humans are similar to chimpanzees, it often comes with some kind of evidence: sometimes, it is about DNA (“we share 98% of our DNA”), sometimes it is about ancestry (“they are our closest relatives”), sometimes it is about phenotype, behaviours like problem-solving or social mores or cultural transmission, etc. Whatever is mentioned, it usually is deemed fundamental, at least in the discursive context, of what it is to be a human or a chimpanzee as species. Of course, if we believe that chimpanzee or human essence (as species, of course) lies in DNA, phenotypic comparisons are not out of question, as genotype is a huge factor in determining phenotype, and, *ceteris paribus*, individuals with similar genotypes also have similar phenotypes. But if we can, we might as well hear it from the horses mouth, and check genotypic similarities. As such, we can assume that philosophers who think that a concept is its intension will also think that concepts which are similar to each other are concepts whose intensions are similar.

Those who see concepts as being intensions typically think of a concept as being its essential properties or predicates (with what is essential being largely dependant on what one believes to be the essential role of a concept in an organism’s cognitive economy). Essentially intensional concepts can be found in a large variety of philosophical and scientific traditions. In traditional conceptual analysis (e.g. King 1998), a concept just is its decomposition into necessary and

²¹An object instantiate a cluster concept if it possesses a certain number of the attributes that are associated to this concept, while none of these attributes is necessary or sufficient for instantiating the concept. Cf. Searle (1958).

²²I do not assume here that all concepts have an extension, as we will be clear in section .

sufficient conditions—a brother just is something that has both the property “sibling” and the property “male”. In cognitive psychology, concepts take different logical forms (Murphy 2004; Machery 2009; Harnad 2009), but they are also characterized by a form of subject-predicate association, even if the predicate is often fuzzy and neither necessary nor sufficient for categorization. Intensional criteria are also common in computer science. Proponents of the method of case are also typically fond of the intensional concept. For instance, Machery (2017) suggests that a concept is a set of belief-like states (“beliefs”) about the substance.

To evaluate similarity between intensions, we can encode them into digital representations. A standard way to do this is to code properties as variables, while a concept can be coded as a data point. In such a case, it might seem that geometrical measures such as the euclidean or the cosine distance would be an obvious choice to give us a good idea of how similar or dissimilar two concepts can be. However, while it would work well for cases like prototype or exemplar concepts, it would not work with concepts that need to be represented using more complex forms of representation, like schemata (Minsky 1975), and that cannot be represented as a point in a high-dimensional space without loss of information.

This said, there are perhaps other ways of measuring similarity and difference between digital representations that could perhaps bridge the gap between representations of very different forms. For instance, a promising avenue might be to think of two concepts as being a few modifications away from each other. By computing the minimum amount of modifications needed to go from one string representation to another, measures like the Levenshtein distance (Levenshtein 1966) and its derivatives can give us an estimation of this drift, and effectively tell us how similar these representations are. These measures can (and often are) easily adapted to measure differences between objects that have different logical forms; they could therefore be adapted to measure differences between representations of the intensions of two concepts. Furthermore, as these measures can be adapted to measure representations in various forms, they could possibly be applied to any concept intension, no matter its form.

However, it isn’t clear that similar intensional traces actually mean similar concepts (let alone identical). Take the FISH/PISCES example: one might describe a fish as an aquatic animal, whereas the PISCES intension also includes other properties, like having a skull, a notochord and gills and lacking digits on the limbs. Intensionally, it would seem that concepts like SIRENS (a family of gilled limbless aquatic salamander) are a lot more similar to PISCES than FISH is to PISCES: sirens also have notochords, skulls, gills and no digits on the limbs, but only the last property can be expected of all fish in the old prescientific meaning of FISH. This problem also arise in more natural settings, as we commonly describe the same things in various ways. Berenice might think that, essentially, water is H_2O , while Charles might think of it as a transparent liquid with the ability to quench thirst. Their intensional pictures of water

have nothing in common, yet they have little difficulty agreeing that they are talking about the same thing. There might be a sense of CONCEPT in which it is relevant to say that this person and I have different concepts of WATER. But if we are trying to understand how the community in which they both live understands the concept of WATER, it seems like both their voices should be included. In other words, for our purposes, intensional similarity does not seem to do the work we would want it to do.

Similarity by extension

The obvious next step after putting intension aside is to take a look at its externalistic twin, similarity by extension. Carnap (1950: 7), reports that this criterion is employed by Karl Menger (1943) with definitions: “A good definition of a word must include all entities which are always denoted and must exclude all entities which are never denoted by the word.” It is worth noting that Carnap, however, does not endorse this view for concepts (cf. Brun 2016).

Nevertheless, similarity by extension has some things going for it. In particular, it would work a little better in practice, at least with the example that we just mentioned. Whether you think of water as H₂O or as a liquid that quenches thirst, the extension remains the same. Indeed, it would seem sensible to think that it is because these two intensions refer to the same thing that Berenice and Charles are talking about the same thing. This said, in the FISH/PISCES case, Carnap notes that the latter is much narrower in extension than the former, and thus, that “they do not even approximately coincide”. However, thanks to work pioneered by Rosch (1973), we now know that, at least in people’s minds, not every instance of a concept counts equally: there is a sense in which a carp is more of a fish than a seahorse, or in which an apple is more of a fruit than a pineapple. Perhaps the proper way of measuring similarity is through a weighted metric that gives more importance to co-extension in instances that are more emblematic of the concept. In such regards, PISCES and FISH are certainly similar, as they conjure the same exemplars of carps and trouts.

However, there are grounds to doubt that it is always fruitful to think of a concept as referential (even sometimes going through great lengths to find a domain where it can be instantiated). Quite often, it is more useful to think of a concept as a tool, say, to structure discourse, knowledge and behaviours. For instance, some concepts are used in ways that suggest that their main or only function is to position an assertion pragmatically or rhetorically (e.g. APOLOGIZE in “We apologize to our readers.”), or to convey mood or attract attention (e.g. IMPORTANT in “This package is very important.”). If we adopted similarity by extension, then we might be unable to use the similarity criterion on those concepts, which might be a problem.

A way out would be to find strategies to assign an extension to every concept. Perhaps we should force ourselves to think of APOLOGIZE as a verbal form

referring to acts of positioning oneself in discourse, and of IMPORTANT as having for extension the set of all things that are deemed important by someone. One worry with this solution is that this might actually change with the meaning of a concept. It does seem that there is something performative in calling something important that goes beyond asserting that something belongs to the set of things that are important. Much like explaining a joke will ruin it, explaining why we think that something is important will not have the same effect as calling it important.

