HOPOS

The International Society for the History of Philosophy of Science

WHAT IS THIS THING CALLED *PHILOSOPHY OF SCIENCE*? A COMPUTATIONAL TOPIC-MODELING PERSPECTIVE, 1934—2015

Author(s): Christophe Malaterre, Jean-François Chartier and Davide Pulizzotto

Source: *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, Fall 2019, Vol. 9, No. 2 (Fall 2019), pp. 215-249

Published by: The University of Chicago Press on behalf of the International Society for the History of Philosophy of Science

Stable URL: https://www.jstor.org/stable/10.2307/26851602

WHAT IS THIS THING CALLED *PHILOSOPHY OF SCIENCE*? A COMPUTATIONAL TOPIC-MODELING PERSPECTIVE, 1934–2015

Christophe Malaterre, Jean-François Chartier, and Davide Pulizzotto

---

What is philosophy of science? Numerous manuals, anthologies, and essays provide carefully reconstructed vantage points on the discipline that have been gained through expert and piecemeal historical analyses. In this article, we address the question from a complementary perspective: we target the content of one major journal in the field—*Philosophy of Science*—and apply unsupervised text-mining methods to its complete corpus, from its start in 1934 until 2015. By running topic-modeling algorithms over the full-text corpus, we identified 126 key research topics that span 82 years. We also tracked those topics' evolution and fluctuating significance over time in the journal articles. Our results concur with and document known and lesser-known episodes in the philosophy of science, including the rise and fall of logic and language-related topics, the relative stability of a metaphysical and ontological questioning (space and time, causation, natural kinds, realism), the significance of epistemological issues about the nature of scientific knowledge, and the rise of a recent philosophy of biology and other trends. These analyses exemplify how computational text-mining methods can be used to provide an empirical large-scale and data-driven perspective on the history of philosophy of science that is complementary to other current historical approaches.

Contact Christophe Malaterre at the Département de philosophie and the Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal, 455 Boulevard René-Lévesque Est, Montréal, Québec H3C 3P8, Canada (malaterre.christophe@uqam.ca). Jean-François Chartier and Davide Pulizzotto are at the Université du Québec à Montréal.

We thank JSTOR for kindly providing access to the complete numerical corpus of *Philosophy of Science*. We also thank Heather Douglas and three anonymous referees for very helpful comments and bibliographical references, as well as the audiences at the 2018 Canadian Philosophical Association meeting (Montreal) and the 2018 Philosophy of Science Association meeting (Seattle), at which earlier versions of the article were presented. Funding from the Canada Foundation for Innovation (grant

---

HOPOS: The Journal of the International Society for the History of Philosophy of Science, vol. 9 (Fall 2019). 2152-5188/2019/0902-0001$10.00.

1. Introduction

Philosophy of science is what philosophers of science do. But what is it that philosophers of science do? Which problems do they focus on? Which topics do they investigate? One way to answer these questions is to turn to manuals, companions or handbooks, and anthologies that offer specific highlights on the subject matter, often pedagogically reconstructed and cleaned up. Another is to appeal to historical studies of the philosophy of science, be they on grand scales or focused on more specific episodes in the constitution and evolution of the discipline. Here, we offer a complementary perspective that focuses on one major journal in the discipline—*Philosophy of Science*—and makes use of computational text-mining methods developed in computer science and the digital humanities. These methods indeed make it possible to comprehensively analyze the semantic content of large corpuses of full-text documents, thereby providing an empirical basis for content-related studies, be they synchronic or diachronic. Such text-mining methods have started to bear interesting results in history and sociology (e.g., Chartier and Meunier 2011; Mimno 2012; DiMaggio et al. 2013; Evans and Aceves 2016; Peirson et al. 2017; Barron et al. 2018), linguistics and the cognitive sciences (e.g., Widdows 2004; Turney and Pantel 2010; Murdock et al. 2017), and philosophy (Buckner et al. 2011; Ramsey and Pence 2016; Hicks and Brister 2018), but they have not yet been used—to the best of our knowledge—to study the history of philosophy of science. In this article, we apply these methods to the complete full-text corpus of *Philosophy of Science* from its very start in 1934 up until 2015 to empirically investigate which research questions philosophers of science have been concerned with and how these questions evolved in the last 82 years. By applying topic-modeling algorithms, we identified 126 key topics that were present in the journal articles during this period. We also analyzed how these topics evolved in significance over time. Our findings concur with well-known episodes in the history of philosophy of science, such as the rise and fall of logical empiricism (1930s–1970s), but they also document other trends, such as the strong appearance of a philosophy of biology in the 1980s and mostly in the first decade of the 2000s as well as the emergence of a significant interest in models and simulations in the 1990s. The article is organized as follows. In section 2, we highlight the text-mining method we followed and the different text-processing and text-analytics algorithms we used. We then present, in section 3, the 126 topics we discovered in the corpus, and we analyze their content. In section 4, we share the results of our dynamic topic modeling and discuss the diachronic patterns

