Diachronic Trends in the Topic Distributions of Formal Epistemology Abstracts

David Kinney

Forthcoming in *Synthese*[*]

October 28, 2021

Abstract

Formal epistemology is a growing field of philosophical research. It is also evolving, with the subject matter of formal epistemology papers changing considerably over the past two decades. To quantify the ways in which formal epistemology is changing, I generate a stochastic block topic model of the abstracts of papers classified by `PhilPapers.org` as pertaining to formal epistemology. This model identifies fourteen salient topics of formal epistemology abstracts at a first level of abstraction, and four topics at a second level of abstraction. I then study diachronic trends in the degree to which formal epistemology abstracts written in a given year are likely to contain words associated with a particular topic, beginning in 2000 and continuing to 2020. My findings suggest that there has been a marked decline in the likelihood of a given formal epistemology abstract being about logical approaches to belief revision (e.g., AGM belief-revision theory). On the other hand, over the past two decades, the salience of probabilistic techniques in formal epistemology has increased, as has the salience of work at the intersection of formal epistemology and some areas of philosophy of science.

1 Introduction

Formal epistemology is a sub-field of epistemology distinguished essentially by its methodology, rather than its subject matter. Like traditional analytic epistemology, formal epistemology is

[*]I am grateful to Julia Lefkowitz and two anonymous reviewers for their insightful comments on early drafts of this manuscript.

concerned with the analysis of norms governing the epistemic attitudes that agents hold towards propositions (e.g., belief and knowledge). What distinguishes formal epistemology is the use of logical and mathematical methods (e.g., modal logic, probability theory, decision theory, computer simulations, or topology, among other methods) to model epistemic norms. In this paper, I present what is (to my knowledge) the first quantitative meta-philosophical study of formal epistemology. Using statistical techniques developed by Gerlach et al. (2018), I analyze a corpus of 1733 abstracts of books, book chapters, and journal articles published between 1937 and 2021 which are classified as pertaining to formal epistemology by the philosophy archiving website `PhilPapers.org`. As the result of my analysis, I identify fourteen salient topics with which formal epistemology abstracts tend to be concerned. I then provide a diachronic analysis of the changes in the comparative salience of these topics over the course of the twenty-first century.

The principle finding of this analysis is that the two decades from 2000 to 2020 have seen a decline in the salience of explicitly logical approaches to the analysis of belief and knowledge in formal epistemology. Such approaches are exemplified by the classic paper, "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions," by Alchourrón et al. (1985), which presented the now-famous "AGM" theory of belief revision (the acronym 'AGM' stands for the paper's three authors: Alchorrón, Gärdenfors, and Makinson). While deeply influential, the data suggest that the overall salience of the AGM approach and other approaches using epistemic logic has declined within formal epistemology over the previous two decades, while probabilistic approaches have seen an increase in salience over the same period.[1]

That the decline of logical techniques would be accompanied by a rise in probabilistic techniques is not surprising from a conceptual perspective; as Hansson (2017) notes, "[i]t has proved to be difficult to construct a reasonably manageable model that covers both the logic-related and the probabilistic properties of belief change." Since it is difficult to model belief change simultaneously

[1] In a paper written at the same time as this one, and recently accepted by this journal, Fletcher et al. (forthcoming) hand-coded a sample of papers from the journal *Philosophical Studies* to detect the use of formal methods and investigate the specific formal methods used in each paper. Their findings cohere in with mine in several respects. They show a significant increase in the use of probabilistic methods in formal philosophy papers over the course of the twenty-first century. While their analysis detected no significant change in the use of logical methods within formal philosophy as a whole, it did detect a decline in the proportion of epistemology papers making use of logical methods. Similarly, my analysis here indicates a decline in the salience of logical approaches within formal epistemology in particular. Ultimately, while there is clearly some connection between my results and theirs, our studies are different in several important respects. Most saliently, their paper and this one use very different techniques for gathering data, and analyze corpora that are constructed using very different principles.

using both logical and probabilistic techniques, it is plausible that a rise in the overall salience of the latter over a given time period would be accompanied by a decline in the overall salience of the former over the same time period, and indeed this is what the data show. Additionally, the data show an increase in the salience of scientific applications of formal epistemology, and of work on the nature of evidence.

Besides demonstrating the aforementioned pattern of change in the salience of various topics in formal epistemology, this paper has two broader goals. First, I aim here to contribute to the nascent field of quantitative study of the research culture of analytic philosophy. To date, topic modeling has been the primary tool for doing this kind of empirical study of the research culture of philosophy in which text-mining techniques are used to model the salience of topics in the philosophical literature (see Malaterre et al. (2020) for a comprehensive topic model of the philosophy of science literature, and Weatherson (2021a) for a similar study of the full philosophy literature, respectively), and this paper contributes to that effort. However, the second goal of the paper is to pilot the use of network-based stochastic block model (SBM) techniques for topic modelling developed by Gerlach et al. (2018). SBM techniques are in some respects an improvement on traditional latent Dirichlet allocation (LDA) techniques for topic modelling developed by Blei et al. (2003) and used in early applications by Griffiths et al. (2004), but they remain under-utilized at present; for example, both Malaterre et al. (2020) and Weatherson (2021a) use LDA techniques in their topic models of philosophy corpora. Thus, a further aim of this paper is to provide an additional "proof of concept" for the utility of network-based techniques for the analysis of textual corpora, and to investigate some of the trade-offs associated with using an SBM approach as compared to an LDA approach.

The remainder of this paper proceeds as follows. In Section 2, I provide details of the specific corpus of abstracts studied and the methods used to analyze the corpus. In Section 3, I present the results of my analysis. In Section 4, I discuss the significance of these results. I conclude in Section 5 by discussing potential avenues for future work. Two technical appendices provide further details of the methods used and the results obtained.

2 Corpus and Methods

2.1 Corpus

I initially analyzed the 2803 journal articles, books, and book chapters categorized by `PhilPapers.org` as pertaining to formal epistemology as of 18 March, 2021. Although PhilPapers uses some automated categorization of materials, most categorization is done by users, who may or may not be the authors of the papers being categorized.[2] Of these 2803 works, only 1733 had text abstracts included as part of their PhilPapers entry, and so only this subset of the formal epistemology literature was actually analyzed. This raises an important caveat about the analysis presented in this paper. Namely, it not really an analysis of the formal epistemology literature in its entirety. Rather, it is an analysis the abstracts of the 1733 works which were categorized by PhilPapers as pertaining to formal epistemology, and for which textual abstracts were entered into PhilPapers. The causal structure of the process whereby works come to be posted to PhilPapers and categorized as pertaining to formal epistemology, as well as the process whereby the text of the abstract of a given work comes to be included in the PhilPapers entry for that work, may introduce significant bias into the corpus that I am unable to account for here. Said biases may be compounded by the fact that the formal epistemology category within the PhilPapers archive currently lacks an editor, and so there is no centralized control over the process whereby abstracts are categorized as pertaining to formal epistemology.

The abstracts, as well as data on the title, author(s), publication venue, and year of publication of the paper associated with each abstract, were scraped from `PhilPapers.org` using the `BeautifulSoup` Python package.[3] So called "stop words" (i.e., uninformative words like 'the' and 'it'), as well as other uninformative words, were removed from the abstracts, also using the `BeautifulSoup` package. In addition, several common punctuation marks, such as quotes and brackets, were manually removed from the text of the abstracts prior to analysis.

Hereafter, I refer to each of the $n = 1733$ abstracts analyzed as a **document** d_i. I refer to the full set of documents $\mathcal{D} = \{d_1, \ldots, d_z\}$ as the **corpus**. Each document d_i is itself a sequence of m_i words $d_i = (w_{i,1}, \ldots, w_{i,m_i})$. In keeping with standard topic modeling practice, we represent each

[2]For full details of PhilPapers' categorization policy, see `https://philpapers.org/help/categorization.html`.
[3]For all code and data used in this paper, see `https://github.com/davidbkinney/formalepistemologySBM`.

document as a "bag of words", in which the ordering of words does not distinguish one document from another, but in which documents can be distinguished solely by the number of times that a given word occurs. The **vocabulary** $\mathcal{W} = \{w_1 \ldots, w_m\}$ is the set of all words that appear in at least one document; for the sake of the analysis here words that appear only once in the entire corpus are removed. In the corpus under study, the vocabulary has cardinality $m = 17138$. Although they are not strictly part of the definition of a document used in the construction of the topic model, for each document d_i we also record data on its title, author(s), publication venue, and year of publication. This allows us to partition the corpus $\mathcal{D}$ into the set $\mathcal{Y} = \{Y_1, \ldots, Y_q\}$, where each $Y_u \in \mathcal{Y}$ contains all and only those documents that share a specific year of publication, which in turn enables diachronic analysis of the corpus.