More importantly, when concepts are abstract, we may be tempted to draw their extensions in more than one place, all of which might be equally adequate. Brun (2016) gives us an example of this from Stalnaker (1976):

“the proposition [a sentence expresses] will be a function taking possible worlds into truth values. Equivalently, a proposition may be thought of as a set of possible worlds [...]”. (Stalnaker 1976: 80)

Here, PROPOSITION₁ extends on functions, while PROPOSITION₂ extends on sets. Therefore the extensions for PROPOSITION₁ and PROPOSITION₂ are disjointed. While we should expect those two *explicata* to be very similar, judging only by extensions would tell us that they are very different. It seems that, at least for concepts which lack extension in the physical world, extensions are not an appropriate way of judging similarity.

Similarity by function

Arguably, one of the best interpretations of Carnap’s own understanding of his criterion of similarity is by way of comparing concept functions. Indeed:

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

This “most cases” ought to be interpreted as “relevant cases” (Brun 2016; Dutilh Novaes and Reck 2017)—i.e. relevant for the problem the explication is meant to solve. Thus, the idea here is that in those relevant cases, the *explicatum* and the *explicandum* are interchangeable in use, which is to say that they perform about the same function²³.

Similarity by function has much going for itself. Firstly, it avoids the problems that we ran into with intension and extension. Stalnaker’s explications for PROPOSITION are equivalent because, even though they have different intensions and extensions, they can still perform the same function (at least in most contexts). While it is possible to have explications where intensions and

²³Here, function has a sense similar to “purpose” or “role”, and is not closely related to the mathematical function or the function in computer science. Therefore, we should not think of this functionalism in the traditional way. Cf. Millikan (1984: 18).

extensions are completely different in *explicandum* and *explicatum*, there still needs to be functional overlap: for a piece of knowledge to be applicable to both, there needs to be a sentence embodying that knowledge where they play the same function.

Secondly, even though these functions are not the traditional functionalist's kind²³, they can be realized in multiple ways. Thus, variations between or within individuals are not an issue. It doesn't matter if different authors in a corpus think of a concept in different ways: if they are using it in a similar way, we can be confident that it is the same (or about the same) concept. This also means that we can encode a concept in different ways (as Machery 2009 suggests) and still be talking about the same thing, so long as there is some functional overlap.

Still, Carnap's account of similarity by function is quite thin, and if we're to apply it systematically, we need more details. Functionalism about concepts can find a more elaborate account if we mobilize Ruth Millikan's (1984, 1998, 2017) works, as it provides a more precise notion of "function" and a system of concepts to go with it. To understand it, we must first understand that Millikan's account is tightly dependent on a peculiar understanding of the ecosystem that houses cultural artefacts such as words and concepts.

Millikan starts from the realization that words, concepts and other cultural artefacts are subject to evolutionary pressures of sorts. On the one hand, they are subject to replication: effectively, we use the words, concepts and other linguistic devices that we have encountered before, so we are replicating the communicative behaviour we have seen in others. On the other, replication is not perfect. If Debra teaches a concept to Eleanor, Eleanor might retain a slightly modified version of the concept, or might even use it in slightly different ways. Debra herself might use a concept in a standard way on most days, but might get creative in certain circumstances, and count on the intelligence and culture of her audience to play with that concept's meaning. Therefore, there is also a space for innovation and semantic drift.

Furthermore, replication does not happen randomly: if we replicate a concept, a word or another language device (as Millikan calls them), it is because we wish to accomplish something, and the language device helps us accomplish that something. Another way of saying it is that we replicate them for a certain function. The function in virtue of which linguistic device is replicated is what Millikan calls this device's "proper function." That is, the proper function is not necessarily the function in virtue of which a specific instance of a linguistic device has been used, but rather the function in virtue of which the linguistic device is being replicated in general in a linguistic community. In other words: the function that ensures that a linguistic devices remains alive in its community. So, to take an example: what identifies a word is its proper function—the function that ensures its being replicated across time and contexts. Depending on the context, "happy" can express a lasting contentment or an ephemeral joy: these two "happy" express different things, have different functions, and therefore are actually different words despite their being associated with the same morpheme.

The nature of linguistic communication, with its requirements for some sort of alignment between speakers and hearers, creates what Millikan calls “stabilizing proper functions” or “standardizing proper functions.” Indeed, under normal conditions, the function of a language device ought to be of value for both the hearer and the speaker, or else the exchange would collapse for lack of cooperation. This implies a certain uniformity in function: if the speaker wants a certain reaction from the hearer, she better stay conservative and employ language devices as they are most employed. Inversely, if the hearer wants to extract the right information from the exchange, she will want to use conservative interpretations.

Now, how should we account for concepts and their functions in such an ecosystem? For simplicity’s sake, let us first consider words once again. To Millikan, a word alone has no proper meaning of its own; rather, meaning is imparted to sentences (1984: 80), and words have meanings in the context of sentences. Sentences are themselves constructed by replicating syntactic forms—that is, patterns of word arrangements that serve specific rhetorical purposes. For instance, “Long live the revolution!” and “Down with the tyrant!” share a simple syntactic form, where the first slot serves to express a sentiment towards the object that is in the second slot. Other syntactic forms, like “Would you . . . ?”, “Could you . . . ?”, “I would like . . . please.” need to be adapted to (and with) other syntactic forms in order to construct a proper sentence. Thus, the role of word in a sentence is mediated by the syntactic form that inserts themselves in the sentence.

While it may happen that a conceptual analysis actually analyzes a word (in Millikan’s terminology) rather than a concept (indeed, Brun 2016 argues that Carnap explicitly accounts for this possibility in the context of explications), prototypical conceptual analyses from Carnap and Haslanger portrait concepts as accomplishing a lot more than just sentences. Millikan’s concept of WORD is tightly associated with lexical forms on the one hand and sentences on the other. It is close to what computational linguists call “sense”: a single semantic unit associated with a word or expression. On the other hand, Carnap and Haslanger see concepts as structuring discursive, scientific and political practices in general. For instance, PISCES organizes entities in such a way that makes it possible to make new true statements about them, and Haslanger’s concept of PARENT acts as a sort of gatekeeper for some social institutions and practices.