---

34555), the Canada Social Sciences and Humanities Research Council (grant 435-2014-0943), and the Canada Research Chair in Philosophy of the Life Sciences at UQAM (CRC-950-230795) is gratefully acknowledged.

exhibited by the most significant topics found in the journal articles over time. We relate these patterns to known and lesser-known episodes in the history of philosophy of science. We then focus, in section 5, on a few specific topics—scientific explanation, natural selection, and models—to exemplify the type of detailed historical studies that are made possible by text-mining methods: we analyze the specific evolution of these topics in terms of significance in the corpus across the whole period, from 1934 until 2015, and conclude by discussing our findings.

2. Dynamic Topic Modeling: Methodological Prerequisites

Text-mining methods, and most notably topic modeling, are based on the simple fact that, in order to convey meaning, texts use words in specific combinations. In turn, these combinations result in repeated word patterns in texts. As a result, studying the patterns that words form in specific texts can be informative about the semantic content of these texts. As the linguist John R. Firth (1957, 11) stated, "You shall know a word by the company it keeps." Topic-modeling algorithms and methods have been developed to exploit this linguistic phenomenon (e.g., Srivastava and Sahami 2009; Aggarwal and Zhai 2012): they identify words with similar associative patterns in text segments of a given corpus and cluster them into topics, thereby making it possible to identify the thematic content of that corpus. They also make it possible to assess the presence of any topic in any specific document—or set of documents—of that corpus, for instance according to publication time slices. It is this methodological approach that we used to identify the topics of *Philosophy of Science* and their evolution since 1934 up until 2015. More specifically, we used a well-known topic-modeling algorithm based on the Latent Dirichlet Allocation model (LDA; Pritchard et al. 2000; Blei et al. 2003).[1] This algorithm is part of a family of unsupervised statistical machine-learning algorithms for topic discovery in texts. Generally speaking, these algorithms are used to explore corpora for which no specific content-related knowledge is available before using any algorithmic approach. By iteratively assessing probability distributions of words within topics and of topics within documents, they make it possible to retrieve the underlying "latent" topical patterns of the corresponding documents. In what follows, we describe the methodology in more detail (sec. 2.1) and discuss methodological

---

[1] We chose this algorithm for its proven reliability for identifying topics in large corpora (e.g., Cohen Priva and Austerweil 2015; Lynam 2016; Nikolenko et al. 2017; Peirson et al. 2017). As a benchmark, we also used an alternative k-means-based algorithm but found no improvement over the LDA modeling, the latter giving better results in terms, notably, of topic interpretability (see sec. 2.2 about methodological limitations).

limitations (sec. 2.2). For direct access to the results, please skip these sections and go to section 4.

2.1. Methodology

For the present study, we retrieved all articles published in *Philosophy of Science* available on JSTOR from 1934 to 2015. The corpus consists of 4,602 full-text articles, totaling 27,544,926 word occurrences, with an average word count of about 6,000 word occurrences per article (fig. 1).[2] These articles include all regular articles published in *Philosophy of Science* as well as the proceedings of the biennial meetings of the Philosophy of Science Association.[3]

The topic-modeling method we followed comprised four main steps: (1) data preprocessing, (2) data modeling, (3) diachronic topic analysis, and (4) topic interpretation. Data preprocessing consisted in preparing the corpus in a suitable way for the topic-modeling computational analysis. This stage included a lemmatization-based spelling normalization step and a word-filtering step based on a part-of-speech (POS) tagging technique and a word frequency sorting. In our study, the lemmatization was done using the TreeTagger algorithm (Schmid 2013). Because topic modeling is based on word co-occurrence in a corpus, it matters whether one keeps only words that reach a certain frequency threshold. Rare words that occurred in fewer than 50 sentences in the corpus thus were filtered out. Moreover, not all kinds of words are proper candidates for expressing topics in a corpus: words such as determinants, prepositions, or pronouns are irrelevant, and it is crucial to filter them out to reduce noise. We therefore used the Penn TreeBank POS tagging algorithm (Marcus et al. 1993) to identify the morphosyntactic category of every word of the corpus and retained only nouns, verbs, modals, adjectives, adverbs, proper nouns, and foreign words. The data-preprocessing stage resulted in a lexicon of 10,658 distinct words distributed among 976,263 sentences.

