What can philosophers really learn from science journals?

Oliver M. Lean[*], Luca Rivelli[*], and Charles H. Pence[*†]

draft of May 11, 2020

[*] Université catholique de Louvain, Institut supérieur de philosophie, Place du Cardinal Mercier 14, bte. L3.06.01, 1348 Louvain-la-Neuve, Belgium
[†] to whom correspondence should be addressed; charles@charlespence.net

Abstract

Philosophers of science regularly use scientific publications in their research. To make their analyses of the literature more thorough, some have begun to use computational methods from the digital humanities (DH). Yet this creates a tension: it's become a truism in science studies that the contents of scientific publications do not accurately reflect the complex realities of scientific investigation. In this paper, we outline existing views on how scientific publications fit into the broader picture of science as a system of practices, and find that none of these views exclude articles as valuable sources for philosophical inquiry. Far from ignoring the gap between texts and practice, proper use of DH tools requires, and can even contribute to, our understanding of that gap and its implications.

Keywords: empirical philosophy; digital humanities; big data; scientific literature; scientific practice; publication

1. Introduction: The Gap

According to many philosophers, we have inherited a largely inadequate view of science – one that is in many ways distorted, and generally out of touch with its rich and varied realities. Criticisms of this kind drove what is often called the "practice turn" (Soler et al. 2014), which in various ways drew attention to the complexities of science as it is in practice, rather than the idealized models proposed in philosophy. As a result, philosophers now increasingly make use of empirical methods such as ethnographic techniques (Nersessian 1995; Leonelli 2016; Soler et al. 2014). The hope is that our philosophical claims about

science may be more legitimate if they are based, at least in part, on such empirical methodology.

In recent years, some philosophers have aimed to apply this empirical spirit to the way they approach scientific literature. While philosophers have always used scientific writings in their work for a wide variety of purposes, it is all too easy to go about this in a way that biases one's conclusions from the start. Indeed, one criticism from the practice turn is that philosophy has based its generalizations about science on idealized examples, cherry-picked from the literature. Given the massive and increasing volume of scientific publications, however, it has been hard to see an alternative that could yield results that are both general and reliable. How can we analyze enough of the literature for our claims about it to have broad validity, and without biasing our conclusions through our search methods?

Recently, philosophers have aimed to remedy this by adapting a range of tools from the digital humanities (DH). Among other methods, DH involves the use of computational techniques to analyze large collections (*corpora*) of documents in order to detect features that would evade unaided examination. These techniques can be deployed in a variety of ways: to detect aspects of language use or style in a corpus (Burrows 2002; Hoover 2010), to construct networks of relationships and interactions between authors (Börner, Maru, and Goldstone 2004; Abbasi, Altmann, and Hossain 2011), or to track large-scale differences between different disciplines (Dias et al. 2018) or conceptual change over time (Chavalarias and Cointet 2013), to name just a few. There is great promise in utilizing these tools to satisfy philosophy of science's increasing interest in empirical rigor, objectivity and scope.

However, there is an apparent tension created by the application of DH methods to empirical philosophy. It is generally acknowledged that there are major disparities between

what is presented in the scientific literature and what actually happens in the process of scientific investigation and discovery. It's been argued that the version of events depicted in research articles – of what was done, how, why, in what order, and so on – is a polished, idealized, and generally inaccurate version of what actually happens in the laboratory or in the field (Medawar 1963; Nickles 1988; Rouse 1990; Schickore 2008). It is partly for this very reason that philosophers of the practice turn have moved away from a reliance on readings of the literature and gone "behind the scenes" to grapple with the complex realities that are left out of the written record. Because of this view of publications, philosophers of scientific practice may regard developments in digital literature analysis with suspicion: the disconnect between the content of scientific publications and actual practice (which for brevity we simply call *the gap*) might completely undermine the very idea of analyzing literature in practice-based philosophy. If such a gap exists, then trying to learn anything about "actual" science from the literature – digitally or otherwise – looks to be a lost cause.

Our aim here is to address this issue, by helping to clarify the nature of the problem, to resolve the tensions it creates, and thus to clear the way for the application of digital corpus methods for philosophical purposes. In doing so, however, our investigation leads us to consider more generally the role of analyses of scientific literature in the service of philosophical goals, whether by digital means or by more traditional methods. Therefore, we hope that our discussion can be of general relevance to all philosophers of science who make use of scientific literature in their research.

2. From Scientific Literature to Philosophical Claims: Clarifying the Problem

Discussion about whether and how post-practice turn philosophers should use scientific literature will not be fruitful without a clear picture of the moving parts in the matter. Our first task, then, is to break the question down into smaller, more philosophically tractable elements. To that end, this section is in effect an extended outline of the paper's structure: it offers a top-down view of our attempt to decompose the problem into parts, each of which is dealt with in order in subsequent sections.

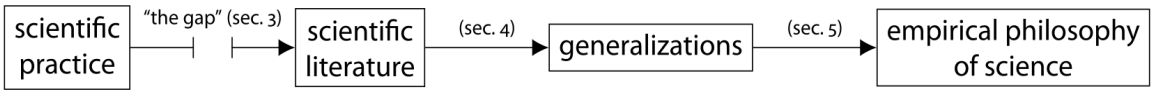

Figure 1. The structure of three related questions about the use of scientific literature in the empirical philosophy of science. Boxes are connected by arrows which indicate that the latter object's content is derived from observation of the content of the former. The first connection, the production of scientific literature from practice, denotes more narrowly what we have called "the gap," while accounts of all three arrows should be considered in order to fully justify philosophers' use of scientific papers.

Our proposed framework for addressing these issues is represented in figure 1, which shows a series of productive or inferential steps, each of which can be dealt with semi-independently. The first step concerns clarifying just what constitutes the gap between scientific literature and practice. It is one thing to acknowledge that scientific publications do not faithfully depict the realities of science, and quite another to spell out what this means. When it is claimed that a journal article doesn't accurately reflect scientific practice, what exactly is changed, added and omitted? With notable exceptions that we will discuss, little has been done to make this question precise enough to illuminate our questions about philosophical methodology. A starting point for this first step is the near-trivial claim that scientific publications are themselves a product of scientific practice. They share this with many of the other aspects of science that are of direct interest to philosophers of scientific

practice: experiments, data, laboratory manuals, and so on. Theories about the gap, then, are in effect theories about publication practices and their implications for the epistemic value of their products. This suggests a different way to frame our questions. How exactly does scientific practice give rise to journal articles? What is the role of publication practices in the doing of science? For which audiences and goals (both epistemic and social) are they prepared? Beginning to develop the problem in this direction constitutes the subject of section 3.