What about concepts? Millikan does have a concept of CONCEPT, but it is not quite what Carnap and Haslanger are talking about when they are talking about concepts. For Millikan, most of the time, “concept” is short for what she calls “empirical concept” (her writing implies that there might be non-empirical concepts, but to my knowledge, she never really develops this notion). Empirical concepts are public ways of referring to kinds or properties: they are shared like words, and they can be used to identify the entity they refer to as well as hold information about it (Millikan 2017: 47). As such, at its core, such a concept is the ability to categorize between instances and non-instances (Millikan 1984:

253), but because this ability is dependent on what I know about the concept, it is also the set of information that we have about its object(s).

However, on the one hand, Millikan herself has recently rejected her concept of CONCEPT (Millikan 2017: 49: “my claim is that there are no such things as empirical concepts”). On the other hand, even if we were to argue against her that there really were such things as empirical concepts, much like Millikan’s concept of WORD, this concept of CONCEPT does not fulfill the role that we need it to fulfill. Not only does it differ with the concept of CONCEPT that is employed in conceptual analysis, but it fails to account for many of the concepts we may wish to study. Indeed, if there are such things as concepts that do not refer, we would like to be able to account for them. Thus, Millikan’s concept of CONCEPT might be too restrictive for our purposes.

This does not mean that we should abandon the idea of a concept of CONCEPT that is functionalist in a millikanian sense. Millikan provides us with an elaborate system of concepts to talk about cultural artifacts in terms of proper functions and selection; there is no reason why we should not be able to construct a concept of CONCEPT out of it that satisfies our needs.

Millikan (1984) puts a strong focus on language, but she addresses questions that are narrowly focused on relatively esoteric topics of philosophy of language using concepts that have a much wider applicability. In the first three chapters, she constructs a set of concepts meant to talk about biological and cultural artifacts in terms of what accounts for their pervasiveness—viz. their systematic reproduction and selection. This is where she defines such concepts as proper functions and stabilizing proper functions. This part then serves as a theoretical ground for the rest of the book, which is more narrowly about language, and in particular the topics of language that drew the interest of the analytic philosophers at the time. Millikan’s interest in, say, proper function, is thus narrow, but her system of concepts have been applied elsewhere—including in philosophy of biology (e.g. Schwartz 1999), epistemology (e.g. Plantinga 1993; Nolfi 2016), meta-ethics (e.g. Wisdom 2017) and semiotics (Menary 2007).

Concepts, as they interest conceptual analysts, touch on a large and discontinuous domain. They are employed in language, but not only at the level of sentences: they are used to construct narratives, stories, tropes, arguments, and other linguistic constructions that structure discourse and lie above the level of sentences. They are also manifested in social, scientific and political practices, rituals and institutions, such as laws, parent-teacher conferences and experimentation best practices. These are all things that are reproduced in a similar way that sentences and language devices are reproduced.

For instance, parent-teacher conferences are events that get replicated regularly in schools across the world. To use Millikan’s terminology, they are members of reproductively established families: because they successfully fulfill a certain proper function, they are allowed to be reproduced, and thus form a family of similar instances. Much like sentences (Millikan 1984: 22), parent-teacher

conferences are formed from a variety of model: they retain forms that are also prevalent in other meetings (like plans for orders of business, presentation rounds, etc.) as well as forms that are relevant to the kinds of discourses or social context which is specific to these kinds of meetings. Furthermore, they articulate various cultural devices (parents, teachers, children, learning, child development, speech, etc.) in a coherent whole that promotes the proper function of the parent-teacher conference. In the same way that the concept PARENT acquires a meaning in a sentence, the concept PARENT also acquires a meaning in a parent-teacher conference. Indeed, concepts relate to higher-level cultural artefacts like parent-teacher conferences or narratives in the same way words relate to sentences: they have no proper function of themselves, but rather, they have derived proper function from their association with other devices to form these higher-level entities.

Therefore, we might think of a concept as the thing that composes higher-level entities, in accordance with the role that is imparted to them by the discursive forms²⁴ that model the higher-level entity. It is an analog to the millikanian word in the context of the sentence—indeed, in a sense, the millikanian word is a special type of concept for which the higher-level entity is the sentences. If grasping a word is grasping its proper function as manifested in the syntactic forms that bind it to sentences, then grasping a concept is grasping its proper function as manifested in the discursive forms (for lack of a better word) that bind it to higher-level entities.

Now that we have placed our concept of CONCEPT in a millikanian system of concepts, we can wonder how it helps us with similarity. If a concept is grasped by way of its proper function, it is also individuated by way of its proper function. Given that it is always a derived proper function—derived from the proper function of the higher-level entities that the concept composes—we can get a sense of a concept's proper function through the contexts in which the concept is used. Hence perhaps Carnap's suggestion: a good way of getting a sense of a concept is to enumerate relevant contexts where it is mobilized; and a good way of comparing two concepts that are suspected to be similar is to test whether they could replace one another in these contexts.

Another way of framing it is to consider a community where two concepts are both used. Let us assume we have access to all the linguistic discourses they produced (we can also assume this corpus to be indefinitely large), and know in each instance which concept was used. Then, we could compare two concepts by comparing their contexts—that is, the higher-level entities in which they were involved—and the roles played within those contexts. The concepts would be similar insofar as they play the same roles as assigned by the same discursive forms, and insofar as they are put in the same relationship with other concepts that are mobilized in their respective contexts.

²⁴I think of the discursive forms as higher-level entities' analog to the syntactic forms for sentences.

While we lack such capacities as that of collecting every instance of every use of every concept in a community, or that of automatically identifying an instance of a concept to its type in every situation, this idealization does suggest some indicators. Carnap's heuristic of producing assertions is one of them, but it only seems appropriate in cases like explication or ameliorative analysis, where the target concept or the *explicatum* has as of yet no real existence. Producing assertions can serve as a form of simulation for the sake of predicting how the new concept will function in its new environment. When we have data on how a concept was actually used, as is the case in Alice's scenario, we can use heuristics to identify concept occurrences in the corpus and construct representations of their contexts. These representations can be compared to produce an index of similarity. This is the principle behind paradigmatic relations in distributional semantics (Sahlgren 2008) and word embeddings in computer sciences (Mikolov, Chen, et al. 2013), which have proved very efficient at predicting word similarities and at uncovering semantic relationships between words. Thus, not only is functional similarity measurable in a corpus, but its measure is a well-established practice in corpus linguistics and natural language processing.