The data-modeling stage consisted in first encoding the word distribution into a word × sentence matrix, $W = [w_{ij}]^{M \times N}$, where $M = 10{,}658$ corresponds to the size of our lexicon, $N = 976{,}263$ is the number of sentences, and w_{ij} is the frequency of word i in sentence j. Following Blei et al. (2003), we then applied an LDA algorithm to this matrix, together with a Gibbs sampling method as described in Griffiths and Steyvers (2004).[4] LDA is a generative probabilistic machine-learning method that models topics as probability distributions over words and documents

---

[2] The corpus was downloaded from JSTOR on May 4, 2017.

[3] The proceedings were published separately from their start in 1970 until 1994 and then jointly with the journal. We chose to include the proceedings in the corpus—in addition to the regular research articles—since the proceedings are also a very relevant source, indicating "what philosophers of science do."

[4] The LDA was performed through an API for Python. See https://pythonhosted.org/lda/api.html.

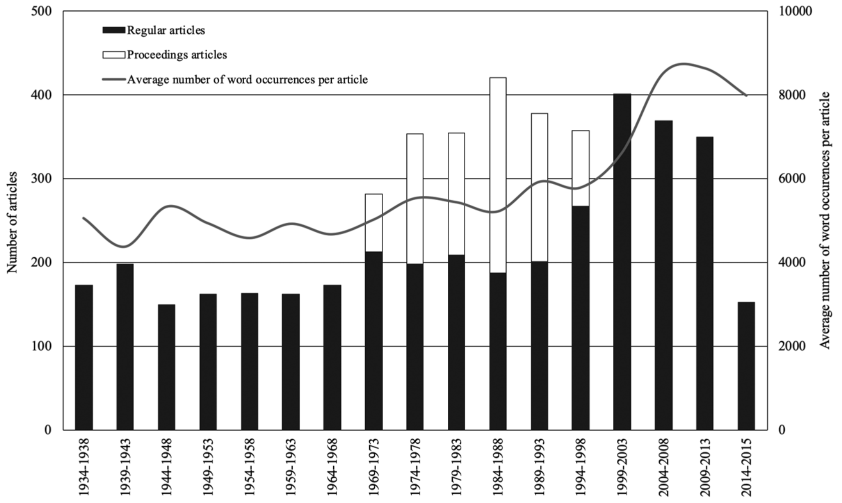


Figure 1. Number of articles per period of 5 years (*left axis*) and average number of words per article (*right axis*).

as probability distributions over topics in a corpus. These probability distributions are assumed to have a sparse Dirichlet prior—hence the name—that encodes the intuition that topics are usually strongly expressed by few words and that documents only express a few topics at a time.[5] LDA is based on the assumption that a corpus is organized according to a hidden "latent" set of topics over its lexicon that generates how words combine into thematic co-occurrence patterns within documents. The goal of LDA is to statistically infer the best possible set of topics that fits these thematic co-occurrence patterns. The method aims at calculating the values of two major probability matrices: $\Phi = [\Pr(w|z)]^{M \times K}$ and $\Theta = [\Pr(z|s)]^{K \times N}$, where M is the size of the lexicon, K is the number of topics, and N is the number of sentences in the total corpus. The matrix Φ indicates which word distributions best express a given topic in the corpus, while the matrix Θ indicates which topics are the most significant in a given sentence. Mathematically, it can be shown that these two probability distributions (which are "latent" or unknown) and the distribution of words within sentences (which is known) are linked. Therefore, taking sentences and words one at a time, one can evaluate which topic to assign to a specific word by using prior estimations of the probability distributions of topics in sentences and of words in topics (and then readjusting the probability distributions over all words and documents). The

---

[5] In this method, two parameters α and β fix the relative selectivity of the distribution over topics and over words respectively (the smaller the value of these parameters, the smaller the number of highly probable topics in a given document or of highly probable words in a given topic).

method is iterative: the topic modeling starts from initial random probability distributions and adjusts (i.e., statistically learns) through Gibbs sampling the two conditional probabilities: (1) the probability $\Pr(w|z)$ that expresses the assignment of a word w to the topic z in the corpus and (2) the probability $\Pr(z|s)$ that corresponds to the proportion of words in a text segment s (in our study, a sentence) assigned to the topic z. Because the probability distributions are constrained by sparse Dirichlet priors, this iterative procedure consists in solving an optimization problem whereby documents should be characterized by as few topics as possible and topics by as few words as possible, as encoded by the Dirichlet parameters. When a convergence criterion is achieved, the method results in populating the two matrices Φ and Θ. One of the main parameters of LDA is the number K of topics, since the model assumes that the dimensionality of the Dirichlet distribution is known and therefore fixed a priori. In the present study, after several runs of trial and error, we chose $K = 200$.[6]