2.2 Methods

The basic idea behind any topic model, whether it is the older, LDA method pioneered by Blei et al. (2003) and Griffiths et al. (2004), or the newer SBM method due to Gerlach et al. (2018), is that a **topic** t is a probability distribution over the vocabulary $\mathcal{W}$ for a given corpus $\mathcal{D}$. In a topic model, it is assumed that each document d_i is generated by selecting a given topic $\hat{t}$ from a set of possible topics $\mathcal{T}$ by sampling from a document-specific probability distribution $p(\cdot|d_i)$ that is defined over $\mathcal{T}$. We call this the **topic distribution** for the document d_i. Next, a word is generated by sampling from $\hat{t}$ (recall that $\hat{t}$, like all topics, is a probability distribution over words in the vocabulary). This process is iterated until the document is filled with m_i words. To infer a topic model, we aim to find the set of probability distributions $p(\cdot|d_i)$ for each document $d_i \in \mathcal{D}$ and the set of topics $\mathcal{T}$ that maximizes the likelihood of the observed corpus without over-fitting. We then examine the words in the vocabulary assigned highest probability by each topic $t \in \mathcal{T}$ and, using domain expertise, determine appropriate labels for each topic. For instance, a topic that assigns highest probability to the words 'probability', 'rational', 'probabilistic', and 'rationality' might be assigned the label 'Probability and Rationality'. Then, for any topic t and document d_i, we interpret the probability $p(t|d_i)$ as the degree to which d_i is about t; the higher this probability, the greater the degree to which the document d_i is about the topic t.

As discussed in the introduction, this paper uses the newer SBM methodology for inferring a topic model from a corpus, following work by Gerlach et al. (2018), which in turn builds on community-detection techniques developed in the network physics literature by Ball et al. (2011) and Peixoto (2015, 2017). See Appendix A for a more detailed description of the SBM method, and an explanation of how it differs from the older LDA method for topic modelling. For now, I note three advantages of the SBM method. First, LDA generates topics by sampling from a unimodal mixture distribution, and then identifying the topics that render the corpus most likely. By contrast, SBM methods do not make any assumptions about the modality of the mixture distribution from which topics are sampled. Second, LDA methods require the modeller to pre-specify the number of topics that they wish to generate when inferring a topic model from a corpus; SBM methods require no such pre-specification, and instead infer the number of topics without modeller input. Third, LDA only generates topics at a single level of abstraction, whereas SBM automatically generates topics at multiple levels of abstraction. The meaning of this third distinction will be made clearer over the course of this section.

One can give a general description of the SBM methodology as follows. Recall that z is the cardinality of the corpus $\mathcal{D}$ and m is the cardinality of the vocabulary $\mathcal{W}$. Consider a **multigraph** $\mathcal{G}$ composed of $(z+m)$ **nodes** $\mathcal{N}$ and an $(z+m)\times(z+m)$ **adjacency matrix A** wherein each entry A_{ij} is a natural number. This natural number is an "edge" relating the nodes n_i and n_j. The nodes n_i and n_j are **adjacent** just in case $A_{ij} > 0$. The set of nodes $\mathcal{N}$ is partitioned into a set $\mathcal{N}_{\mathcal{D}}$ representing each document in the corpus, and a set $\mathcal{N}_{\mathcal{W}}$ representing each word in the corpus' vocabulary. Importantly, we stipulate that no node in $\mathcal{N}_{\mathcal{D}}$ is adjacent to another node in $\mathcal{N}_{\mathcal{D}}$, nor is any node in $\mathcal{N}_{\mathcal{W}}$ adjacent to another node in $\mathcal{N}_{\mathcal{W}}$. That is, if $n_i \in \mathcal{N}_{\mathcal{D}}$ and $n_j \in \mathcal{N}_{\mathcal{D}}$, or $n_i \in \mathcal{N}_{\mathcal{W}}$ and $n_j \in \mathcal{N}_{\mathcal{W}}$, then $A_{ij} = 0$. The value of any positive A_{ij} represents the number of times that word w_j appears in document d_i, or the number of times word w_i appears in document d_j, depending on whether n_i and n_j represent a word or a document, respectively. See Figure 1 for an illustration of this kind of multigraph for a very small corpus.

Next, we define a "higher-level" set of word nodes $\mathcal{N}^{\uparrow}_{\mathcal{W}} = \{N^{\uparrow 1}_{\mathcal{W}}, \ldots, N^{\uparrow v}_{\mathcal{W}}\}$ such that each node in $\mathcal{N}_{\mathcal{W}}$ is an element of at least one $\mathcal{N}^{\uparrow}_{\mathcal{W}}$. This allows us to define a multigraph $\mathcal{G}^{\uparrow}$ containing the nodes $\mathcal{N}_{\mathcal{D}} \cup \mathcal{N}^{\uparrow}_{\mathcal{W}}$ and the adjacency matrix $\mathbf{A}^{\uparrow}$ in which each positive entry $A^{\uparrow}_{ir}$ is the number of words in the set $N^{\uparrow r}_{\mathcal{W}}$ that appear in the document d_i.

A **stochastic block model** is a statistical model of the process whereby the lower-level multigraph $\mathcal{G}$ representing the actual corpus is generated from the upper-level multigraph $\mathcal{G}^{\uparrow}$. Specifically,

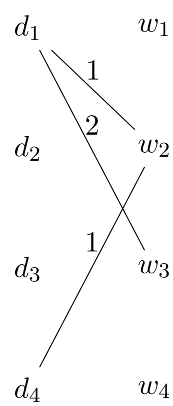

Figure 1: Simple four-document, four-word multigraph in which the document d_1 contains the word w_2 once and w_3 twice, and the document d_4 contains the word w_3 once.

we model $\mathcal{G}$ as generated via the following process:

1. Obtain the degree of each node $n_i \in \mathcal{N}_{\mathcal{D}}$ representing a document d_i (i.e., the number of words in each document).

2. Obtain the degree of each node $n_j \in \mathcal{N}_{\mathcal{W}}$ representing a word w_j (i.e., the the number of times that word appears in corpus).

3. Infer a set of subsets $\mathcal{N}^{\uparrow}_{\mathcal{W}}$ of the vocabulary $\mathcal{W}$.

4. Obtain the higher-level adjacency matrix $\mathbf{A}^{\uparrow}$ for the multigraph with nodes $\mathcal{N}_{\mathcal{D}} \cup \mathcal{N}^{\uparrow}_{\mathcal{W}}$.

5. Randomly pair document-nodes to word-nodes to generate a multigraph $\mathcal{G}$ that is consistent with the higher-level adjacency matrix $\mathbf{A}^{\uparrow}$, given the grouping of words $\mathcal{N}^{\uparrow}_{\mathcal{W}}$.

When inferring a stochastic block model, the goal is to infer a word grouping $\mathcal{N}^{\uparrow}_{\mathcal{W}}$ that maximizes the likelihood of generating the lower-level multigraph $\mathcal{G}$ without over-fitting. How precisely this is done is described in more detail in Appendix A.

The probability distributions $p(\cdot|\mathcal{N}^{\uparrow r}_{\mathcal{W}})$ (i.e., the topics), can be calculated via the following equation:

$$p(w_j|\mathcal{N}^{\uparrow r}_{\mathcal{W}}) = \begin{cases} \frac{\sum_{i=1}^{z} A_{ij}}{\sum_{i=1}^{z} A^{\uparrow}_{ir}} & \text{if } w_j \in N^{\uparrow r}_{\mathcal{W}} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

That is, the probability of word w_j being sampled from topic $\mathcal{N}_{\mathcal{W}}^{\uparrow r}$ is the ratio between the degree of the node representing w_j in the lower-level multigraph $\mathcal{G}$ and the degree of the node $\mathcal{N}_{\mathcal{W}}^{\uparrow r}$ in the higher-level multigraph $\mathcal{G}^{\uparrow}$. Similarly, we can calculate topic distributions $p(\cdot|d_i)$ over the set of higher-level word nodes via the following equation:

$$p(N_{\mathcal{W}}^{\uparrow r}|d_i) = \frac{A_{ir}^{\uparrow}}{\sum_{r=1}^{z} A_{ir}^{\uparrow}}. \tag{2}$$

That is, $p(N_{\mathcal{W}}^{\uparrow r}|d_i)$ is the ratio between the number of words in d_i that are elements of the group $N_{\mathcal{W}}^{\uparrow r}$, and the total number of words in d_i. Said probability distributions $p(\cdot|d_i)$ over $\mathcal{N}_{\mathcal{W}}^{\uparrow}$ are the topic distributions.

Crucially, we can understand $\mathcal{G}^{\uparrow}$ as itself generated by a higher-level stochastic block model. That is, we introduce a set of even higher-level word nodes $\mathcal{N}_{\mathcal{W}}^{\uparrow\uparrow} = \{N_{\mathcal{W}}^{\uparrow\uparrow 1}, \ldots, N_{\mathcal{W}}^{\uparrow\uparrow o}\}$, where $o < v$ and $\mathcal{N}_{\mathcal{W}}^{\uparrow\uparrow}$ is also a set of subsets of $\mathcal{W}$. This yields a higher-level adjacency matrix $\mathbf{A}^{\uparrow\uparrow}$ for a multigraph $\mathcal{G}^{\uparrow\uparrow}$ in which nodes represent either documents or elements of $\mathcal{N}_{\mathcal{W}}^{\uparrow\uparrow}$. At the second level of abstraction, the goal of stochastic block modeling is to infer the grouping $\mathcal{N}_{\mathcal{W}}^{\uparrow\uparrow}$ of $\mathcal{W}$ that maximizes the likelihood of generating the multigraph $\mathcal{G}^{\uparrow}$ without over-fitting. We can continue iterating this process at higher levels of abstraction until we generate a model with a single word node from which all lower-level word nodes can be sampled.