As we have seen, a functionalist's similarity avoids the pitfalls that similarity by intension or extension fall into. Furthermore, it explains why Carnap's heuristic for judging of similarity is a good one, and it hints at an explanation for the success of similarity indices based on paradigmatic relations in natural language processing. And as we mentioned, it also comes with an additional perk in that a millikanian framework can afford a clear-cut criterion for a concept instance to belong to a concept type.

In the spirit of flexibility, Carnap did not suggest any way of finding a cut-off point beyond which the *explicatum* is too far from the *explicandum* for there to be a sort of identity between the two. In the context of an explication, it might not be too important of a problem: the idea is that any loss in similarity ought to be offset with gains in fruitfulness and precision, so we don't necessarily need a cut-off point. But in the context of a corpus, it can be important to know where a concept begins and ends, and which contexts mobilize a concept of interest and which only mobilize a similar, but distinct, concept.

Millikan has a proposition for a cutoff point in the context of words (cf. Millikan 1984: 72-5). Words, in her account, are individualized by reference to their genealogy—they are to be categorized with words that are reproduced from the same lineage of previous uses of the same word type. The verb "to mean" comes from a different etymology as the noun "mean", which entails that they are obviously not the same word. But "to mean" can take a variety of meanings: to convey meaning, to intend, to be important, to be sincere or even to bring about. Until the 19th century, "to mean" as "to be important" was not a standard usage of the verb "to mean". Certainly, in some context, English-speakers would have been able to make sense of a usage of "to mean" in this sense, but they would have understood by inferring from the proximity in meanings and from the context. Therefore, even in the case of an abnormal usage, reproduction

of the verb “to mean” would have been driven by the proper function of “to mean” when it means “to convey meaning”. When it became standard for “to mean” to mean “to be important”, then what was driving this use of “to mean” was not the same proper function—communicative acts were not successful in virtue of an inference from a similar meaning, but rather because audiences were habituated to see “to mean” as meaning “to be important”. Therefore, the driving force of the reproduction of “to mean” as meaning “to be important” was now a new proper function. This, to Millikan, is the birth of a new word. This kind of stabilizing proper function is what Millikan calls a *least type*: the narrowest proper function that manages to drive its reproduction. The same reasoning can be applied to concepts in general. Therefore, to determine whether two concepts are identical, one simply needs to determine if their function can be boiled down to the same least type.

To recapitulate, we have argued that similarity between concepts should be understood as similarity by function. In order to clarify this proposition, we have turned to millikanian teleosemantics and its concept of proper function, which contributes a more precise account of function (as proper function) while grounding it into the dynamics of communications within a community. From this basis, we have introduced a concept of CONCEPT using Millikan’s ontology, as an analog to her concept of WORD which can play the roles of a concept within the realm of conceptual analysis. As such, this concept articulates its roots in Millikan’s system of concepts with the requirements of conceptual analysis, as we have come to understand it in section section . Furthermore, it is haslangerian in the way we defined higher-level entities to include not only the usual descriptive structures like propositions, theories and models, but also non-descriptive linguistic structures and even structures that shape our political interactions like schools and parent-teacher conferences. But it is also carnapien inasmuch as it addresses a problem in carnapien explication.

Millikanian concepts for corpus-based conceptual analysis

Now, we might wonder if the millikanian framework described above can actually help with corpus-based conceptual analysis.

From a certain perspective, it could appear that this account of functional similarity is bad news for corpus-based conceptual analysis, because functions are not what we observe directly. In practice, we can never be certain that the apparition of a word, for example, has been driven by the reproduction of a certain proper function rather than another. Furthermore, a function in this sense harder to express than a subject-predicate association or a subset of an extension.

However, there are reasons to think that concept as function actually makes

things easier from Alice’s perspective. Firstly, a proper function is not about what caused something, but about what normally causes something in a certain environment under normal, everyday conditions. Therefore, it is not opaque: it could not be transmitted or replicated if it were private. Furthermore, given that we are talking about communication events, function cannot lie, say, in the emitter alone; the reader, as a participant in the communication, also has in principle a privileged access to the function of its components. Secondly, while there is no straightforward way to represent a function, it might be possible to find proxies for it. For instance, in a corpus, we might expect functions to be associated with distributional patterns.

We can explain and illustrate these points by showing how they can be applied in corpus-based conceptual analysis, and in the task of detecting concept in particular.

On the one hand, we mentioned in the introduction that computer scientists use human judgments as a way to evaluate and improve the efficiency of an algorithm, and thus to give us confidence in its judgment. In the case of concept detection, there is no material obstacle to asking humans to do exactly the kind of task that we are asking the algorithm to do, and then comparing their answers. Thus, while Alice might not have to detect concept presence by herself if an algorithm does it for her, someone at some point has to be able to make those judgments.

In normal conditions, if there is a shared body of linguistic devices between receivers and emitters, humans usually have an intuitive grasp of when a concept is present in discourse. Thus it makes sense to ask participants “Is concept *C* present here?” However, there are various shades and variations to this perception: not all concepts feel present in the same way or the same degree. A concept may be present in the theme—for example, it might be the very subject we are discussing about—but it can also play a supporting role in the argumentation, or be vaguely alluded to.

So what task is it exactly that people making judgments about concept presence are making?

It is relevant here to recall that higher-level entities can only play their linguistic role properly if all of their constituents are taken into account. Concepts, as we have described them above, are constituents of higher-level entities which, in the context of a text corpus, means higher-level discursive entities, such as sentences, of course, but also narratives, arguments, and other higher-level entities that structure text. These higher-level entities are what embody the message—what is being communicated. They are thus what needs to be understood in order for the communication event to be successful. And since they are constituted by concepts, it is necessary (but not sufficient) that the receiver and emitter share a grasp of the concepts that constitute the higher-level entity. In other words, higher-level entities and their message cannot be transmitted without mobilizing their constituting concepts.

Thus, if we are faced with a message, we are faced not only with the concepts explicitly mentioned or set on the centre stage, but also with concepts that play supporting roles, without which the message would be different. It stands to reason that these supporting roles are both essential to the message, and qualitatively different from centre-stage roles. Therefore, if we wish to draw a portrait of how a concept is being used, or of its overall function in language and behaviour in general, our portrait of it should account for supporting roles as much as the more glamorous ones. Therefore, our task when detecting concept use is to get all the concepts constituting the higher-level discourse entities that structure the text, be they centre-stage or not, explicit or not.