Suppose now that we have an answer to these first questions – that we have an account according to which the journal literature is itself a product of scientific practice with a well-theorized relationship to science as a whole, and hence an idea of how we can find out what is going on in scientific practice from what is going on in the literature. Even so, we are left with the question of how to find out what is going on in the literature in the first place. Again, given its sheer size, it is hard to see how to generate and test generalizations about the literature with sufficient rigor and scope. Of course, philosophers have long experience with methods of close reading and the construction of case studies from the scientific literature, though they are increasingly recognized to face a variety of limitations and problems (Mizrahi 2020). With the advent of the digitization of the journal literature, the tools of the digital humanities offer philosophers a different and potentially powerful perspective. However, there are different theoretical views of how the directly accessible features of the literature – its *syntax* – relates to the *semantic* content in which philosophers are generally interested. Importantly, different theoretical views about language and meaning and their relationship to corpora hold different implications for how to go about corpus-based analysis

of science. We outline some of these views and discuss their implications for philosophical research in section 4.

Finally, section 5 summarizes the discussion by tying together these various threads. Ultimately, our goal as philosophers is to establish how to approach the scientific literature in the service of specifically *philosophical* interests. Again, even with an answer to the first two questions – of how to make reliable generalizations about features of the literature, and of how those features relate to science-as-practice – there remains the question of what philosophers can do with all of that. To that end, we offer suggestions to this effect.

In short, the task before us is to discuss ways to think about the three arrows in figure 1. We will begin by carefully considering the manner in which scientific practice might be thought to give rise to journal articles, paying particular attention to various accounts of how the gap might arise or fail to arise. We then consider how we draw generalizations from that literature, in our case roughly equivalent to validating our use of the methods of the digital humanities in the philosophy of science. Finally, we will briefly consider in the conclusion what it is to construct "good philosophy of science" on the basis of these general claims about the literature.

3. Characterizing the Gap Between Literature and Practice

We argued in the last section that if we view scientific publications as a product of scientific practice, we can understand them in equivalent terms to other things that empirical philosophers study, namely, as sources of evidence whose significance to philosophers should be based on theoretical understanding of how that product is produced. With that in

mind, the next step is to explore alternative views of the relationship between publications and practice, what we have called "the gap."

We should begin by noting the obvious fact that scientists themselves make use of publications to keep in touch with the landscape of their field and inform their own practical work. So unless they are in some way deeply misguided in doing so, we can take for granted that the literature does carry content that is informative to scientific practitioners. Yet this does not mean that this content is transparent: properly understanding what is written may require considerable contextual knowledge. Indeed, scientists learn how to write and read publications through a combination of deliberate training and learned skill acquired from experience. What's more, we cannot simply take scientists at their word when they describe what they do and why. First, much of what one needs to know may be implicit and unconscious, and, second, we mustn't assume that the whole process works in the way it is understood to work. Hence the need for the third-person perspective of science studies, which aims to make explicit and critically analyze these practices. In short, even if the prospect of analyzing scientific texts philosophically from a practice perspective isn't futile, it is not simple.

To that end, this section considers several accounts of how and why scientific publications are produced, and how their content consequently departs in various ways from the reality of scientific practice. It is only with answers to these questions that we can say what, if anything, we as empirical philosophers can learn about scientific practice from the literature. As we'll see, although different views of this relationship have importantly different implications for how philosophers should interpret their content, all of them attribute a valuable role to scientific literature in some sense.

Importantly, these ways of understanding the gap are far from exhaustive. Despite appearances, they are also not necessarily mutually exclusive: scientific writings are multi-faceted, and some of the views only concern influences which affect certain aspects of the literature. And even where they seem to address the same aspect, more than one of the influences they describe may be at work, or they may apply in varying degrees to different disciplines or types of literature. Our sketch here should be seen as a toolkit of possible perspectives for philosophers to take in their work.

One final aside is worthy of note before we continue. We will assume in the following that the scientific articles at issue are produced in what we might call optimal working conditions – that is, that scientists are publishing all and only the papers that they want to be publishing, when they want to be publishing them. It is clear that even in these (impossible) conditions, there will still be a worry about the connection between articles and practice. Our discussion neglects, however, what we might call "ecosystem-level" effects of scientific publishing, such as file-drawer effect (the fact that negative, uninteresting, or unwanted results are not published), popularity of particular "hot" topics in high-impact journals, and so forth (see, e.g., Ioannidis 2005; Smaldino and McElreath 2016). While we believe that these are clearly important problems which, if pervasive, will make the connection between published articles and practice more tenuous, we lack the space to pursue them here.

We begin with views that deal specifically with scientific journal articles' portrayal of scientific reasoning and justification. This matter has been discussed at length by Schickore (2008) – to our knowledge the most detailed treatment to date of any dimension of the gap that considers its implications for philosophers. Schickore's focus is primarily on the gap's implications for epistemology, that is, on where philosophers should look to understand the

logic of scientific reasoning and the loci of scientific or theoretical knowledge. We recount here three broad interpretations of scientific justification in journal articles, which we will call the *logical, sociological,* and *fraud* theories.

The *logical* view is represented in various forms by Reichenbach (1938), Suppe (1998) and others. These authors assert that, in one way or another, journal articles contain what a discipline considers to be legitimate forms of justification for empirical claims. Reichenbach famously distinguished the context of discovery from the context of justification. Applying the same distinction in the domain of scientific publication, one might think of articles as being written in the "justificatory" mode, as vehicles by which scientists intend to justify their findings and conclusions to each other. The context of discovery – the process by which those conclusions were actually reached through theorizing and experiment – may not be represented in the finished product of research articles. Here, though, the gap offers us an important signal: it marks out the way in which scientists have prepared their findings for presentation, and is evidence of the logical standards by which scientists believe they will be judged. In that vein, Suppe argues that philosophers have so far failed in their task of explicating scientific reasoning precisely because our theories of this process do not capture what we find in the scientific literature.