The third stage of the method consisted in inferring the diachronic distributions of topics over the 82 years of *Philosophy of Science*. We chose to split the corpus into 17 periods of 5 years (except for the last period, which included only 2 years). This decision was motivated by the granularity of the analysis we aimed at. A new matrix $\Omega = [\Pr(z|p)]^{K\times T}$ was computed, with $\Pr(z|p)$ being the probability of finding topic z in period p and T corresponding to the total number of periods. $\Pr(z|p)$ was simply obtained by averaging $\Pr(z|s)$ for all sentences s in period p. This matrix thereby indicates which topics were the most significant for each period.

The final stage of the method consisted in interpreting the topics. Technically speaking, topics in a statistical topic model are just probability distributions over words in a corpus. These distributions are called "topics" because they are interpretable as such: by looking at the most likely words assigned to a particular probability distribution in the matrix Φ, one can usually recognize a co-occurrence pattern specific to the expression of a topic and label it with a synthetic predicate. In the case when ambiguities remain in the interpretation of the words assigned to a topic, one can also retrieve, from the matrix Θ, the text segments in which that topic is the most likely in order to confirm the interpretation.

Take, for instance, topic 135: the most likely words assigned to this topic include *explanation*, *hempel*, *law*, *explanandum*, *explanans*, *deductive*, and *cover*

---

[6] We ran different analyses with different values of K: lower values of 50 or 100 led to topics that appeared too broad in scope and difficult to interpret, while higher values of 250 or 300 led to too much redundancy between topics. Other key parameters included fixing the Dirichlet parameter for distribution over topics $\alpha = 0.1$ and fixing the Dirichlet parameter for distribution over words $\beta = 0.01$ (which are fairly standard values). The number of Gibbs iterations was set to 2,000, large enough to reach convergence.

Table 1. List of Most Probable Terms for Topic 135, ‘DN-Explanation’

Top-20 Most Probable Terms	Topic ID
Explanation; hempel; explain; law; model; explanandum; statistical; explanans; explanatory; salmon; account; scientific; deductive; cover; probabilistic; provide; require; event; generalization; particular	135

(table 1). These words form a fairly recognizable co-occurrence pattern that can legitimately be associated with the Hempelian model of explanation (i.e., the deductive-nomological or covering-law model of explanation).[7] This interpretation was confirmed by looking at the most strongly associated articles with that topic, which typically refer to scientific explanation and to the DN-account (table 2). We thereby chose to give the label DN-EXPLANATION to topic 135. We interpreted and labeled all 200 topics modeled from our corpus by using this approach.

From the 200 topics, we identified 126 that we found to be directly relevant for our objective of mapping what philosophy of science is about. Among the 74 discarded topics, 47 appeared to be either too generic or polysemic to be precisely related to any meaningful issue in philosophy of science. We therefore grouped these 47 topics under the label “Jargon” and set them aside.[8] Among the other 27 remaining topics, we typically found editorial noise about HTML or LaTeX code for mathematical expressions and publication status (e.g., with terms such as *note*, *section*, *acknowledgement*, *figure*, etc.). These topics were set aside as well.

2.2. Methodological Discussion

Topic modeling has been shown to be very reliable in identifying topics from corpuses of texts (Griffiths and Steyvers 2004; Griffiths et al. 2007; Blei and Lafferty 2009; DiMaggio et al. 2013). Note again how such topic identification is done in a nonsupervised, data-driven way, that is to say without a priori knowledge of which topics populate the corpus. This does not mean, though, that the researcher does not intervene at all in the methodology: the researcher

---

[7] This interpretation also made sense when compared to other topics that included explanation-related words (see table 3 below).

[8] Although of no direct use for identifying the specific research topics that interest philosophers of science, these jargon topics could, however, provide interesting insights on the generic features of the philosophical discourse.