SBM thereby adopts a decidedly "top-down" understanding of the process whereby the topics covered in a particular corpus determine the specific content of that corpus. At a given level of abstraction, the SBM aims to find a graphical structure at the next-highest level of abstraction that renders the graphical structure of the corpus at more concrete levels of description most likely. Thus, in an SBM, one conceives of the fine-grained or concrete structure of the corpus as *generated by* a more coarse-grained or abstract description of the same corpus. In what follows, I use a variation of the software developed by Gerlach et al. (2018) to infer a stochastic block model for the formal epistemology corpus described above. I conduct inference at two levels of abstraction. Results are described in the next section.

3 Results

3.1 The Topics

3.1.1 Level 1

At the first level of abstraction, the model identified twenty-one topics. Of these, four were deemed to have no clear interpretation. That is, the words assigned highest probability in that topic were not evocative of any known sub-field of formal epistemology. In three separate cases, two topics were deemed to be similar enough that they should be combined into a single topic, resulting in a total of fourteen topics with a clear interpretation. The three pairs of topics that were combined were topics 2 and 13, topics 16 and 17, and topics 17 and 18. The complete list of topics is as follows:

0. Evidence
1. No Clear Interpretation
2/13. Logical Approaches to Belief Revision
3. Logical Approaches to Knowledge
4. Probability and Rationality
5. Inductive Risk
6. Applications to General PhilSci
7. No Clear Interpretation
8. Abstracts not in English
9. Judgement
10. Coherence and Conditionalization
11. Applications to Scientific Modelling
12. Hypothesis Testing and Confirmation
14. No Clear Interpretation
15. Data Science Applications
16/17. Applications to Scientific Practice[4]
18/19. Causal Modelling
20. No Clear Interpretation

In Appendix B, I list the twenty words in the vocabulary with highest probability within each of these topics. An exception is topic 19, in which only five words have positive probability. As indicated in the list above, topic 19 was combined with topic 18 due to significant overlap in their content. Note that there is room for disagreement with respect to the correctness of my proposed

[4]While there is room for disagreement about how any topic is to be interpreted, with respect to Topics 16 and 17 there is an especially strong case to be made that the high probabilities assigned to words like ‘functional’, ‘brain’, ‘mental’, ‘disease’, and ‘physical’ suggest an association with philosophy of mind or psychology. At the same time, words like ‘scientists’, ‘treatment’, ‘confirm’, ‘caused’, and ‘estimation’, which are also assigned high probability by these topics, do not have any special relevance to philosophy of mind, but are relevant to scientific practice. For this reason, I have chosen the topic title ‘Applications to Scientific Practice,’ but flag here that “scientific practice” should be taken to include the practice of neuroscience, cognitive science, and psychology.

typology of the inferred topics in this corpus. This highlights the ineliminable subjective component of the interpretation of the findings of topic models, as it is unlikely that any two practitioners will be in total agreement as to the correct characterization of every inferred topic.

The ability of SBM methods to learn a topic model without pre-specifying the number of topics gives SBM techniques an important advantage over LDA techniques. Namely, the user of an SBM model does not need to make a difficult and consequential decision about the number of topics to be found, which in some cases can influence the overall picture of the corpus that emerges from a topic modelling analysis. On the other hand, the high prevalence of "no clear interpretation" topics in this analysis demonstrates one downside of an SBM topic model; it can return topics that do not admit of clear interpretation, and, unlike in an LDA approach, the user cannot fine-tune the number of topics to avoid results in which some topics must be declared meaningless.

Nevertheless, it is important to emphasize that no abstracts are in any sense "excluded" from this study as the result of not analyzing certain topics without a clear interpretation. Recall that for each abstract, there is a probability distribution over all topics representing the probability that a word in that abstract is generated by sampling from that topic. This is that abstract's topic distribution. Thus, choosing not to analyze uninterpreted topics would only exclude an abstract if that abstract's topic distribution only assigned positive probability to the four excluded topics at Level 1, and the single excluded topic at Level 2. No abstracts in our corpus satisfy this condition, and so no abstracts are excluded from our analysis through the omission of topics without a clear interpretation.

3.1.2 Level 2

At the second level of abstraction, the model identified five topics. Of these, one was judged to lack a clear interpretation, while four did have clear interpretations. The topics inferred at the second level of analysis are:

0. Belief and Belief Revision
1. No Clear Interpretation
2. Applications to Scientific Modelling
3. Abstracts not in English
4. Causal Modelling

As in the case of the topics inferred at the lower level of abstraction, the twenty words with highest

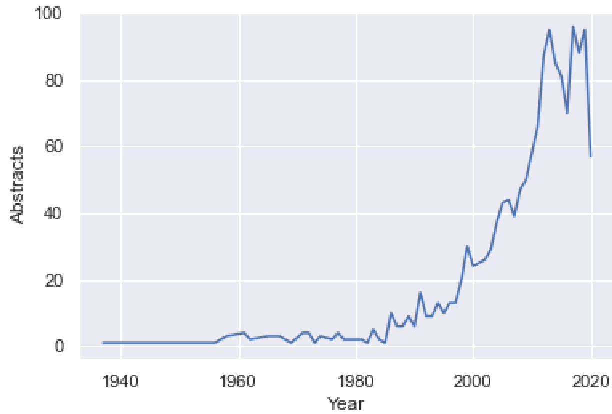

Figure 2: Number of formal epistemology abstracts on PhilPapers published in each year 1937-2020.

probability in each topic are listed in Appendix B.

3.2 Diachronic Analysis

3.2.1 Overall Growth in Formal Epistemology

Using the topic distributions for each document in the corpus, we are able to perform diachronic analysis of changes in the salience of various topics in formal epistemology. However, I begin by making a synoptic point: the data collected here suggest that formal epistemology is indeed a growing field. The growth in formal epistemology abstracts by publication year is shown in Figure 2. From 2000 to 2020, average year-on-year growth in the number of formal epistemology abstracts published, according to the PhilPapers data, is 6.01%. Note that this analysis excludes abstracts in the corpus whose listed publication year is 2021 (10), 'forthcoming' (87), 'manuscript' (49), and 'unknown' (115). Moreover, it is possible that the apparent significant decline in formal epistemology papers with a publication year of 2020 is due to the fact that many publications uploaded to PhilPapers and published that year remain listed as having a publishing date of 'forthcoming', 'manuscript', or 'unknown', or have not yet included an abstract in their PhilPapers entry. If we exclude 2020 data from the analysis, average year-on-year growth in the number of formal epistemology abstracts from 2000-2019 is 8.43%.

3.2.2 Level 1

To analyze changes in the salience of topics over time, recall first that we can partition the corpus of abstracts $\mathcal{D}$ into the set $\mathcal{Y}$, where two documents are in a given $Y_u \in \mathcal{Y}$ if and only if they have the same publication year. As before, let $\mathcal{N}_{\mathcal{W}}^{\uparrow}$ be the set of topic nodes at the first level of abstraction. We define an **annual salience function** $s : \mathcal{Y} \times \mathcal{N}_{\mathcal{W}}^{\uparrow} \to [0, 1]$ as follows:

$$s(Y_u, N_{\mathcal{W}}^{\uparrow r}) = \frac{1}{|Y_u|} \sum_{d_i \in Y_u} p(N_{\mathcal{W}}^{\uparrow r} | d_i) \tag{3}$$

That is, the annual salience of a topic $N_{\mathcal{W}}^{\uparrow r}$ in a year Y_u is the average probability that a document with the publication year Y_u is about the topic $N_{\mathcal{W}}^{\uparrow r}$. Where two topics are combined, we calculate their total salience by adding together the individual salience of each topic.

Figure 3 shows the annual salience of each topic in each year from 2000 to 2020. Years prior to 2000 were excluded from chronological analysis of individual topics due to the relatively small number of annual publications in formal epistemology prior to 2000, which results in highly volatile trends in the annual salience of various topics throughout the twentieth century. Table 1 shows the regression coefficient β (i.e., the slope of the line of best fit) for the annual trend in the salience of each topic from 2000 to 2020, the square of the value of Pearson's r, which provides a normalized measure of the overall linearity of the correlation captured by the value of β, and the p-value of each trend. It also shows the average salience of each topic over the five-year period from 2015-2020, as a measurement of the recent salience of the topic.

We find that the topic with the strongest negative trend in annual salience is Logical Approaches to Belief Revision. Logical Approaches to Knowledge also has a negative trend overall, though this trend is not significant at the 5% level. Moreover, Logical Approaches to Belief Revision has the highest value of r^2 of any annual salience trend for any topic, suggesting that the line of best fit captures well the overall declining trend in the salience of this topic. This fits with a narrative in which formal epistemology is moving away from epistemic logic as the primary method used to model epistemic attitudes and norms. Nevertheless, these topics remain relatively salient within the contemporary formal epistemology corpus, in part because they were so dominant in the past: from 2000 to 2005, the average salience of Logical Approaches to Belief Revision was .127, and