The *sociological* view, as we call it, takes a less elevated view of the matter. This perspective is represented, for example, in the work of Karin Knorr-Cetina (1981) and Latour and Woolgar (1987), and emphasizes the social role of journal articles in scientific communities. According to Knorr-Cetina, scientific findings are not just produced but largely tested and credentialed in the lab, in ways that are messy, contextual, pragmatic, and opportunistic. The reconstruction performed in publication, in contrast, is intended to play a

kind of recruiting role: the aim there is to gain converts to the research program, fend off criticisms from hostile competing factions, and so on. On this view, journal articles contain neither the actual facts of discovery nor their actual justification, but simply attempts to gain followers. This is done in a particular social context involving the kind and degree of skepticism or hostility that the claims are likely to be met with. On this view, what we see in journal articles is to a large extent a kind of power struggle, and so the language within is to be interpreted at least partly in political rather than epistemic terms.

Finally, the *fraud* theory represents the most pessimistic view of journal articles and the reasoning presented in them. This view was notoriously taken by Peter Medawar in his "Is the scientific paper a fraud?" (1963). Articles, Medawar says, are written as though scientific inferences are operations of induction – deriving generalizations from passively and dispassionately acquired observational data. Yet this bears no resemblance to how scientists actually reason, which he understands as a Popperian hypothetico-deductive method of testing preconceived hypotheses. Medawar sees no useful purpose to this distortion, and proposes that scientists should instead write in ways that more truly reflect the reasoning they actually accept.

Each of these views holds different implications for what philosophers can infer from some observed feature of scientific literature. According to the logical view, for example, the argumentation we find in journals in fact captures what the scientists themselves consider valid reasons – perhaps even the strongest reasons – for accepting their conclusions, even if it fails to capture how the authors themselves reached them. If one's philosophical interest is in what scientific communities accept as valid justification, then of course an analysis of journal articles would be an acceptable means of finding this out. While there remain questions on

this view about which publications to use, and which parts (Schickore 2008), it remains the case that philosophically interesting material is to be found somewhere among them.

However, accepting the logical view does cast doubt on the value of journal articles for addressing other philosophical questions, including many that interest philosophers concerned with scientific practice. Since the practice turn, many philosophers now take direct critical interest in the strategies scientists use to investigate and manipulate their objects of study (Hacking 1982; Wimsatt 2007; Waters 2008; Baxter 2019). Many philosophers of science now view science as skillful problem-solving – the overcoming of various practical, material, and social challenges in order to produce knowledge. For the particular purpose of understanding scientific practices as solutions to these challenges, the literature is indeed likely to be of limited use, since much of this problem-solving occurs behind the scenes and isn't reported in the final presentation.

In contrast to the logical view, the sociological view suggests that we cannot, or not easily, discover what scientists accept as good scientific reasoning from analyzing papers. The encroachment of sociopolitical factors cautions us not to take any claim in a paper at face value, since it may be at least partly a rhetorical move rather than a dispassionate presentation of reasons. Such considerations have implications for other areas of interest to philosophers, such as the role and importance of certain kinds of language. For example, Lily Kay (2000) argues that talk of "information" in molecular biology took hold because of its rhetorical power rather than its importance to actual scientific investigation, most of which led to dead ends when guided by informational principles.

Claims of this sort are in effect claims about the gap between journals and practice as driven by the former's status as the "public face" of research programs. If true, they suggest

that we cannot safely infer the importance of a concept to scientific practice writ large from its prevalence in the literature. Of course, we can acknowledge the role of irrational or non-rational forces of social psychology at work in literature exchanges, while also supposing that scientists are at least sometimes or partly swayed by reason over rhetoric. In any case, this makes the task of discovering these motivations in scientific articles more complex.

Despite the complications this holds for the scientific epistemologist, the sociological view of publication practices also reveals a positive role for journal articles. For if social forces manifest themselves in scientific writing, we should be able to detect signals of these forces in the literature. For example, it may be that the amount and strength of rhetorical language an article uses reflects the degree of hostility that the authors expect towards their findings. Learning to detect these signals would, of course, also require cross-referencing with actual sociological observations. In any case, even the sociological view suggests that the language used in papers offers clues about something of value to philosophers interested in scientific practice; namely, the sociological context of the science and its development over time.

We move now to views about how scientific publications relate to other aspects of science. What we will call the *narrative* view focuses on the complexities surrounding the perceived significance of scientific findings. Variations on this view are represented in Rouse (1990) and Nickles (1988). Rouse argues that scientific practice depends essentially on a narrative. Like all purposeful activity, doing science requires a sense of how the past has led to the present and how one's actions might extend into the future. In other words, what a scientist or group of scientists does is always seen as a response to the situation they are currently in, including what they know and don't know, their capabilities, and so on. The

narrative is neither linear nor universal in a particular field; the view into both the past and future is constantly changing with new developments, and a field is characterized by competing narratives (e.g., of the "orthodoxy" and challengers to it).

This narrative view of science affords a very particular role for literature – namely, that it serves to feed the scientific community's sense of narrative that is necessary for science to exist at all. A journal article presents in its introduction a "state of play" consisting of previous work necessary for understanding its findings and their significance for future research. Crucially, this means that what is considered significant isn't fixed – results previously thought important can later be seen as ignoble, and vice versa (a fact discussed particularly by Nickles 1988).

What can we say about the use of scientific literature for philosophical purposes from the narrative perspective? Like the sociological view, conceiving of a scientific article as meaningful only in relation to a narrative certainly complicates the problem, particularly when it comes to judging whether an individual work or body of work is significant. This is something philosophers will be interested to know: for many purposes, the philosophical significance of scientific work will depend largely on its scientific significance. If this significance itself is only relative to a narrative, we must understand the narrative and, crucially, its evolution over time in order to make proper use of it in philosophical research. This diachronic view is especially needed if, with Nickles (1988, 37–38), we agree that changes in a narrative can lead to different ways of using and understanding concepts over time, even if they continue to be named by the same words. Nevertheless, given the important role that publications play in constituting and reconstituting the narratives at play in scientific

practice, they remain vital sources for interpreting those narratives and their change over time.