In normal conditions, readers should be able to pick out concepts even when they are implicit and play supporting roles, because the understanding of higher-level discourse entities depends on it. However, given how language purposefully draws our attention to centre-stage concepts, this demands that efforts be made in order to get our attention at the right place. In an annotation protocol, this means coming up with devices to force the annotator to focus on supporting role concepts. For example, Chartrand, Cheung, and Bouguessa (2017) came up with a two-step annotation process. The first step is meant to build a pool of concepts that can be used in a second step that is designed to limit the bias against supporting role concepts: a text segment is presented to an annotator, and she is tasked with providing five concepts that she deems to be present in the text segment. In the second step, on the one hand, the annotator is given a concept and a text segment, so that she cannot discriminate in favour of centre-stage concepts. On the other hand, annotators are asked for the concept's *degree* of presence—this way, they can express that a concept is not centre-stage without being tempted to express it by marking the concept as absent.

So, this is how our concept of CONCEPT translated into annotation protocols, but can we leverage it for automatic processing? As we have alluded in the previous section, millikanian linguistics offer a natural ground for distributional semantics, which can in turn be used to make indirect representations of concepts' proper functions.

Firstly, it useful to explain two fundamental concepts in distributional semantics: *syntagmatic* and *paradigmatic* relations (cf. Sahlgren 2008).

When words are *syntagmatically* related, it usually means that they are encountered in the same documents, text segments or sentences—in other words, syntagmatically related words co-occur a lot. This is significant because, while full sentences don't repeat themselves in a corpus, people usually use about the same words to talk about the same things, and discourse about a topic is something that is typically repeated. Thus, co-occurring words are typically found in texts which are thematically similar (they talk about the same things). Therefore, syntagmatic similarity between two words signals unity in the themes we can express with those two words.

On the other hand, *paradigmatically* related words are words that cooccur with

the words. Typically, paradigmatically related words can play similar roles in the same clauses: replacing one by another might change the meaning, but it will still mean something, and, often, this will form another sentence that is susceptible to be found in the corpus. Paradigmatic relations thus approximate relations of synonymy, with the caveat that antonyms are usually paradigmatically very close, given that they differ only on one dimension and that they will play similar roles in sentences.

Syntagmatic and paradigmatic relations have been leveraged in various way by researchers in natural language processing. Syntagmatic relations are often leveraged in vector representations where words are represented by the documents where they occur, or vice-versa. Through clustering, we can get groups of documents or words that are thematically related, and groups of documents, although topic models are now more commonly used to represent thematic units which are linked to both words and documents. Syntagmatic relations are also used in a large variety of tasks, including automatic summarization, information retrieval (finding a document from a keyword query), recommendation engines, etc. Paradigmatic relations are leveraged to make vector representations where words are represented through other words in terms of their propensity to co-occur, typically within a very short window. On top of finding synonyms, these representations can be used for tasks that involve word composition, for word-sense disambiguation (disambiguating different meanings or sense for a single morpheme), to enhance some topic models (it is particularly useful for inferring topics for short texts, like twitter statuses), for language models (e.g. predicting which will be the next word), etc.

Coming back to millikanian linguistics, higher-level entities tend to reproduce themselves (not as exact copies, as words do, but rather in the same fashion as sentences reproduce themselves), for the same reason any linguistic device does: they successfully serve a purpose in the social and discursive landscape where they are enacted. Now, there are a large variety of factors affecting word use—simply overhearing someone use a word in a conversation nearby certainly makes us more likely to reuse this word. However, *ceteris paribus*, that we are expressing a certain narrative or story, for example, will strongly determine the words we will use to express it. This is partly because the concepts that constitute them condition lexicon by way of favouring words that can be used to express them, but also because playing a certain role in discourse is more readily achieved using some types of words rather than others. For example, while they may be argued for the same conclusions about the same themes, racist discourses from far-right extremists and from mainstream conservative politicians usually will not share the same vocabulary, because they are not staged in the same settings (Van Dijk 1993).

This association of a recurring vocabulary to recurring higher-level entities could at least partly explain the phenomenon of syntagmatic relations in distributional semantics (Sahlgren 2008). Two documents, two sentences, or two text segments are similar to the degree that they share the same words. This may very

well be because, as higher-level entities condition vocabulary, sharing the same vocabulary indicates a common involvement in expressing the same higher-level discourse entity. Thus, *ceteris paribus*, similar vocabulary means shared involvement in the same higher-level entities.

These syntagmatic relations can then be leveraged by topic models (e.g. Blei, Ng, and Jordan 2003) and clustering algorithms (Meunier, Biskri, and Forest 2005) to find clusters of textual units that can be read as expressing the same higher-level entity. In other words, higher-level entities are traceable in the text because of the way they condition it and because of the lexical trace that they leave.

Conversely, concepts, words and other linguistic devices that participate in constructing higher-level discourse entities can be described in terms of the other linguistic devices that participate in the same higher-level entities. This is because variations in higher-level entities typically conserve the same discursive forms, and these discursive forms select their constituent by their broad functions. For instance, in “Long live the King!”, “King” can easily be replaced with “Queen”, as they have similar functions. To a lesser degree, the same can be said of any figure or entity that has a strong authority. Thus, association with “Long live” in a large corpus might indicate that the word “king” and “queen” can fulfill the same function of being the object of approval as an authority. Given that this form can take variations, “king” and “queen” might also be associated with “Down with”, which would indicate that they can also both function as objects of disapproval as authority. And so on with other variations, and other discursive forms.

As these cooccurrences accumulate, we can have a decent portrayal of the propensity of two linguistic devices to embody the same functions: this would correspond to the paradigmatic relations in distributional semantics (Sahlgren 2008). This explains why counting word cooccurrences in large generalist corpora is such a good indicator of synonymy, as well as various semantic properties of representations made this way (cf. Mikolov, Chen, et al. 2013; Mikolov, Sutskever, et al. 2013; Pennington, Socher, and Manning 2014).