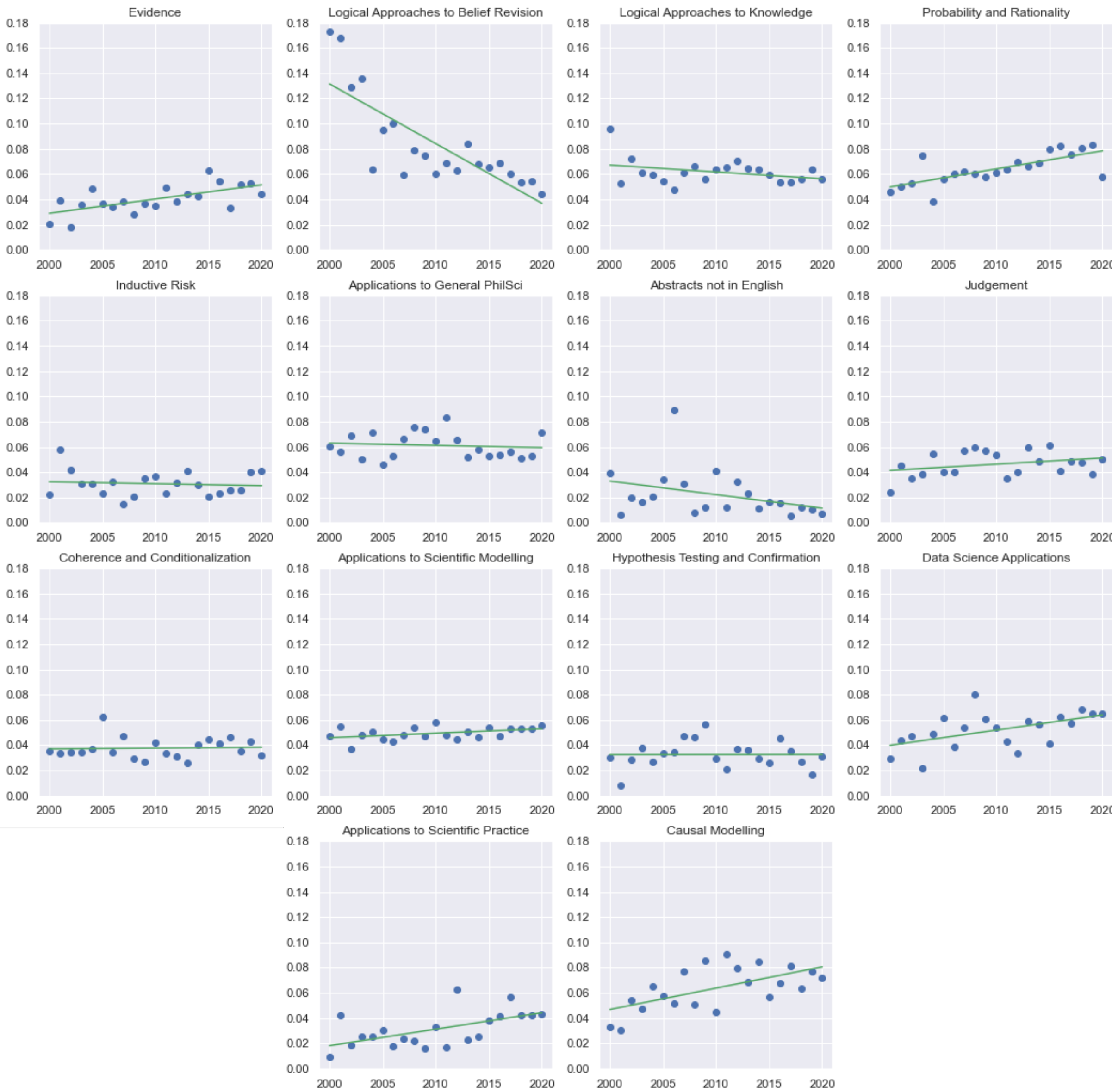

Figure 3: Salience of each of the Level 1 topics from 2000-2020.

Topic	β	r^2	p-value	2015-20 Avg. Salience
0. Evidence	1.12×10^{-3}	.41	1.87×10^{-3}	.050
2/13. Logical Approaches to Belief Revision	-4.72×10^{-3}	.63	1.55×10^{-5}	.058
3. Logical Approaches to Knowledge	-5.44×10^{-4}	.11	1.35×10^{-1}	.057
4. Probability and Rationality	1.43×10^{-4}	.52	2.35×10^{-4}	.076
5. Inductive Risk	-1.55×10^{-4}	.01	6.78×10^{-1}	.029
6. Applications to General PhilSci	-1.76×10^{-4}	.01	6.45×10^{-1}	.056
8. Abstracts not in English	-1.07×10^{-3}	.13	1.16×10^{-1}	.011
9. Judgement	4.93×10^{-4}	.09	1.76×10^{-1}	.048
10. Coherence and Conditionalization	6.64×10^{-5}	.002	8.30×10^{-1}	.041
11. Applications to Scientific Modelling	3.57×10^{-4}	.19	4.91×10^{-2}	.053
12. Hypothesis Testing and Confirmation	6.30×10^{-6}	$< .001$	9.88×10^{-1}	.030
15. Data Science Applications	1.20×10^{-3}	.28	1.43×10^{-2}	.060
16/17. Applications to Scientific Practice	1.30×10^{-3}	.34	5.84×10^{-3}	.044
18/19. Causal Modelling	1.69×10^{-3}	.38	3.10×10^{-3}	.070

Table 1: Regression coefficient for the annual salience of formal epistemology topics from 2000-2020, Pearson's r^2 and p-value for each annual salience trend, and the average salience of each topic from 2015-2020.

the average salience of Logical Approaches to Knowledge was .066. The former figure is greater than the average salience of *any* topic over the five years from 2015-20, and the latter figure is higher than the salience of almost all topics over the same period. Thus, the data suggest that over the past twenty years logical approaches to belief and knowledge have declined from being the dominant topics in the formal epistemology corpus to being one topic among many.

On the other hand, the fastest-growing topics in the corpus over the past twenty years are: Evidence, Data Science Applications, Applications to Scientific Practice, and Causal Modelling, each of which also has a relatively high value for r^2, and a growth trend that is statistically significant at the 5% level. This supports the conclusion that areas of formal epistemology that allow for greater integration with philosophy of science, the sciences themselves, and areas of law and ethics have been ascendant over the previous two decades. The two topics with highest average salience from 2015-2020 are: i) Probability and Rationality and ii) Causal Modelling. This supports the conclusion that the decreasing prominence of logical approaches in formal epistemology over the past two decades has been offset by an increase in the importance of probability theory as a technique in formal epistemology, especially since many causal modelling techniques popular in formal epistemology, such as those developed by Spirtes et al. (2000), are essentially probabilistic in nature.

Figure 4: Salience of each of the Level 2 topics from 2000-2020.

3.2.3 Level 2

Diachronic trends in topic salience at the second level of abstraction largely cohere with the findings at the first level of abstraction. Figure 4 shows the annual salience of each topic in each year from 2000 to 2020, while Table 2 shows the regression coefficient for the annual trend in the salience of each topic from 2000 to 2020, the value of Pearson's r^2 for each annual trend, and the p-value for each trend. It also shows the average salience of each second-level topic topic over the five-year period from 2015-2020, as a measurement of the recent salience of the topic. As the statistics for the second-level versions of the topics Abstracts not in English and Causal Modelling are very close to the statistics observed for their first-level versions, we can conclude that most of the meaningful coarse-graining when moving to the second level of abstraction involves sorting more fine-grained categories into the coarse-grained categories Belief and Belief Revision and Applications to Science.

Like the data produced by the first-level analysis, the second-level data are also suggestive of a decline in the salience of logical approaches to belief revision within formal epistemology as a whole, and an increase of roughly the same magnitude in work related to scientific applications.

Topic	β	r^2	p-value	2015-20 Avg. Salience
0. Belief and Belief Revision	-4.34×10^{-3}	.54	1.60×10^{-4}	.264
2. Applications to Science	4.03×10^{-3}	.53	1.86×10^{-4}	.401
3. Abstracts not in English	-1.07×10^{-3}	.13	1.16×10^{-1}	.011
4. Causal Modelling	1.43×10^{-3}	.41	1.64×10^{-3}	.078

Table 2: Regression coefficient for the annual salience of second-level topics from 2000-2020, Pearson's r^2 and p-value for each annual salience trend, and the average salience of each second-level topic from 2015-2020.

Topic	β	r^2	p-value	2015-20 Avg. Salience
1	-2.36×10^{-4}	.02	.56	.158
7	-1.14×10^{-4}	.005	.26	.087
14	-6.52×10^{-4}	.10	.16	.056
20	7.33×10^{-6}	5.57×10^{-5}	.97	.015

Table 3: Regression coefficient for the annual salience of uninterpreted Level 1 topics from 2000-2020, and Pearson's r^2 for each annual salience trend, and the average salience of each uninterpreted topic from 2015-2020.

At the same time, the fact that the coarse-graining process retains causal modelling as a distinct topic at the second level of abstraction suggests that causal modelling is a distinct and specialized sub-topic within the broader formal epistemology literature.

3.3 Uninterpreted Topics

One might worry that the strength of my conclusions in this section are undermined to some degree by the presence of four uninterpreted topics, i.e., Topics 1, 7, 14, and 20 at Level 1 and Topic 1 at Level 2. Some of the words with relatively high probability in these topics, it might be argued, are more evocative of logical approaches than probabilistic approaches in formal epistemology. Examples of such words include 'information', 'set', 'conditions', and 'relation' in Topic 1 and 'foundational' and 'disjunction' in Topic 14. While such worries speak to both the inherent subjectivity of topic interpretation, and the drawbacks of the SBM approach discussed above, I do not believe that the uninterpreted topics threaten my overall analysis here. This is because none of these topics show particularly strong positive or negative trends in their average overall salience from 2000-2020, as compared to many topics with a clear interpretation. As shown in Table 3, none of the uninterpreted Level 1 topics have a regression coefficient with an absolute value greater than .001, whereas six interpreted Level 1 topics have a regression coefficient with an absolute value greater than .001. No uninterpreted topic at Level 1 has an r^2 value greater than .1, whereas nine interpreted topics have an r^2 value greater than .1, and none of the Level 1 uninterpreted topics show an annual average salience trend that is significant at the 5% level. This suggests that, in general, the uninterpreted topics do not exhibit clear trends with respect to the growth or decline of their average annual salience during the period from 2000-2020. Moreover, the two Level 1 topics that assign high probability to words arguably associated with epistemic logic, Topics 1 and 14, do still show declines over the period of study, in keeping with our overall conclusions above.