Finally, we should mention the well-developed literature critically discussing the role of *citations* in publication practices. The idealized view of citations is that they are supposed to indicate what previous works influenced the intellectual or practical development of the present work. But this ideal is not necessarily reflected in reality. MacRoberts and MacRoberts (1989), for example, highlight several shortcomings of real-world citation practice: influences can be omitted, citations are biased in various ways, especially in favor of the author's own work, citation norms vary between disciplines and across time, and so on.

These criticisms, while well known, have far-reaching implications, including for philosophers aiming to use scientific literature. This is especially true of the potential role in philosophy for digital methods such as the construction of citation networks, but applies equally to traditional methods of the reconstruction of scientific literatures, which often crucially rely upon article bibliographies. Suppose we are interested in a particular scientific finding or idea and its growth, development, and influence in or across scientific fields. The above suggests that this cannot be straightforwardly inferred from the prominence of the relevant publications in citation networks: that article B cites article A does not necessarily imply that article A provided significant practical or intellectual context for article B's development. Yet, again, accepting these criticisms needn't lead to a wholesale rejection of any attempt to quantify influence by citations. Instead, it calls for philosophers to be cautious, and elaborates on what they should be cautious about.[1]

1 We will discuss some of the ways in which citation networks could be used despite the presence of these critiques in section 4.2.

This section has outlined several ways of conceiving of the role of scientific literature within scientific practices as a whole, and the ways these various perspectives bear on what philosophers can learn from the literature and how. The key lesson is that these are markedly different views of what journal articles are written for, how they are read, what contributes to their uptake, and so on. It bears repeating that these alternatives are not mutually exclusive – each of them may simply capture distinct but equally important aspects of science and the role of publication practices within it. In any case, all these views reveal complexities to publication practices that we as philosophers mustn't ignore. The general lesson is simply that a scientific article should never be taken at face value. We should always understand a research paper in pragmatic terms, as a kind of speech act whose meaning depends on the role it plays within a larger social activity, which may vary dramatically from one case to another.

None of this means, of course, that papers themselves are useless to philosophers of scientific practice. Instead, our conclusion is simply that their interpretation requires an understanding of the larger social activity in which they occur. Crucially, there are alternative ways of conceptualizing this social activity, and so one's interpretation of the literature should always be tied to a particular conceptualization. This is no more or less true of digital methods of literature analysis in particular, as we now discuss.

4. Getting from Corpora to Generalizations About Science

The last section considered the ways in which we might understand the first arrow in figure 1, connecting scientific practice to corpora. We now turn our focus to the second arrow of that

diagram. Assuming that we have a particular body of scientific text, how can we extract from it general claims about what goes on in science? These generalizations, in turn (as we will discuss in the conclusion), often serve as either source material for, or crucial validations of, our philosophical claims about science itself.

As mentioned earlier, such a discussion could easily be targeted not only at the novel methods of digital humanities, but also at our classic use of close-reading and the way in which philosophers have traditionally converted scientific journal articles into "case studies" (though see Mizrahi 2020 for a critical discussion of this methodology). This, however, will not be our focus in this section. Rather, we will consider the use of computer-aided, corpus-driven methods to generate such claims. On the one hand, such tools are potentially an incredibly rich source of claims about science – able to avoid many classic problems of bias and cherry-picking that may render the use of particular case studies problematic (though these same worries remain for corpus analysis, albeit in different guises; see Allen and Murdock forthcoming). On the other hand, such methods might result in a spate of unjustified and facile conclusions, drawn unreflectively from the output of misapplied computer-based tools (a worry which has often been raised concerning the digital humanities; see, e.g., Pechenick, Danforth, and Dodds 2015). We thus need to provide a justification of how these new methods can produce a high-level understanding of the process and practice of science. For our purposes, we will consider two different approaches that could offer us insight into the content of scientific publications, as representative examples: the use of textual analysis and the construction of citation networks.

4.1. Textual Analysis

Some digital humanities methods offer us the promise of evaluating, at a large scale and across vast corpora of articles, the very meaning or conceptual structure of the articles at issue. If we could extract information about the use of a concept, for instance, from a large corpus, we could follow its usage across journals and across time, formulating and testing a variety of hypotheses about conceptual dynamics and theory change. Broadly, such methods are described as different sorts of *textual analysis*. Let's begin by considering a simple example, *co-occurrence analysis*. By measuring which words are found near others throughout the corpus (e.g., in the same paper, in the same paragraph, or in the same sentence), we can create a kind of weighted, undirected network that connects the terms within the corpus.

How does this network allow us to understand the meaning of texts? If we look at the structure of the network – in particular, the way in which it clusters into "communities," or subsets of terms within which terms are more densely connected to one another than to terms outside the community (Girvan and Newman 2002) – we find surprising possibilities for the automated extraction of novel knowledge from already existing data. To mention just one impressive case, Wilkinson and Huberman (2004), focusing on a network of co-occurrence of the names of genes in the abstracts of a corpus of biomedical literature, were able to show that detection of communities in that network could be used to automatically discover previously unknown functional modules of genes related to colon cancer. While these were "latently" documented (so to speak) in the extant literature, such functional modules had

escaped the attention of human researchers, due to the enormous number of publications which would have had to be considered in order to find them.[2]

For a more complex methodology, consider the use of *topic modeling* (well described for a philosophical audience in Malaterre, Chartier, and Pulizzotto 2019). Topic models break down documents into clusters of "topics," here a technical term, defined as a set of probability distributions across all words in the corpus. Such topics, then, can be interpreted as topics in the vernacular sense, subjects that each document is "about." Charting the changes in topic distribution in the corpus over time can give us an idea of evolving conceptual emphasis or focus, while examining the probability distributions that make up the topics themselves might help us understand the ways in which authors talk about different subjects in different corpora.