Now, if the repetition of higher-level entities can predict word distributions, then, conversely, from word distributions, we can infer, at the very least, the probable presence of a higher-level entity. This is essentially how probabilistic topic model work: topics are higher-level structures that are inferred to explain word distributions. And if we assume that topics are composed of concepts, and that their influence over word distributions is a function of the concepts they are composed with, then we can use these word distributions not only to identify which topics are present in a document, but also which concepts compose these topics. This is the hypothesis that is followed in Chartrand, Cheung, and Bouguessa (2017) and Chartrand (2019), using different models of both concepts and topics.

Similarly, paradigmatic relations ensure that at least words can be represented

in another way, through their close neighbours, which, as we mentioned, can act as a proxy for function. However, we know that not only words can fruitfully be represented on these vector spaces, but new meaningful vectors can be constructed from word vectors. Furthermore, other entities can be usefully modelled in the same vector space as word embeddings, as shown by the success of algorithms like *doc2vec* (Lau and Baldwin 2016), which represents documents as vectors, *sense2vec* (Trask, Michalak, and Liu 2015) which deals with word sense and *LCTM* (Hu and Tsujii 2016), which uses a vector model for concepts to construct topics. Therefore, there are good reasons to be optimistic concerning the modellization of paradigmatic relations for concepts using distributional semantics.

Thus, distributional semantics offers two ways of getting at concepts by observing its traces in textual corpora: through the higher-level entities that they construct, and through the modellization of functional similarity through the space of paradigmatic relations.

In this light, it appears that accepting concept similarity by function actually quite compatible with the main insights of distributional semantics. This, in turns, opens up the possibility of using technology built on these foundations for detecting concept presence automatically. As Chartrand et al. (2016) and Chartrand, Cheung, and Bouguessa (2017) have shown, we can use syntagmatic relations to find higher-level entities, which will in turn tell us where their constituting concepts are likely to be present. And as Chartrand (2019) suggests, recent progresses on paradigmatic relations might enable us to determine which concepts constitute these higher-level entities. Furthermore, as works in computer-assisted reading and conceptual analysis of text suggest (Meunier, Biskri, and Forest 2005; Chartier et al. 2008; Sainte-Marie et al. 2011; Le et al. 2016) these techniques can also be leveraged for other aspects of corpus-based conceptual analysis, like for representing certain aspects of the target concept.

Conclusion

In this article, two main objectives were sought. On the one hand, I pursued the general goal of providing a theoretical framework for corpus-based conceptual analysis. On the other hand, this general, operative objective was pursued through a question: which account of similarity is best adapted to answer the challenges of corpus-based conceptual analysis?

In the first section, I sought to give a general picture of how an empirically-based conceptual analysis might be conceived and theorized. I did this through an overview of the leading accounts of conceptual analyses as method or philosophical endeavour in analytic philosophy—in particular, I reviewed Carnapian explication, Haslangerian analysis, and some perspectives on the method of cases. I noted that the contributions of these accounts are largely complementary,

and used this insight to distill these contributions into a general account of corpus-based conceptual analysis as a method.

This first section acts as a sort of introduction for the second section, whereby the hermeneutical resources necessary for formulating the problem of the second section are presented and put together. The second section builds on this, first by formulating its driving question—which account of conceptual similarity is best adapted to corpus-based conceptual analysis? Given how corpus-based conceptual analysis has a foot in conceptual analysis and another in observation of concepts in corpora, this question plays on two very different contexts: the application of the similarity criterion to evaluate the target concept (or *explicatum*, to use Carnap’s terminology) on the one hand, and the distinction between the concepts’ representation in corpora on the other. In both contexts, I argued that concepts’ similarity should be based on resemblance of their proper functions rather than comparison of extensions or intensions. To do so, I overviewed Millikan’s (1984) system of concept and adapted her notion of WORD in order to account for and describe concepts as they are studied by empirically based conceptual analyses.

Finally, in order to illustrate how the framework developed in section contributes to the more general objective of providing theoretical grounds for corpus-based conceptual analysis, I showed in section how the millikanian framework and concept similarity as function can be leveraged to operationalize concept presence detection. This was done both for concept presence detection as performed by a human and by a computer algorithm.

As such, my contribution is threefold: (1) I provided an argumentation in favour of assessing concept similarity by way of comparing proper functions, (2) I provided a framework that formulates accounts of concept and conceptual analysis for corpus-based conceptual analysis, and (3), I illustrated how this framework is leveraged in concept presence detection.

On the one hand, this provides a theoretical basis that justifies both the way the concept presence detection problem is formulated in Chartrand, Cheung, and Bouguessa (2017) and the way annotations were performed. Furthermore, it helps us better formulate the assumptions behind the algorithmic approaches defended by Chartrand, Cheung, and Bouguessa (2017) and Chartrand (2019), and lends some support to it.

While some work has been done to promote corpus-based conceptual analysis (Bluhm 2013; Andow 2016), it still represents a new way of approaching philosophical method within analytic philosophy, and, as such, disposes of very few hermeneutical resources to account for itself. One can only hope that the work presented in this paper can contribute to addressing this want—and perhaps inspire further work in this direction.

References

- Alexander, Joshua, and Jonathan M. Weinberg. 2007. “Analytic Epistemology and Experimental Philosophy.” *Philosophy Compass* 2 (1):56–80.
- Andow, James. 2016. “Qualitative tools and experimental philosophy.” *Philosophical Psychology* 29 (8):1128–41.
- Bealer, George. 1998. “Intuition and the Autonomy of Philosophy.” In *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, edited by Michael DePaul and William Ramsey, 201–40. Rowman & Littlefield.
- Bealer, George, and Peter Frederick Strawson. 1992. “The incoherence of empiricism.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 66. JSTOR:99–143.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of machine Learning research* 3 (Jan):993–1022.
- Bluhm, Roland. 2013. “Don’t Ask, Look! Linguistic Corpora as a Tool for Conceptual Analysis.” In *Was dürfen wir glauben?: Was sollen wir tun? Sektionsbeiträge des achten internationalen Kongresses der Gesellschaft für Analytische Philosophie e.V.*, edited by Migue Hoeltje, Thomas Spitzley, and Wolfgang Spohn, 7–15. DuEPublico.
- BonJour, Laurence. 2006. “Kornblith on Knowledge and Epistemology.” *Philosophical Studies* 127 (2):317–35.
- Braddon-Mitchell, David. 2009. “Naturalistic Analysis and the a Priori.” In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola. MIT Press.
- Brun, Georg. 2016. “Explication as a Method of Conceptual Re-Engineering.” *Erkenntnis* 81 (6):1211–41.
- Cappelen, Herman. 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.
- Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.
- . 1963. “The Philosophy of Rudolf Carnap.” In *The Philosophy of Rudolf Carnap*, 859–1013. La Salle: Open Court.
- Carus, A. W. 2008. *Carnap and Twentieth-Century Thought: Explication as Enlightenment*. Cambridge: Cambridge University Press.
- Chalmers, David J. 2014. “Intuitions in Philosophy: A Minimal Defense.” *Philosophical Studies* 171 (3):535–44.
- Chalmers, David J., and Frank Jackson. 2001. “Conceptual Analysis and

Reductive Explanation.” *Philosophical Review* 110 (3):315–61.