As for Level 2, the single uninterpreted topic, Topic 1, has a regression coefficient of -3.50×10^{-4}, which has lower absolute value than any interpreted topic at Level 2, an r^2 of .02, which is also lower than any interpreted topic at Level 2, and a p-value of .57. So, at this level of analysis we also find that the uninterpreted topics do not show clear trends of either growth or decline in the period under study.

Nevertheless, one may remain troubled by the high levels of overall salience of uninterpreted topics. In response, I note that one must keep in mind that the annual salience of a topic for a given year is the average of the probability that each abstract published in that year contains a word generated by sampling from that topic. An abstract may, with high probability, sample from a given topic, despite that topic not being especially meaningful. To illustrate, consider a topic that assigns highest probability to the words 'the', 'it' and 'and'. Many abstracts published in a given year might be best represented as generating words by sampling, with high probability, from this topic. However, one would not thereby infer that the topic is especially meaningful within the corpus, since it contains stop words common to all abstracts. Thus, the relatively high salience of some excluded topics does not thereby imply that, by excluding them from our analysis, we are missing out on important conclusions regarding diachronic changes in the topics covered in formal epistemology.

4 Discussion

On 8 February, 2021, Jonathan Weisberg, formal epistemologist and Associate Professor at University of Toronto, tweeted:

> [P]repping my modal logic lectures and i'm stuck at "section 1: motivations" (Weisberg, 2021).

Brian Weatherson, another formal epistemologist and Professor at University of Michigan, responded with the following:

> Remember when formal epistemology meant epistemic logic, not everything in probabilities? That stuff had value, and it's worth knowing the language it's written in (Weatherson, 2021b).

To summarize the exchange, Weisberg suggests (in a lighthearted way) that there is no point to teaching and learning modal logic, and Weatherson responds by saying that teaching modal logic to students is worthwhile if only so that those students can understand the literature in earlier formal epistemology, which he implies primarily consisted in the application of logic, especially modal logic, to epistemological problems. Moreover, Weatherson explicitly states that this older, logical approach to formal epistemology has been replaced by an approach more heavily centered around probability theory.

I take this informal exchange between two leading practitioners of formal epistemology to be indicative of a broader anecdotal consensus among members of the formal epistemology community that epistemic logic is somewhat passé as a methodology, whereas probability theory is now the standard methodology in the field. The empirical results collected in this paper corroborate this anecdotal consensus. They do this by demonstrating: i) the greater overall salience of probability-based topics, as compared to logical topics, in formal epistemology abstracts, ii) the rise of probability-based methods and areas of interest such as causal modelling, and iii) the decline in logic-based methods from their prominence in the early 2000s.

If, as the data presented here suggest, there has been a shift towards probability theory as the primary methodology of formal epistemology, then this may have implications for how the broader philosophy curriculum ought to be designed. Graduate (i.e., PhD) programs in Philosophy presumably aim in part to prepare students to contribute to growing areas of philosophical research. The data collected here suggest that formal epistemology is such a field. Nevertheless, a course in formal logic is a cornerstone of graduate training in philosophy. By contrast, if probability theory is formally taught at all in the graduate philosophy curriculum, it remains an ancillary topic, despite the demonstrable salience of probability in one of the primary formal fields of contemporary analytic philosophy. Thus, there is an argument to be made that probability theory, causal modelling, and other probabilistic methods should play a more central role than they currently do in graduate education in philosophy, perhaps as an alternative to or substitute for requirements in logic.

At the undergraduate level, while preparing students for a research career in philosophy is not necessarily the central goal of curricular design, there remains a desire to impart broadly useful skills to undergraduate students as part of their philosophy education. Probability theory is a core component of machine learning, data science, and other skill sets deployed in a wide variety of academic, industrial, governmental, and philanthropic spaces in which undergraduate philosophy students may work after graduation. As such, there is a case to be made for bringing undergraduate training *and* graduate training in philosophy into line with trends in formal epistemology, and increasing the amount of probability theory in the philosophy curriculum at all levels. For more on the increasing salience of probability theory in philosophy, and the importance of including probability theory in the philosophy curriculum, see Staffel (ms).

To be clear, nothing in the previous paragraph should be read as denigrating the philosophical or broader intellectual value of logic, including epistemic logic. Indeed, as discussed in the previous section, the data do *not* indicate that epistemic logic is a dead topic in formal epistemology; the average annual salience for 2015-2020 of Logical Approaches to Belief Revision (.058) and Logical Approaches to Knowledge (.057) are both above the median of .051 for all Level 1 topics. Thus, the data are best understood as suggesting that epistemic logic is now one topic among many rather than the central topic in the field. Moreover, despite the reservations expressed by Hansson (2017) and quoted above, there has been a spate of recent work aimed at unifying probabilistic and non-probabilistic approaches to modelling the norms of both partial and full belief, e.g., Buchak (2014), Ross and Schroeder (2014), Pettigrew (2015), Jackson (2018), and Collins (2020). Nevertheless, there remains a case to be made that, given limited resources of time and energy, a student interested in formal epistemology would be better served, or at least as well served, by dedicated coursework in probability theory, rather than in logic.

Additionally, as mentioned above, the rise in the salience of probability theory in formal epistemology has been accompanied by an increase in topics that concern the overlap between formal epistemology and both philosophy of science and the sciences themselves. There is a sense in which these two trends can be read as part of a common shift in the conceptual orientation of formal epistemology. Probabilistic techniques are used widely across the sciences, whereas techniques from epistemic logic are rarely explicitly used outside of philosophy, logic, and some more theoretical areas of computer science. Thus, one can understand the increased salience of probabilistic techniques within formal epistemology as part of a broader trend in the field towards naturalism and integration with the natural and social sciences, as well as philosophy of science.

5 Conclusion

The preceeding study suggests several intriguing avenues for further research. First, it is possible that the data presented here do not reflect a shift in the overall importance of probability theory within the philosopher's toolkit broadly construed, but instead only reflect a shift in the semantics of the term 'formal epistemology'. It may be that papers written in the past, even though they use tools from probability theory to analyze epistemic problems, are less likely to be categorized as pertaining to formal epistemology than similar papers written more recently. As such, it may be that what is shown in our data is not so much a rise in the salience of probabilistic methods for the study of epistemic norms, but rather a bias towards classifying more recent philosophy papers that use probabilistic techniques as papers in formal epistemology. Testing this hypothesis would require further analysis of the philosophical corpus, going beyond those papers tagged on PhilPapers as pertaining to formal epistemology.

Second, I have not been concerned here with the role of journals in the dissemination of formal epistemology, or the role that journals and other publication venues might play in shaping the salience of various topics in formal epistemology over time. This choice on my part is due to the heterogeneity of the publication venues in the formal epistemology corpus compiled here, which spans journals, books, and edited volumes. By contrast, other topic modelling studies of philosophy, such as Malaterre et al. (2019) and Malaterre et al. (2020), have defined their corpus by specifically focusing on a set of journals, and studying diachronic trends in the topics published by those journals. This is difficult when studying formal epistemology because there are, to my knowledge, no major journals devoted solely or even primarily to formal epistemology, so that it would be difficult to see how to build such a corpora without intensive manual sorting of the catalogue of a large body of journals to identify a corpus of formal epistemology papers published in various journals over a given time frame.

Finally, a key limitation of this analysis is that the corpus analyzed here is composed solely of the text of formal epistemology abstracts, rather than the text of full papers. Important details of the content of formal epistemology papers may not be included in the abstracts of those papers, such that the analysis presented here does not capture the true distribution of topics covered in the formal epistemology literature. While I do take the abstracts of most papers to be a decent proxy for determining the topics covered in that paper, this is indeed a limitation of the current approach, especially since norms about how much content, and what sort of content, to include in the abstract of a paper as opposed to in the main text of a paper may have changed over the time period studied here. For this reason, it is clear that one could gain a fuller picture of trends in the topics covered in the formal epistemology literature by conducting a more expansive study in which the entire texts of journal articles, books, and book chapters in formal epistemology are digitized and analyzed using topic modeling techniques.

Despite these limitations of the present study and suggestions for further research, I take it to be a strength of the results presented here that they cohere with the anecdotal belief often expressed in formal epistemology circles (e.g., in the tweets quoted above), that the field is turning away from epistemic logic, and towards probability theory and further integration with the sciences. Moreover, these results provide further proof-of-concept for the use of SBM techniques in text-mining, science-of-science, and humanities analytics applications. This in turn is suggestive of a common conceptual thread between topic modeling and the kinds of network-based community-detection tasks for which SBM techniques were first developed by Ball et al. (2011). My hope is that in the future, more research in topic modelling, including work in empirical meta-philosophy, will use these newer methods.

References

C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of symbolic logic*, pages 510–530, 1985.

B. Ball, B. Karrer, and M. E. Newman. Efficient and principled method for detecting communities in networks. *Physical Review E*, 84(3):036103, 2011.

D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of machine Learning research*, 3:993–1022, 2003.