4.2. Citation Networks

Another of the oldest and most successful methods applied in the digital humanities, citation networks are networks of papers, where papers are related by the citation relation occurring between a paper and the ones that directly cite it. These networks can also be transformed into networks of citations between other kinds of entities, such as authors, journals or research institutions. Citation networks can also be constructed starting from indirect relations, such as co-citation (two entities are related when they are cited together by another entity; Small 1973) or bibliographic coupling (two entities are related when they both cite a

2 A similarly striking example can be found in Tshitoyan et al. (2019), who modeled a set of abstracts from articles in materials science. This model, they claim, selected materials that would have been eight times more likely to be discovered as thermoelectrics than a material chosen at random. In other words, latent information about future discoveries was found in the abstracts of the then-current literature.

third entity; Kessler 1963). Intuitively, while co-citation networks detect commonality of topic between related entities, as they connect entities that are found in the same bibliographies and thus serve as common sources for later works, bibliographic coupling could reveal latent or indirect commonality of interests between entities which do not cite one another but rely upon the same kinds of sources, a particularly interesting relation between authors, who might be unaware of these commonalities.

Several kinds of analysis can be performed on basic citation networks, the most common one being *community detection*, where the goal is to detect areas of mutual citation (groups of authors or other higher-level entities which all regularly cite one another) reflecting, again, shared interests. Once these communities are found, a coarse-grained version of the original network, in which each node corresponds not to a single paper (or author, etc.) but to a single community, can be produced, easing understanding of the network structure at a large scale by letting the observer see the field as a whole and avoid the information overload of a direct observation of the complex original network.

When applied to citation networks built from relationships other than direct citation, community detection could help reveal hidden patterns. In co-citation networks, communities should in general reveal topics grouping together entities which often “travel together” in later literature, while, in bibliographic coupling networks, the same technique could reveal communities of researchers with common interests, possibly in turn indicating ongoing latent processes of unification or splitting of research fields.

The topology of a citation network can shed light both on the general process by which scientific literature grows, and on the ways in which scientific information flows through the network during its dissemination. Citation networks tend in general to grow by

preferential attachment: the more a paper is cited, the more it gets cited subsequently. This mechanism gives rise to what is called a network with *scale-free* topology (Barabási and Albert 1999) – where a limited percentage of nodes form poles of attraction (the hubs) cited by many other nodes, with most other nodes considerably less cited. From an information-flow point of view, hyper-cited hubs are the primary sources of information for subsequent papers.

Specific dynamics of network evolution can be detected by analyzing sequences of snapshots of the citation network over time. Diachronic analysis, especially of coarse-grained networks, could reveal large-scale phenomena in the development of corpora of scientific literature, producing fruitful results both from an historical perspective – seeing when such processes have already occurred – and from a philosophical one. As an example of the latter, in a network of communities of authors who cite the same common entities, the real nature and scope of a debate could be clarified by studying the "latent" extension, both in space and in time, of the communities which fuel it.

Scientific papers can, of course, be further classified into kinds (research articles, review articles, letters, etc.), each with a typical range of uses. Diachronic reconstruction of the movement of ideas through citation networks – for example, from the initial announcement in the form of a letter, to the appearance of the polished results in the research paper, to the subsequent inclusion in a review paper and then in further research papers – could help model the ever-changing narrative landscape of science stressed by authors like Nickles (1988).

When modeling the evolution of citation networks at shorter temporal scales, machine learning techniques (as envisioned in Shibata, Kajikawa, and Sakata 2012) could be applied

to predict which papers will most likely be cited at a later time. This is a way of modeling the dynamics of an author in the process of choosing which papers to cite when writing a new paper, with the potential to reveal in general how the biases induced by the structure of the extant corpus of literature shape its subsequent growth through their influence on the practice of scientific writing. Thus biases, far from being purely negative, can be viewed as useful data for modeling the production of papers: biases in the literature influence the (mis)perception of the information conveyed to the scientists by the literature itself, providing philosophers with information about how to start tackling the "gap" problem.

As mentioned above, citation networks have received their fair share of criticism (MacRoberts and MacRoberts 1989; but see Wouters 2007). In part, these critiques arise as a result of the fact that citations are objects with meanings, just like any other element of a scientific text: they occur in a certain context, and express the possibly biased intentions of an author when referencing other papers. There is thus a distinction between syntax and semantics at work in the analysis of citations, just as there is in the textual analysis of scientific concepts. We must, of course, be precise about just what it is that we hope to analyze. If our object is the citation network as a whole, then we are not directly interpreting the intent of an author in citing a particular paper – this network is a product of those aggregated intentional behaviors, as performed by all of the writers. It can then be defended on statistical grounds (e.g., van Raan 2004) that when working with large datasets of reasonably highly cited papers, sets of citations get their random semantic biases averaged out, and in this sense they can be treated in a purely formal way, without regard to their semantic content. This runs parallel to the consideration that while semantics is likely

significant for the theory of the production of single papers, it can be reasonably ignored when algorithmically analyzing certain features of an entire corpus.

This line of reasoning alone seems to provide a sufficient rationale for advocating that analysis of patterns in large citation networks can still offer us a variety of insights into the structure and evolution of the scientific community. It is even likely that citation networks themselves will be indispensable for studying the disconnect between the varied and possibly biased intentions of authors driving their use of references and the "objective" influence of articles on one another, as the citation network should indirectly retain, in its structural and statistical features, information about the intentions behind the formation of references.[3]

4.3. Analyzing Meaning

In all these methods, whether analyzing text or citation networks, an automated algorithm, with access only to the syntax of the corpus and no access to the meaning of the terms within, reveals general or high-level structure and regularity in that syntax, which is at least interpreted by us to suggest its meaning. This offers us yet another justificatory challenge: how should we understand the jump that is made here from data like frequencies of terms within documents or citation relations to the meaning of texts? Just as inferences from what's said in the literature to scientific practice are tied to theories about the gap between them, inferring meaning from text depends on theories about the relationship between syntactical structure and semantics.

3 The most trivial example is the typical scale-free topology of citation networks, which reveals a pervasive bias in producing references.