Chartier, J. F., J. G. Meunier, J. Danis, and M. Jendoubi. 2008. “Le travail conceptuel collectif: une analyse assistée par ordinateur du concept d’ACCOMMODEMENT RAISONNABLE dans les journaux québécois.” *Actes des JADT 2008*, 297–307.

Chartrand, Louis. 2019. “Mixing syntagmatic and paradigmatic information for concept detection.” *arXiv e-prints*, {Apr}, –1904.

Chartrand, Louis, Jackie C. K. Cheung, and Mohamed Bouguessa. 2017. “Detecting Large Concept Extensions for Conceptual Analysis.” In *Machine Learning and Data Mining in Pattern Recognition*, 78–90. Lecture Notes in Computer Science. Springer, Cham.

Chartrand, Louis, Jean-Guy Meunier, Davide Pulizzotto, José López González, Jean-François Chartier, Ngoc Tan Le, Francis Lareau, and Julian Trujillo Amaya. 2016. “CoFiH: A heuristic for concept discovery in computer-assisted conceptual analysis.” In *JADT 2016 : 13ème Journées internationales d’Analyse statistique des Données Textuelles*. Vol. 1. Nice, France.

Clark, A., and D. Chalmers. 1998. “The extended mind.” *Analysis* 58 (1):7.

Clark, Andy. 2008. *Supersizing the mind*. Oxford University Press.

Deutsch, Max Emil. 2015. *The myth of the intuitive: Experimental philosophy and philosophical method*. MIT Press.

Dutilh Novaes, Catarina. 2018. “Carnapian explication and ameliorative analysis: a systematic comparison.” *Synthese*, 1–24.

Dutilh Novaes, Catarina, and Erich Reck. 2017. “Carnapian Explication, Formalisms as Cognitive Tools, and the Paradox of Adequate Formalization.” *Synthese* 194 (1):195–215.

Evans, Vyvyan. 2006. “Lexical Concepts, Cognitive Models and Meaning-Construction.” *Cognitive Linguistics* 17 (4).

Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Gettier, Edmund L. 1963. “Is justified true belief knowledge?” *Analysis* 23 (6):121–23.

Goldman, Alvin. 2005. “Kornblith’s Naturalistic Epistemology.” *Philosophy and Phenomenological Research* 71 (2):403–10.

Goldman, Alvin I. 1986. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

———. 2007. “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status.” *Grazer Philosophische Studien* 74 (1):1–26.

Gopnik, Alison, and Eric Schwitzgebel. 1998. “Whose Concepts Are They,

- Anyway?: The Role of Philosophical Intuition in Empirical Psychology.” In *Rethinking Intuition*, edited by M. R. DePaul and William Ramsey, 75–91. Lanham: Rowman and Littlefield.
- Harnad, Stevan. 2009. “Concepts: The Very Idea.” Canadian Philosophical Association Symposium on Machery on Doing without Concepts. {<http://eprints.soton.ac.uk/268029/>}.
- Haslanger, Sally. 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press.
- Hu, Weihua, and Jun’ichi Tsujii. 2016. “A latent concept topic model for robust topic inference using word embeddings.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2:380–86.
- Hurley, Susan L. 1998. “Vehicles, Contents, Conceptual Structure and Externalism.” *Analysis* 58 (1). Oxford University Press:1–6.
- Ichikawa, Jonathan Jenkins. 2009. “Intuitions and Begging the Question.”
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, Magdalena Balcerak. 2013. “Conceptual Analysis and Epistemic Progress.” *Synthese* 190 (15):3053–74.
- King, Jeffrey C. 1998. “What is a Philosophical Analysis?” *Philosophical Studies* 90 (2):155–79.
- Knobe, J. 2003. “Intentional Action and Side Effects in Ordinary Language.” *Analysis* 63 (3):190–94.
- Knobe, Joshua. 2016. “Experimental Philosophy is Cognitive Science.” In *A Companion to Experimental Philosophy*, edited by Justin Sytsma and Wesley Buckwalter. Blackwell.
- Koch, Steffen. 2019. “Carnapian Explications, Experimental Philosophy, and Fruitful Concepts.” *Inquiry: An Interdisciplinary Journal of Philosophy* 0 (0):1–18. <https://doi.org/10.1080/0020174X.2019.1567381>.
- Kornblith, Hilary, and others. 2002. *Knowledge and its Place in Nature*. Oxford: Oxford University Press.
- Lau, Jey Han, and Timothy Baldwin. 2016. “An empirical evaluation of doc2vec with practical insights into document embedding generation.” *arXiv preprint arXiv:1607.05368*.
- Le, Ngoc Tan, Jean-Guy Meunier, Louis Chartrand, Davide Pulizzotto, Jose A. Lopez, Francis Lareau, and Jean-François Chartier. 2016. “Nouvelle méthode d’analyse syntactico-sémantique profonde dans la lecture et l’analyse de