L. Buchak. Belief, credence, and norms. *Philosophical studies*, 169(2):285–311, 2014.

J. Collins. Simple belief. *Synthese*, 197(11):4867–4885, 2020. doi: 10.1007/s11229-018-1746-3.

S. C. Fletcher, J. Knobe, G. Wheeler, and B. A. Woodcock. Changing use of formal methods in philosophy: Late 2000s vs. late 2010s. *Synthese*, forthcoming.

M. Gerlach, T. P. Peixoto, and E. G. Altmann. A network approach to topic models. *Science advances*, 4(7):eaaq1360, 2018.

T. L. Griffiths, M. Steyvers, D. M. Blei, and J. B. Tenenbaum. Integrating topics and syntax. In *NIPS*, volume 4, pages 537–544, 2004.

S. O. Hansson. Logic of Belief Revision. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition, 2017.

E. Jackson. Belief, credence, and evidence. *Synthese*, pages 1–20, 2018.

C. Malaterre, J.-F. Chartier, and D. Pulizzotto. What is this thing called philosophy of science? a computational topic-modeling perspective, 1934–2015. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 9(2):215–249, 2019.

C. Malaterre, F. Lareau, D. Pulizzotto, and J. St-Onge. Eight journals over eight decades: a computational topic-modeling approach to contemporary philosophy of science. *Synthese*, pages 1–41, 2020.

T. P. Peixoto. Model selection and hypothesis testing for large-scale network models with overlapping groups. *Physical Review X*, 5(1):011033, 2015.

T. P. Peixoto. Nonparametric bayesian inference of the microcanonical stochastic block model. *Physical Review E*, 95(1):012317, 2017.

R. Pettigrew. Accuracy and the credence-belief connection. *Philosopher's Imprint*, 15(16), 2015.

J. Ross and M. Schroeder. Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, 88(2):259–288, 2014.

P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, prediction, and search*. MIT press, 2000.

J. Staffel. Probability without tears. ms. URL https://philpapers.org/rec/STAPWT.

B. Weatherson. A history of philosophy journals. volume 1: evidence from topic modelling, 1876-2013, 2021a. URL http://www-personal.umich.edu/~weath/lda/.

B. Weatherson. Tweet on 8 february, 2021b. URL https://twitter.com/bweatherson/status/1358813210741264386?s=20.

J. Weisberg. Tweet on 8 february, 2021. URL https://twitter.com/jweisber/status/1358809726079885315?s=20.

Appendix A. LDA and SBM Approaches to Topic Modeling

This appendix presents a more detailed comparison of the LDA and SBM approaches to topic modelling than would be appropriate in the main body of the paper, although there is some conceptual overlap between the presentation of SBM given here and that given in Section 2.2. Specifically, I give an exegetical overview of the LDA and SBM approaches to topic modelling, highlighting the differences between the two approaches where appropriate.

A1. LDA Topic Modelling

As is done above, consider a corpus $\mathcal{D} = \{d_1, \ldots, d_z\}$ of documents in which each document is a sequence of words $d_i = (w_{i,1}, \ldots, w_{i,m_i})$ from a vocabulary $\mathcal{W} = \{w_1, \ldots, w_m\}$. In an LDA topic model, we must also pre-specify a number of topics K to be discovered. We then assume that each document d_i in the corpus is generated via the following process (Blei et al. 2003, p. 996):

1. Obtain the number m_i of words in the document by sampling from a Poisson distribution with parameter ξ.

2. Obtain a document-specific probability distribution $p(\cdot|d_i)$ over topics $\mathcal{T}$ by sampling from a Dirichlet distribution with $\boldsymbol{\alpha}$ as its concentration parameter and K as the parameter fixing the number of elements in $\mathcal{T}$. Note that the Dirichlet distribution is a unimodal distribution, meaning that in an LDA the distribution over topics $p(\cdot|d_i)$ is always "either concentrated around the mean value, or spread away uniformly from it towards pure components" (Gerlach et al. 2018, p. 5).

3. To determine each word $w_{i,j}$ in the document d_i:

   (a) Choose a topic t_j by sampling from a single-trial multinomial distribution that takes as its parameter the topic mixture $p(\cdot|d_i)$.

   (b) Choose a word $w_{i,j}$ by sampling from a single-trial multinomial distribution whose parameter is a mixture of the topic t_j and a distribution $\boldsymbol{\beta}$.

Inferring an LDA topic model involves finding settings for the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that produce topics and topic distributions that render the corpus $\mathcal{D}$ maximally likely while avoiding over-fitting, where parameters are said to avoid over-fitting if they not only render the corpus $\mathcal{D}$ highly likely but also render corpora that are similar but not identical to $\mathcal{D}$ sufficiently likely. While solving this inference problem exactly is computationally intractable, there are several algorithms for efficient approximation of suitable LDA parameters for a given input corpus.

A2. SBM Topic Modelling

As stated in the main body of the paper, an SBM represents a corpus as a multigraph in which nodes represent either documents or words, such that all documents and words in the corpus are represented, and the adjacency matrix $\mathbf{A}$ is such that if two nodes n_i and n_j both represent documents or both represent words, then $A_{ij} = 0$. Otherwise, A_{ij} represents the number of times that word w_j appears in document d_i, or the number of times word w_i appears in document d_j, depending on whether n_i and n_j represent a word or a document, respectively. We group the set of nodes $\mathcal{N}_{\mathcal{W}}$ representing words in the vocabulary into v groups $\mathcal{N}_{\mathcal{W}}^{\uparrow} = \{N_{\mathcal{W}}^{\uparrow 1}, \ldots, N_{\mathcal{W}}^{\uparrow v}\}$ (in Gerlach et al.'s presentation of their technique they also group the documents, but I dispense with this part of their model here as it is not used in my analysis). This yields a high-level adjacency matrix $\mathbf{A}^{\uparrow}$ such that whenever node $n_i^{\uparrow}$ represents a document and $n_r^{\uparrow}$ represents a document group, the matrix entry $A_{ir}^{\uparrow}$ represents the number of words in $N_{\mathcal{W}}^{\uparrow r}$ included in d_i. Inferring an SBM topic model involves finding a set of groups of word nodes $\mathcal{N}_{\mathcal{W}}^{\uparrow}$ and the upper-level adjacency matrix $\mathbf{A}^{\uparrow}$ such that the observed lower-level adjacency matrix $\mathbf{A}$ is maximally likely while avoiding over-fitting.

To describe this inference process, I follow the presentation in Peixoto (2017) and Gerlach et al. (2018), with some adjustments for the topic modelling context. I begin by introducing some notation. Let $\mathcal{N}_{\mathcal{W}}^{l}$ be the grouping of word nodes at the l-th level of abstraction, so that $\mathcal{N}_{\mathcal{W}}^{0} = \mathcal{N}_{\mathcal{W}}$, $\mathcal{N}_{\mathcal{W}}^{1} = \mathcal{N}_{\mathcal{W}}^{\uparrow}$, $\mathcal{N}_{\mathcal{W}}^{2} = \mathcal{N}_{\mathcal{W}}^{\uparrow\uparrow}$, and so on. Let $\mathbf{A}^{l}$ be the adjacency matrix for the multigraph with nodes $\mathcal{N}_{D} \cup \mathcal{N}_{\mathcal{W}}^{l}$. Let $\{\mathcal{N}_{\mathcal{W}}^{l}\}$ denote the set of word groupings at all levels of abstraction $l \in \{0, \ldots, L\}$. We aim to find the set of world groupings $\{\mathcal{N}_{\mathcal{W}}^{l}\}$ that maximizes the joint probability $P(\mathbf{A}^{0}, \{\mathcal{N}_{\mathcal{W}}^{l}\})$. To do this, we aim to *minimize* the description length Σ, which is defined via the equation:

$$\Sigma = -\log_2 P(\mathbf{A}^{0}, \{\mathcal{N}_{\mathcal{W}}^{l}\}). \tag{4}$$

To do this, we note first that $P(\mathbf{A}^{0}, \{\mathcal{N}_{\mathcal{W}}^{l}\})$ can be expanded as follows (recall that $\mathbf{k}$ is still the degree vector for the lowest-level multigraph $\mathcal{G}$):

$$P(\mathbf{A}^0, \{\mathcal{N}_{\mathcal{W}}^l\}) = P(\mathbf{A}^0|\mathbf{k}, \mathbf{A}^1, \mathcal{N}_{\mathcal{W}}^1)P(\mathbf{k}|\mathbf{A}^1, \mathcal{N}_{\mathcal{W}}^1)P(\mathcal{N}_{\mathcal{W}}^1)\prod_{l=1}^{L-1} P(\mathbf{A}^l|\mathbf{A}^{l+1}, \mathcal{N}_{\mathcal{W}}^{l+1})P(\mathcal{N}_{\mathcal{W}}^{l+1}). \tag{5}$$

We then proceed by defining each of the components of this expansion of $P(\mathbf{A}^0, \{\mathcal{N}_{\mathcal{W}}^l\})$.