This is, of course, by no means the first time that such a question has been broached in the literature. "How can we get from syntax to semantics?" is a refrain perhaps most commonly found within the literature on corpus linguistics, where the entire *raison d'être* of the field is, in some sense, tied up in an answer to this question. Unfortunately, unlike traditional philosophy of language, philosophical approaches to linguistics are still somewhat in their infancy (though see, e.g., Santana 2016 for a philosophy-of-language approach highly informed by interplay with linguistics). Our goal in this subsection will simply be to point out two approaches to the syntax-semantics relation in the corpus linguistics community and, similar to our discussion of the gap in section 3, demonstrate that on neither reading is it impossible to draw interesting generalizations from the scientific literature.

One approach is grounded in the *social use of language* (we take here as representative Teubert 2005). On this view, the meanings of fragments of text are entirely defined by past segments of text which themselves are also to be found within the discourse. "Meaning," Teubert writes, "does not concern the world outside the discourse" (2005, 3). But because text is diachronically connected – because "what is said today is a reaction to what has been said before, an argument in a simultaneous debate and an anticipation of what we expect to be said tomorrow" (Teubert 2005, 4) – we can still extract meanings from texts by looking at the ways in which players in that discourse paraphrase, self-describe, define, even "negotiate" what they consider "meanings" as they arise within a particular discourse tradition.

Such a perspective on semantics has both obvious resonances and tensions with the traditional ways in which philosophers of science have treated the meanings of scientific terms. On the one hand, there is a welcome emphasis on use present here: we shouldn't trust

our own reconstructions of a concept, nor our potentially Whiggish or naive suppositions about how such a concept connects to the outside world. On the other hand, this account requires a very subtle understanding of the relationship between these concepts and the remainder of scientific practice. If meanings are objects within the discourse itself, then we need to carefully build our account of the ways that our other practical and social uses of language can affect those meanings. It is here that, on a sociolinguistics perspective, we can find room for scientific practice and the social structure of science to re-enter. Such a view requires a quite subtle account of the nature and function of scientific concepts, and their accessibility to study by linguistic means.

If this approach strikes the reader as having too-far divorced meaning within the corpus from questions of reference and mental content, consider instead the perspective on corpus analysis derived from *cognitive* approaches to linguistics. Here, by contrast both with the social approach described above and a Chomskyan view on which language processing is handled by a dedicated mental "module," cognitive linguists view syntax, semantics, pragmatics, and the other components of the structure of language as fundamentally mental, and connected to our other processes of cognition (Glynn 2014, 8). What, then, is the role of corpora on such a perspective, where meanings, at least in part, are grounded within the heads of the users of a language rather than lying solely within the discourse itself? Corpora of real-world text are essential in order to falsify or verify hypotheses that we derive about semantics. We can operationalize claims about the meaning and use of concepts by investigating when they become common as linguistic units in a particular corpus, that is, when they function in regular and acceptable uses of language (their "entrenchment" as "grammatically acceptable" in a particular corpus, in Glynn's phrasing).

This sort of picture, again, comes with both advantages and disadvantages. On the one hand, the consistent focus on operationalization of linguistic hypotheses as claims about corpora is welcome in a digital humanities context, within which we can imagine directly implementing ways to test claims about the meanings of scientific concepts within dedicated systems of text analysis (such as Pence 2016). On the other hand, such freedom comes with a significant price, as developing such operationalizable hypotheses about scientific concepts will by no means be a simple endeavor. We hope that DH methods can be widely accessible and easy to use, and thus developing "best practices" surrounding these kinds of syntax-semantics inferences in a philosophy of science context should, if we adopt this linguistic perspective, be a significant and important task for digital philosophy of science in the future.

In sum, while there are again multiple approaches to making the jump from the syntactical aspects of journal articles to their meaning (just as there were multiple ways of theorizing the connection between practice and corpora), on none of these approaches are we unable to make semantic generalizations arising from careful study of the journal literature. The kinds of meta-philosophical work needed to ensure that those claims are justified will differ depending on the other kinds of commitments we have in play, but, in our view, no such meta-level work seems impossible.

5. Conclusion: From Scientific Literature to Empirical Philosophy

Let's step back and take stock. We began by arguing in section 2 that the general question of how empirical philosophers can use scientific literature may be decomposed into a number of separate, related questions. One piece of the puzzle, as discussed in section 3, is to consider

how the literature is derived from science-as-practice, which requires treating that literature as an artifact of publication practices and considering what those practices are like and what they are for. The next step discussed in section 4 – that of producing generalizations about what is in the literature – involves a justification of the ability of DH tools to extract meanings from scientific corpora. It is important to underline that these are two separate but related questions. We owe ourselves answers to each, which can work in concert, to produce a coherent picture of what kinds of philosophical inferences the literature can be taken to support. We conclude here by examining the very last arrow in figure 1, in pursuit of a few glimpses of this broader picture: how can we get from empirical observation of the literature to philosophical arguments or conclusions?

What exactly counts as a philosophical question is of course wide open to debate. For our purposes, we take it to be at least partly a matter of generality. A philosophical theory of science – a theory of "how science works" – is more than a laundry list of empirical particulars: while we needn't insist on anything like universality, we aim to abstract and generalize over particular empirical facts about scientific practice, to at least some degree. In any case, we take it as uncontroversial that philosophical claims about science should be based on well-supported empirical generalizations about what science is really like. For those who might not be convinced, we want to close by hinting at a few ways in which such generalizations could serve as an important tool for philosophical work. In doing so, however, we do not intend to minimize the importance of "cross-validating" the kind of work we describe here with other methods. For that, it seems that one would need to get behind the scenes and observe the practical realities of the research that the literature doesn't show. Even then, however, the comparison between the face of the research presented by the literature

and what is behind it is philosophically interesting in itself. Analyses of the literature, including those made possible with DH tools, would lend themselves to a principled comparison of that kind.