- textes assistées par ordinateur (LATAO).” In *JADT 2016 : 13ème Journées internationales d’Analyse statistique des Données Textuelles*. Vol. 2.
- Levenshtein, V. I. 1966. “Binary Codes Capable of Correcting Deletions, Insertions and Reversals.” *Soviet Physics Doklady* 10 (feb):707.
- Lewis, David. 1970. “How to Define Theoretical Terms.” *Journal of Philosophy* 67 (13):427–46.
- Ludwig, Kirk. 2007. “The Epistemology of Thought Experiments : First Person Versus Third Person Approaches.” In *Midwest Studies in Philosophy*, edited by Peter A. French and Howard K. Wettstein, 128–59. Blackwell.
- Machery, Edouard. 2009. *Doing without Concepts*. Oxford University Press.
- . 2017. *Philosophy Within its Proper Bounds*. Oxford University Press.
- Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen P. Stich. 2004. “Semantics, Cross-Cultural Style.” *Cognition* 92 (3):1–12.
- Menary, Richard. 2007. *Cognitive Integration: Mind and Cognition Unbounded*. Palgrave-Macmillan.
- Menger, Karl. 1943. “What is dimension?” *The American Mathematical Monthly* 50 (1):2–7.
- Meunier, Jean Guy, Ismail Biskri, and Dominic Forest. 2005. “Classification and categorization in computer assisted reading and analysis of texts.” In *Handbook of categorization in cognitive science*, edited by Henri Cohen and Claire Lefebvre, 955–78. Elsevier.
- Meunier, Jean-Guy. 2017. “Theories and Models: Realism and Objectivity in Cognitive Science: Objectivity and Truth in Science.” In, edited by Evandro Agazzi, 331–52. Cham: Springer International Publishing.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient estimation of word representations in vector space.” *arXiv preprint arXiv:1301.3781*.
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. “Distributed representations of words and phrases and their compositionality.” In *Advances in neural information processing systems*, 3111–9.
- Millikan, Ruth Garrett. 1984. *Language, thought, and other biological categories: New foundations for realism*. MIT press.
- . 1998. “A common structure for concepts of individuals, stuffs, and real kinds: More Mama, more milk, and more mouse.” *Behavioral and Brain Sciences* 21 (1):55–65.
- . 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford University Press.
- Minsky, Marvin. 1975. “A framework for representing knowledge.” In *The psychology of computer vision*, edited by P. Winston, 211–77. New-York: McGraw-

Hill.

Murphy, Gregory. 2004. *The big book of concepts*. Cambridge, MA: MIT press.

Murphy, Taylor. 2014. “Experimental Philosophy: 1935-1965.” In *Oxford Studies in Experimental Philosophy*, edited by Tania Lombrozo, Joshua Knobe, and Shaun Nichols, 1–325. Oxford University Press.

Nado, Jennifer. 2016. “The intuition deniers.” *Philosophical Studies* 173 (3):781–800.

Naess, Arne. 1938. “*Truth*” as Conceived by Those who are Not Professional Philosophers. Edited by Jacob Dybwad. Oslo.

Nolan, Daniel. 2009. “Platitudes and Metaphysics.” In *Conceptual Analysis and Philosophical Naturalism*, edited by David Braddon-Mitchell and Robert Nola. MIT Press.

Nolfi, Kate. 2016. “Epistemically Flawless False Beliefs.” Montréal, Québec: Fillosophie.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. “Glove: Global vectors for word representation.” In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–43.

Pettit, Dean, and Joshua Knobe. 2009. “The pervasive impact of moral judgment.” *Mind & Language* 24 (5):586–604.

Pinder, Mark. 2017. “Does Experimental Philosophy Have a Role to Play in Carnapian Explication?” *Ratio* 30 (4):443–61.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. Oxford: Oxford University Press.

Pohlhaus, Gaile. 2015. “Different Voices, Perfect Storms, and Asking Grandma What She Thinks: Situating Experimental Philosophy in Relation to Feminist Philosophy.” *Feminist Philosophy Quarterly* 1 (1).

Pulizzotto, Davide, José A. Lopez, Jean-François Chartier Jean-Guy, Meunier1 Louis Chartrand1 Francis Lareau Tan, and Le Ngoc. 2016. “Recherche de «périsegments» dans un contexte d’analyse conceptuelle assistée par ordinateur: le concept d’«esprit» chez Peirce.” In *JEP-TALN-RECITAL 2016*. Vol. 2. Paris.

Quine, Willard V. 1971. “Epistemology naturalized.” *Akten des XIV. Internationalen Kongresses für Philosophie* 6:87–103.

Rey, Georges. 2018. “The Analytic/Synthetic Distinction.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University.

Rosch, Eleanor H. 1973. “Natural categories.” *Cognitive psychology* 4 (3):328–50.

Rudolf, Carnap. 1928. *Der logische Aufbau der Welt*. Berlin-Schlachtensee:

Weltkreis Verlag.

Rupert, R. D. 2009. *Cognitive systems and the extended mind*. Oxford University Press, USA.

Rysiew, Patrick. 2017. "Naturalism in Epistemology." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University.

Sahlgren, Magnus. 2008. "The distributional hypothesis." *Italian Journal of Disability Studies* 20:33–53.

Sainte-Marie, Maxime B., Jean-Guy Meunier, Nicolas Payette, and Jean-François Chartier. 2011. "The concept of evolution in the Origin of Species: a computer-assisted analysis." *Literary and linguistic computing* 26 (3):329–34.

Schwartz, Peter H. 1999. "Proper Function and Recent Selection." *Philosophy of Science* 66 (3):210–22.

Searle, John R. 1958. "Proper names." *Mind* 67 (266):166–73.

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

Sosa, Ernest. 2007. "Experimental Philosophy and Philosophical Intuition." *Philosophical Studies* 132 (1). Springer:99–107.

Stalnaker, Robert. 1976. "Propositions." *Issues in the Philosophy of Language*, 79–91.

Swain, Stacey, Joshua Alexander, and Jonathan Weinberg. 2008. "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp." *Philosophy and Phenomenological Research* 76 (1):138–55.

Trask, Andrew, Phil Michalak, and John Liu. 2015. "sense2vec-A fast and accurate method for word sense disambiguation in neural word embeddings." *arXiv preprint arXiv:1511.06388*.

Van Dijk, Teun Adrianus. 1993. "Principles of critical discourse analysis." *Discourse & society* 4 (2):249–83.

Von Eckardt, Barbara. 1995. *What is cognitive science?* Cambridge, MA: MIT press.

Williamson, Timothy. 2008. *The Philosophy of Philosophy*. Malden, MA: Wiley-Blackwell.

———. 2013. "How Deep is the Distinction Between A Priori and A Posteriori Knowledge?" In *The A Priori in Philosophy*, edited by Albert Casullo and Joshua C. Thurow, 291–312. Oxford University Press.

Wisdom, Jeffrey. 2017. "Proper Function Moral Realism." *European Journal of Philosophy* 25 (4):1660–74.