We begin with each $P(\mathcal{N}_{\mathcal{W}}^l)$, which can be defined as follows:

$$P(\mathcal{N}_{\mathcal{W}}^l) = \prod_{r=1}^{|\mathcal{N}_{\mathcal{W}}^l|} \frac{|N_{\mathcal{W}}^{l,r}|!}{|\mathcal{N}_{\mathcal{W}}^{l-1}|!} \binom{|\mathcal{N}_{\mathcal{W}}^{l-1}|-1}{|\mathcal{N}_{\mathcal{W}}^l|-1}^{-1} \frac{1}{|\mathcal{N}_{\mathcal{W}}^{l-1}|}. \tag{6}$$

where $N_{\mathcal{W}}^{l,r}$ is the r-th element of the level-l grouping $\mathcal{N}_{\mathcal{W}}^l$. The RHS of this equation is the proportion of possible groupings of word nodes at level l that are consistent with the size of the grouping $\mathcal{N}_{\mathcal{W}}^{l-1}$. Sampling each grouping of the vocabulary from this hierarchical prior at each level allows us to avoid the unimodality constraint on topic distributions imposed under the LDA framework.

Next, we move to the probabilities of the form $P(\mathbf{A}^l|\mathbf{A}^{l+1}, \mathcal{N}_{\mathcal{W}}^{l+1})$, which can be defined as follows:

$$P(\mathbf{A}^l|\mathbf{A}^{l+1}, \mathcal{N}_{\mathcal{W}}^{l+1}) = \prod_{i=1}^{z} \prod_{r=1}^{|\mathcal{N}_{\mathcal{W}}^{l+1}|} \left(\!\!\binom{|N_{\mathcal{W}}^{l+1,r}|}{A_{ir}^{l+1}}\!\!\right)^{-1} \prod_{r=1}^{|\mathcal{N}_{\mathcal{W}}^{l+1}|} \Big(\frac{|N_{\mathcal{W}}^{l+1,r}|(|N_{\mathcal{W}}^{l+1,r}|+1)}{2}\Big)^{-1} \tag{7}$$

where $\left(\!\binom{x}{y}\!\right) = \binom{x+y-1}{y}$. The RHS of this equation is a fraction with 1 as its numerator and the number of adjacency matrices $\mathbf{A}^l$ consistent with the higher-level adjacency matrix and higher-level grouping $\mathbf{A}^{l+1}$ and $\mathcal{N}_{\mathcal{W}}^{l+1}$ as its denominator.

Moving to the probability $P(\mathbf{k}|\mathbf{A}^1, \mathcal{N}_{\mathcal{W}}^1)$, let η_k^r be the number of nodes in $N_{\mathcal{W}}^{1,r}$ with degree $k \in \{1, \ldots, \kappa\}$. We define the probability as follows:

$$P(\mathbf{k}|\mathbf{A}^1, \mathcal{N}_{\mathcal{W}}^1) = \prod_{r=1}^{|\mathcal{N}_{\mathcal{W}}^1|} \frac{\prod_{k=1}^{\kappa} \eta_k^r!}{|N_{\mathcal{W}}^{1,r}|!} q\Big(\sum_{i=1}^{z} A_{ir}^1, |N_{\mathcal{W}}^{1,r}|\Big)^{-1} \tag{8}$$

where $q(m, n)$ is the number of ways of writing m as the sum of n positive integers. Although $q(m, n)$ is computationally intractable for larger m and n, efficient approximations are given in Peixoto (2017, p. 6). The RHS of this equation is the ratio between the number of lowest-level multigraphs with degree vector $\mathbf{k}$ and the number of lowest-level multigraphs that are consistent with the higher-level adjacency matrix and grouping $\mathbf{A}^1$ and $\mathcal{N}^1_{\mathcal{W}}$.

Finally, to define $P(\mathbf{A}^0|\mathbf{k}, \mathbf{A}^1, \mathcal{N}^1_{\mathcal{W}})$, we begin by calculating the number $\Omega(\mathbf{A}^1, \mathcal{N}^1_{\mathcal{W}})$ of lower-level multigraphs consistent with $\mathbf{A}^1$ and $\mathcal{N}^1_{\mathcal{W}}$:

$$\Omega(\mathbf{A}^1, \mathcal{N}^1_{\mathcal{W}}) = \frac{\prod_{r=1}^{|\mathcal{N}^1_{\mathcal{W}}|} (\sum_{i=1}^{z} A^1_{ir})!}{\prod_{i=1}^{z} \prod_{r}^{|\mathcal{N}^1_{\mathcal{W}}|} A^1_{ir}!} \tag{9}$$

For any given lowest-level adjacency matrix $\mathbf{A}^0$ and degree vector $\mathbf{k}$, the number $\Xi(\mathbf{A}^0, \mathbf{k})$ of higher-level multigraphs consistent with $\mathbf{A}^0$ and $\mathbf{k}$ is given by the equation:

$$\Xi(\mathbf{A}^0, \mathbf{k}) = \frac{\prod_{i=1}^{n+m} k_i!}{\prod_{j=1}^{n+m} \prod_{i<j} A^0_{ij}!} \tag{10}$$

Thus, the probability of observing a multigraph with the lower-level adjacency matrix $\mathbf{A}^0$, given parameter settings $\mathcal{N}^{\uparrow}_{\mathcal{W}}$, $\mathbf{k}$, and $\mathbf{A}^1$, is given by the following ratio:

$$P(\mathbf{A}^0|\mathbf{k}, \mathbf{A}^1, \mathcal{N}^1_{\mathcal{W}}) = \frac{\Xi(\mathbf{A}, \mathbf{k})}{\Omega(\mathbf{A}^1, \mathcal{N}^1_{\mathcal{W}})}. \tag{11}$$

Taking stock, we see that the RHS of the equations for all the component probabilities in the product in the RHS of (5) can be calculated using either the observed lowest-level adjacency matrix $\mathbf{A}^0$ for the corpus, or from the set of groupings $\{\mathcal{N}^l_{\mathcal{W}}\}$. This entails in turn that one can infer the SBM at all levels of abstraction by finding the set of groupings $\{\mathcal{N}^l_{\mathcal{W}}\}$ that minimizes the description length Σ of the probability $P(\mathbf{A}^0, \{\mathcal{N}^l_{\mathcal{W}}\})$. In Gerlach et al. (2018), and in the code used in the analysis in this paper, this is done through Markov Chain Monte Carlo simulation. Said inference algorithms do *not* require pre-specification of the cardinality of the grouping of the word nodes at each level of abstraction. Instead, this can learned from data; this is another advantage of the SBM methodology over LDA methods.

Appendix B. Formal Epistemology Topics

B1. Level 1

In this appendix, I list the twenty words with highest probability according to each of the twenty-one first-level topics discovered in this analysis. The probability of each word in that topic is given in parenthesis next to the word in question. Note that for the purpose of this study, topics 2 and 13, 16 and 17, and 18 and 19 were deemed to be sufficiently similar that each pair was combined into a single topic. Note also that in topic 19, the five words listed are the only five words that had positive probability with respect to that topic.

0. Evidence
   - evidence (.129)
   - disagreement (.021)
   - response (.018)
   - correct (.015)
   - right (.014)
   - thought (.014)
   - law (.014)
   - think (.013)
   - always (.013)
   - procedure (.013)
   - positive (.011)
   - procedures (.011)
   - lead (.011)
   - seem (.011)
   - consensus (.011)
   - majority (.010)
   - legal (.009)
   - act (.008)
   - significance (.008)
   - confidence (.008)
1. No Clear Interpretation
   - information (.024)
   - results (.018)
   - set (.016)
   - terms (.013)
   - view (.012)
   - important (.011)
   - notion (.010)
   - conditions (010)
   - principle (.009)
   - certain (.009)
   - second (.008)
   - number (.008)
   - states (.008)
   - standard (.008)
   - relation (.008)
   - order (.008)
   - kind (.008)
   - role (.008)
   - often (.007)
   - rather (.007)
2. Logical Approaches to Belief Revision I
   - belief (.214)
   - epistemic (.121)
   - beliefs (.093)
   - revision (.082)
   - change (.067)
   - agent (.052)
   - agents (.046)
   - state (.037)
   - AGM (.028)
   - postulates (.017)

- iterated (.017)
- operators (.014)
- doxastic (.014)
- epistemically (.013)
- minimal (.011)
- operator (.011)
- update (.010)
- base (.010)
- operations (.009)
- inconsistent (.009)

3. Logical Approaches to Knowledge

   - knowledge (.062)
   - logic (.045)
   - common (.031)
   - logical (.025)
   - truth (.019)
   - system (.017)
   - context (.014)
   - normative (.013)
   - attitudes (.012)
   - without (.011)
   - inquiry (.011)
   - dynamic (.011)
   - language (.011)
   - action (.009)
   - level (.009)
   - available (.009)
   - prior (.008)
   - appropriate (.008)
   - logics (.008)
   - full (.008)

4. Probability and Rationality

   - probability (.044)
   - rational (.042)
   - probabilistic (.026)
   - rationality (.022)
   - group (.020)
   - principles (.018)
   - true (.017)
   - functions (.017)
   - rule (.016)
   - individual (.016)
   - propositions (.016)
   - rules (.016)
   - aggregation (.016)
   - function (.015)
   - justification (.015)
   - thesis (.015)
   - probabilities (.015)
   - proposition (.015)
   - credences (.014)
   - norms (.013)

5. Inductive Risk

   - error (.014)
   - risk (.013)
   - core (.011)
   - practical (.010)
   - Hume (.010)
   - Glymour (.009)
   - nonmonotonic (.008)
   - published (.008)
   - failure (.008)
   - fails (.008)
   - drawing (.007)
   - parameter (.007)
   - program (.007)
   - warrant (.007)
   - deductive (.007)
   - chance (.006)
   - agency (.006)
   - QCA (.006)
   - efficacy (.006)
   - reading (.005)

6. Applications to General PhilSci

   - science (.043)