How could generalizations about practice be useful for philosophy of science? First, we might think of them as playing the role of *hypothesis generators*, or *exploratory tools.* Consider, for example, the crucial role played by historical reflection on the Fresnel-Maxwell debate in the development of ontic structural realism (Worrall 1989; French 2011), which was used as a way to demonstrate the possibility – via its actual use by historical scientists – of a novel way of thinking about the status of unobservable entities as grounded in facts about their relational structure. Such a hypothesis, while of course relying on a host of other concerns raised in the literature and the varying configurations of the debate over scientific realism, was in no small part spurred by the realization that another mode of thinking about science could be extracted by generalizing from this particularly fruitful instance of historical theorizing. Generalizations from the literature, we think, have often played such a catalytic role.

Second, we often rely upon such generalizations as ways in which we might test our philosophical hypotheses. As was seen by the tight connection between Medawar's fraud view of the gap and Popperian falsificationism as the "true" theory of scientific knowledge generation, the kinds of claims made by philosophers and historians of science about issues like theory change or scientific progress often come with a variety of very important implications for the actual process of science. Our ability to test these philosophical conjectures thus relies largely on our ability to extract well supported evidence from the varied products of the scientific process itself. In short, on either of these views – or any of

the numerous others that one might propose – the derivation of empirical generalizations from the scientific literature should be a key element in our studies of science.

On the approach we presented here, how could the questions we have raised work together in defense of philosophical claims? Consider, for example, what might happen if we combine one "package" of responses to the various questions we have raised. If one takes the social view of the gap, and adds to it a social-linguistics understanding of the connection between syntax and semantics, the following kind of justification for philosophical work might emerge. Claims about conceptual analysis are indeed accessible via the literature, but those concepts as they appear in the literature are a sort of socially negotiated entity, constantly in flux, subject to definition and redefinition over the course of a particular scientific episode. Our emphasis would need to be on diachronic methods for reconstructing scientific debates, and we would need to support that with a clear understanding of the scientific communities at play, for which citation-network methods might offer a very useful indicator. Connecting such claims to theoretical frameworks or real-world entities would be a messy, piecewise undertaking, likely at least somewhat historically- and community-relative. Of course, other packages of responses would produce dramatically different uses of these tools. We want only to insist that in the absence of such a set of responses, it will remain difficult to determine whether the conclusions drawn in any given study are properly justified (see also Allen and Murdock forthcoming, for similar worries).

We fully acknowledge that there is a long road ahead for the integration of corpus analysis into the methodological toolkit of empirical philosophy of science. However, we hope to have shown that these problems are not qualitatively different from other empirical methods – at least not in a way that is fatal for the possibility of a role for them in philosophy.

In fact, the "gap" between literature and practice is, in effect, an instance of the well-known gap between data and phenomenon: processing and interpreting the data properly simply requires a clear theoretical understanding of the relationship between the two. In turn, this is helped by understanding of the local context of that relationship acquired by other empirical means such as ethnographic studies. Rather than replacing those other methods, they can and should coexist in a relationship of mutual scrutiny and enrichment. In short, DH and its challenges are simply business as usual for empirical study of science (with significant advantages, we think, thanks to the unprecedented scale of the data involved). Again, the way we go about the empirical part of our philosophical research, in this case as in all cases, must be adapted, consciously and repeatedly, to the types of questions we ask and to the particular environments of the scientific fields about which we ask those questions.

In this sense, we hope to be joining a larger chorus of authors within the science studies community who, inspired by the recent insights and successes of practice, empirical, experimental, and other similar approaches to the study of science, have turned to dedicated studies of the justification of our own philosophical practices. We think such a turn is long overdue, and if our work here can encourage further reflection on these important questions, we will consider it successful.

Acknowledgments

This work was supported by the Fonds de la Recherche Scientifique – FNRS under grant no. F.4526.19. For comments on previous versions of this paper, we thank Mark Alfano and Carlos Santana. This paper benefited from an audience at the XPhiSci 2019 conference at

Aarhus University, especially Sam Schindler, Edouard Machery, Moti Mizrahi, Deena Skolnick Weisberg, Carlos Santana, Michiru Nagatsu, and Andrew Shtulman; and at the CEFISES Seminar at the Université catholique de Louvain, especially Michel Ghins, Gwenael Laurent, Alexandre Guay, and Peter Verdée.

References

Abbasi, Alireza, Jörn Altmann, and Liaquat Hossain. 2011. "Identifying the Effects of Co-Authorship Networks on the Performance of Scholars: A Correlation and Regression Analysis of Performance Measures and Social Network Analysis Measures." *Journal of Informetrics* 5 (4): 594–607. https://doi.org/10.1016/j.joi.2011.05.007.

Allen, Colin and Murdock, Jaimie. Forthcoming. "LDA Topic Modeling: Contexts for the History & Philosophy of Science." In De Block, Andreas and Ramsey, Grant, eds., *The Evolution of Science.* Pittsburgh, PA: University of Pittsburgh Press.

Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–12. https://doi.org/10.1126/science.286.5439.509.

Baxter, Janella. 2019. "How Biological Technology Should Inform the Causal Selection Debate." *Philosophy, Theory, and Practice in Biology* 11: 2. https://doi.org/10.3998/ptpbio.16039257.0011.002.

Börner, Katy, Jeegar T. Maru, and Robert L. Goldstone. 2004. "The Simultaneous Evolution of Author and Paper Networks." *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5266–73. https://doi.org/10.1073/pnas.0307625100.

Burrows, John F. 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17 (3): 267–87.

Chavalarias, David, and Jean-Philippe Cointet. 2013. "Phylomemetic Patterns in Science Evolution—the Rise and Fall of Scientific Fields." *PLOS ONE* 8 (2): e54847. https://doi.org/10.1371/journal.pone.0054847.

Dias, Laércio, Martin Gerlach, Joachim Scharloth, and Eduardo G. Altmann. 2018. "Using Text Analysis to Quantify the Similarity and Evolution of Scientific Disciplines." *Royal Society Open Science* 5 (1). https://doi.org/10.1098/rsos.171545.

French, Steven. 2011. "Shifting to Structures in Physics and Biology: A Prophylactic for Promiscuous Realism." *Studies in History and Philosophy of Biological and Biomedical Sciences* 42 (2): 164–73. https://doi.org/10.1016/j.shpsc.2010.11.023.

Girvan, M., and M. E. J. Newman. 2002. "Community Structure in Social and Biological Networks." *Proceedings of the National Academy of Sciences* 99 (12): 7821–26. https://doi.org/10.1073/pnas.122653799.

Glynn, Dylan. 2014. "Polysemy and Synonymy: Cognitive Theory and Corpus Method." In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, edited by Dylan Glynn and Justyna A. Robinson, 7–38. Amsterdam: John Benjamins.

Hacking, Ian. 1982. “Experimentation and Scientific Realism.” *Philosophical Topics* 13 (1): 71–87. https://doi.org/10.5840/philtopics19821314.

Hoover, David L. 2010. “Authorial Style.” In *Language and Style: Essays in Honour of Mick Short*, edited by Dan McIntyre and Beatrix Busse, 250–271. Palgrave.

Ioannidis, John P. A. 2005. “Why Most Published Research Findings Are False.” *PLoS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

Kay, Lily E. 2000. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press.

Kessler, M. M. 1963. “Bibliographic Coupling between Scientific Papers.” *American Documentation* 14 (1): 10–25. https://doi.org/10.1002/asi.5090140103.

Knorr-Cetina, Karin D. 1981. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon.

Latour, Bruno, and Steve Woolgar. 1987. *Laboratory Life: The Construction of Scientific Facts*. 2nd ed. Princeton, NJ: Princeton University Press.

Leonelli, Sabina. 2016. *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.

MacRoberts, Michael H., and Barbara R. MacRoberts. 1989. “Problems of Citation Analysis: A Critical Review.” *Journal of the American Society for Information Science* 40 (5): 342–49.

Malaterre, Christophe, Jean-François Chartier, and Davide Pulizzotto. 2019. “What Is This Thing Called Philosophy of Science? A Computational Topic-Modeling Perspective, 1934–2015.” *HOPOS* 9 (2): 215–49. https://doi.org/10.1086/704372.

Medawar, Peter B. 1963. “Is the Scientific Paper a Fraud?” *The Listener* 70 (September): 377–378.

Mizrahi, Moti. 2020. “The Case Study Method in Philosophy of Science: An Empirical Study.” *Perspectives on Science* 28 (1): 63–88. https://doi.org/10.1162/posc_a_00333.

Nersessian, Nancy J. 1995. “Opening the Black Box: Cognitive Science and History of Science.” *Osiris* 10 (1): 194–211. https://doi.org/10.1086/368749.

Nickles, Thomas. 1988. “Reconstructing Science: Discovery and Experiment.” In *Theory and Experiment: Recent Insights and New Perspectives on Their Relation*, edited by Diderik Batens and Jean Paul van Bendegem, 33–53. Dordrecht: D. Reidel.

Pechenick, Eitan Adam, Christopher M. Danforth, and Peter Sheridan Dodds. 2015. “Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution.” *PLoS ONE* 10 (10): e0137041. https://doi.org/10.1371/journal.pone.0137041.

Pence, Charles H. 2016. “RLetters: A Web-Based Application for Text Analysis of Journal Articles.” *PLoS ONE* 11 (1): e0146004. https://doi.org/10.1371/journal.pone.0146004.

Raan, Anthony F. J. van. 2004. “Measuring Science.” In *Handbook of Quantitative Science and Technology Research*, edited by Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch, 19–50. Dordrecht: Springer.

Reichenbach, Hans. 1938. *Experience and Prediction: An Analysis of the Foundations and Structure of Knowledge*. Chicago: University of Chicago Press.

Rouse, Joseph. 1990. “The Narrative Reconstruction of Science.” *Inquiry* 33 (2): 179–96.

Santana, Carlos. 2016. “What Is Language?” *Ergo* 3 (19). https://doi.org/10.3998/ergo.12405314.0003.019.

Schickore, Jutta. 2008. “Doing Science, Writing Science.” *Philosophy of Science* 75 (3): 323–43.

Shibata, Naoki, Yuya Kajikawa, and Ichiro Sakata. 2012. “Link Prediction in Citation Networks.” *Journal of the American Society for Information Science and Technology* 63 (1): 78–85. https://doi.org/10.1002/asi.21664.

Smaldino, Paul E., and Richard McElreath. 2016. “The Natural Selection of Bad Science.” *Royal Society Open Science* 3 (9): 160384. https://doi.org/10.1098/rsos.160384.

Small, Henry. 1973. “Co-Citation in the Scientific Literature: A New Measure of the Relationship between Two Documents.” *Journal of the American Society for Information Science* 24 (4): 265–69. https://doi.org/10.1002/asi.4630240406.

Soler, Léna, Sjoerd Zwart, Michael Lynch, and Vincent Israel-Jost, eds. 2014. *Science after the Practice Turn in the Philosophy, History, and Social Studies of Science*. New York: Routledge.

Suppe, Frederick. 1998. “The Structure of a Scientific Paper.” *Philosophy of Science* 65 (3): 381–405. https://doi.org/10.1086/392651.

Teubert, Wolfgang. 2005. “My Version of Corpus Linguistics.” *International Journal of Corpus Linguistics* 10 (1): 1–13. https://doi.org/10.1075/ijcl.10.1.01teu.

Tshitoyan, Vahe, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder, and Anubhav Jain. 2019. “Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature.” *Nature* 571 (7763): 95–98. https://doi.org/10.1038/s41586-019-1335-8.

Waters, C. Kenneth. 2008. “Beyond Theoretical Reduction and Layer-Cake Antireduction: How DNA Retooled Genetics and Transformed Biological Practice.” In *The Oxford Handbook of Philosophy of Biology*, edited by Michael Ruse, 238–62. Oxford: Oxford University Press.

Wilkinson, Dennis M., and Bernardo A. Huberman. 2004. “A Method for Finding Communities of Related Genes.” *Proceedings of the National Academy of Sciences* 101 (suppl 1): 5241–48. https://doi.org/10.1073/pnas.0307740100.

Wimsatt, William C. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Worrall, John. 1989. “Structural Realism: The Best of Both Worlds?” *Dialectica* 43 (1–2): 99–124. https://doi.org/10.1111/j.1746-8361.1989.tb00933.x.

Wouters, P. 1998. “The Signs of Science.” *Scientometrics* 41 (1–2): 225–41.