- explanation (.028)
- scientific (.027)
- Causal (.017)
- philosophers (.013)
- explanatory (.013)
- explanations (.012)
- practice (.010)
- author (.010)
- psychology (.009)
- thinking (.009)
- ideas (.008)
- psychological (.008)
- cognition (.008)
- power (.007)
- authors (.007)
- diversity (.007)
- domain (.007)
- tools (.007)
- property (.006)

7. No Clear Interpretation
   - social (.024)
   - question (.022)
   - research (.021)
   - philosophical (.018)
   - philosophy (.018)
   - well (.016)
   - decision (.015)
   - cognitive (.014)
   - questions (.013)
   - empirical (.012)
   - book (.012)
   - idea (.012)
   - concept (.011)
   - among (.011)
   - alternative (.011)
   - interpretation (.011)
   - value (.011)
   - choice (.010)
   - issues (.010)
   - explain (.009)

8. Abstracts not in English
   - de (.043)
   - la (.023)
   - que (.012)
   - en (.011)
   - un (.009)
   - die (.009)
   - John (.008)
   - der (.007)
   - und (.007)
   - las (.006)
   - es (.005)
   - el (.005)
   - una (.004)
   - se (.004)
   - testimonio (.004)
   - los (.004)
   - des (.004)
   - como (.003)
   - del (.003)
   - En (.003)

9. Judgment
   - judgments (.027)
   - people (.024)
   - features (.016)
   - judgment (.015)
   - influence (.014)
   - multiple (.013)
   - actual (.013)
   - networks (.012)
   - moral (.012)
   - factors (.012)
   - temporal (.011)
   - found (.011)
   - behavior (.011)

- considerations (.010)
- ability (.010)
- likely (.010)
- identify (.010)
- children (.009)
- across (.009)
- network (.008)

10. Coherence and Conditionalization

    - coherence (.063)
    - Bayesian (.045)
    - measure (.027)
    - version (.022),
    - measures (.020)
    - changes (.016)
    - coherent (.013)
    - mind (.010)
    - external (.009)
    - intuitive (.009)
    - illustrate (.009)
    - Dutch (.009)
    - foundations (.009)
    - sequence (.008)
    - updating (.007)
    - acceptance (.007)
    - pragmatic (.007)
    - internal (.007)
    - conditionalization (.007)
    - components (.006)

11. Applications to Scientific Modelling

    - model (.102)
    - models (.066)
    - inference (.046)
    - possible (.033)
    - methods (.031)
    - method (.025)
    - conditional (.023)
    - world (.022)
    - time (.022)
    - relevant (.020)
    - modeling (.020)
    - assumption (.017)
    - inferences (.016)
    - constraints (.016)
    - allows (.015)
    - consistent (.014)
    - test (.014)
    - relevance (.013)
    - strong (.012)
    - apply (.012)

12. Hypothesis Testing and Confirmation

    - condition (.032)
    - hypothesis (.024)
    - confirmation (.020)
    - Bayes (.014)
    - Markov (.014)
    - assessment (.014)
    - evaluation (.014)
    - prediction (.013)
    - performance (.012)
    - opinion (.012)
    - correlations (.012)
    - quantum (.011)
    - nets (.011)
    - discussions (.010)
    - deterministic (.008)
    - systematic (.008)
    - counterexamples (.007)
    - universal (.007)
    - intended (.007)
    - mechanics (.007)

13. Logical Approaches to Belief Revision II

    - contraction (.054)
    - sets (.029)
    - class (.019)
    - partial (.017)

- meet (.016)
- conditionals (.016)
- represented (.016)
- sentences (.015)
- & (.013)
- et (.013)
- Gärdenfors (.012)
- equivalent (.012)
- entrenchment (.011)
- sentence (.011)
- operation (.011)
- al (.011)
- semantic (.010)
- accepted (.010)
- obtained (.010)
- theorems (.010)

14. No Clear Interpretation

    - informal (.003)
    - foundational (.003)
    - price (.002),
    - 2013 (.002)
    - written (.002)
    - suggestion (.002)
    - Paul (.001)
    - disjunction (.001)
    - Keith (.001)
    - essence (.001)
    - advancing (.001)
    - Time (.001)
    - defeasibility (.001)
    - machinery (.001)
    - substitute (.001)
    - depart (.001)
    - pervasive (.001)
    - raising (.001)
    - Basic (.001)
    - thoroughly (.001)

15. Data Science Applications

    - data (.034)
    - structure (.030)
    - cause (.029)
    - effect (.028)
    - variables (.028)
    - effects (.026)
    - learning (.024)
    - relations (.022)
    - studies (.020)
    - assumptions (.019)
    - statistical (.016)
    - structural (.015)
    - experimental (.012)
    - structures (.012)
    - outcome (.011)
    - discovery (.010)
    - selection (.010)
    - independence (.010)
    - distribution (.010)
    - underlying (.010)

16. Applications to Scientific Practice I

    - treatment (.033)
    - functional (.016)
    - select (.014)
    - personal (.013)
    - brain (.013)
    - confirm (.012)
    - policies (.011)
    - toward (.011)
    - agree (.010)
    - sharing (.009)
    - balance (.009)
    - formats (.009)
    - Find (.008)
    - asked (.008)
    - estimation (.008)
    - robust (.008)

- simulation (.008)
- articles (.008)
- format (.007)
- processing (.007)

17. Applications to Scientific Practice II
    - mental (.0470)
    - scientists (.044)
    - difference (.024)
    - disease (.023)
    - physical (.022)
    - challenges (.016)
    - pooling (.015)
    - already (.014)
    - helps (.014)
    - caused (.013)
    - seek (.013)
    - nt (.012)
    - intermediate (.010)
    - try (.010)
    - communication (.010)
    - looks (.010)
    - fixed (.010)
    - needs (.010)
    - assigned (.009)
    - care (.009)

18. Causal Modelling I
    - causation (.064)
    - causes (.030)
    - mechanisms (.025)
    - counterfactual (.022)
    - causality (.022)
    - events (.019)
    - sciences (.019)
    - semantics (.018)
    - mechanism (.018)
    - relationships (.016)
    - interventionist (.014)
    - counterfactuals (.013)
    - Lewis (.011)
    - event (.011)
    - dependence (.010)
    - modelling (.008)
    - policy (.009)
    - Pearl (.008)
    - definitions (.008)
    - trials (.008)

19. Causal Modelling II
    - causal (.875)
    - experiments (.048)
    - interventions (.030)
    - intervention (.028)
    - Cartwright (.010)
    - net (.009)

20. No Clear Interpretation
    - th (.010)
    - pain (.006)
    - variance (.004)
    - Futurium (.003)
    - crisis (.003)
    - Semmelweis (.003)
    - deaths (.003)
    - observe (.003)
    - convincing (.003)
    - creature (.003)
    - continuing (.003)
    - Aufbau (.003)
    - Digital (.003)
    - COVID (.003)
    - reductions (.002)
    - KCIT (.002)
    - neglects (.002)
    - J-revision (.002)
    - implausibility (.002)
    - ca (.002)

B2. Level 2

Here I list the twenty words with highest probability according to each of five second-level topics discovered in this analysis.

0. Belief and Belief Revision
   - belief (.033)
   - evidence (.019)
   - epistemic (.019)
   - beliefs (.014)
   - revision (.013)
   - change (.0102)
   - coherence (.008)
   - agent (.008)
   - agents (.007)
   - contraction (.007)
   - Bayesian (.006)
   - state (.006)
   - AGM (.004)
   - measure (.004)
   - sets (.004)
   - condition (.004)
   - disagreement (.003)
   - version (.003)
   - measures (.003)
   - hypothesis (.003)
1. No Clear Interpretation
   - information (.015)
   - results (.011)
   - set (.010)
   - social (.009)
   - terms (.008)
   - question (.008)
   - research (.008)
   - view (.008)
   - important (.007)
   - philosophical (.007)
   - notion (.007)
   - philosophy (.007)
   - conditions (.006)
   - well (.006)
   - principle (.006)
   - certain (.005)
   - decision (.005)
   - number (.005)
   - second (.005)
   - cognitive (.005)
2. Applications to Scientific Modelling
   - model (.014)
   - knowledge (.010)
   - models (.009)
   - causation (.007)
   - logic (.007)
   - probability (.007)
   - rational (.007)
   - science (.007)
   - inference (.006)
   - common (.005)
   - data (.005)
   - possible (.004)
   - structure (.004)
   - probabilistic (.004)
   - cause (.004)
   - explanation (.004)
   - effect (.004)
   - methods (.004)
   - scientific (.004)
   - variables (.004)
3. Abstracts not in English
   - de (.043)
   - la (.023)
   - que (.012)

- en (.011)
- un (.009)
- die (.009)
- John (.008)
- der (.007)
- und (.007)
- las (.006)
- es (.005)
- el (.005)
- una (.004)
- se (.004)
- testimonio (.004)
- los (.004)
- des (.004)
- como (.003)
- del (.003)
- En (.003)

4. Causal Modelling
   - causal (.204)
   - experiments (.011)
   - treatment (.011)
   - mental (.010)
   - scientists (.010)
   - interventions (.007)
   - intervention (.007)
   - functional (.005)
   - difference (.005)
   - disease (.005)
   - physical (.005)
   - select (.005)
   - personal (.004)
   - brain (.004)
   - confirm (.004)
   - policies (.004)
   - challenges (.004)
   - toward (.004)
   - agree (.004)
   - pooling (.003)