Table 2. List of Most Strongly Associated Articles for Topic 135, 'DN-Explanation'

Top-20 Most Strongly Associated Articles	Article ID
Hempel, Carl G., Oppenheim, Paul (1948) "Studies in the Logic of Explanation"	504
Grünbaum, Adolf (1962) "Temporally-Asymmetric Principles, Parity between Explanation and Prediction, and Mechanism versus Teleology"	954
Omer, I. A. (1970) "On the D-N Model of Scientific Explanation"	1239
Nickles, Thomas (1971) "Covering Law Explanation"	1297
Cupples, Brian (1977) "Three Types of Explanation"	1528
Railton, Peter (1978) "A Deductive-Nomological Model of Probabilistic Explanation"	1556
Hanna, Joseph F. (1978) "On Transmitted Information as a Measure of Explanatory Power"	1582
Forge, John (1980) "The Structure of Physical Explanation"	1645
Gardenfors, Peter (1980) "A Pragmatic Approach to Explanations"	1655
Woodward, James (1984) "Explanatory Asymmetries"	1822
Achinstein, Peter (1984) "The Pragmatic Character of Explanation"	4135
Jobe, Evan K. (1985) "Explanation, Causality, and Counterfactuals"	1858
Fetzer, James H. (1992) "What's Wrong with Salmon's History: The Third Decade"	2123
Glymour, Bruce (1998) "Contrastive, Non-probabilistic Statistical Explanations"	2444
Strevens, Michael (2000) "Do Large Probabilities Explain Better?"	2549

notably is involved in optimizing the algorithms' control parameters through several cycles of feedback loops between parameter setting, computer simulations, and careful inspection of results (Hu et al. 2014). Researcher intervention takes place at all four stages of the methodology (as described in the previous section). For step 1, data preprocessing, this included inspecting the lexicon that resulted from lemmatization processing and word filtering based on POS tagging (visual inspection aided by specific queries for stop-words, special characters, and word frequencies). For step 2, data modeling, alternative topic-modeling algorithms can be used to check for result robustness. Here, we chose to compare our results with an approach based on the k-means clustering algorithm (Aggarwal and Zhai 2012). The high similarity of the topic models of both approaches gave us confidence in our initial results (the LDA topics being, however, somehow sharper and easier to interpret). As mentioned earlier (see n. 6), several parameters need to be chosen for the LDA, notably the Dirichlet parameters for which we chose standard values, and the number of topics, for which we ran different tests at 50, 100, 200, 250, and 300 topics before settling at $K = 200$, a number that gave good results in terms of granularity and interpretation. We also chose to implement the LDA at the syntactic level of the sentence, which is a fairly standard way of proceeding. For step 3, diachronic topic analysis, we checked whether similar topics tended to occur at similar times; we also picked random topics, retrieved the articles in which they were likely to occur, and checked the content and dates of these articles. As for step 4, topic interpretation, we retrieved text excerpts to corroborate the authorial interpretation we could make on the basis of the most related words. Finally, the overall fit of the findings with known episodes in the philosophy of science also gives credence to the methodology and the quality of its implementation.

One of the limitations of topic modeling is that the methodology only results in identifying topics (which are no other than ordered lists of words chosen for the patterns of co-occurrence they display) and their relative probability of occurrence in documents, and nothing more. The methodology indeed is not meant to capture more sophisticated relationships between words or between topics such as entailment or causal relations (to this aim, other methodologies should be implemented, among which are conceptual analysis and argument mining methodologies [Peldszus and Stede 2013; Swanson et al. 2015]). The methodology therefore makes possible certain analyses in terms of occurrence and evolution of topics in the corpus, but it cannot reveal the deeper relationships between topics that the original authors intended in their papers in support of their own specific philosophical arguments. For instance, one can assess that a given paper is about explanation and causation, because it is associated with both topics, but not whether this is a paper in which the authors argue in favor of a causal model of explanation or against the relevance of causation in explanation. One should therefore be careful not to infer too much from the simple co-occurrence of topics in documents or periods.

That being said, the topic-modeling methods provide empirical grounding to specific theses about the content of *Philosophy of Science* and its evolution over the past 80 years, while offering the advantage of fallibility. The topics were found according to purely data-driven approaches. There was no guarantee at the start that the topics resulting from the application of topic-modeling methods would make certain topics salient at certain times according to patterns that indeed match what is known of the history of philosophy of science. An advantage of the methods is therefore that they are data driven and capable of efficiently inferring patterns in corpuses that are too large for manual scholarly analyses. Note that, in addition, the methods can easily be implemented on other corpuses, notably other philosophy of science journals, thereby broadening the inference base for historical analyses of the field.

3. What Is Philosophy of Science?

We focused our analyses on the 126 relevant topics. These topics are presented in table 3.[9] For ease of analysis, we also clustered the topics into 16 categories on the basis of our topic interpretation.[10] These 126 topics show a great diversity of interests exhibited by philosophers of science, ranging from questions about logic and philosophy of language to issues that are usually thought of as more metaphysical, for example, natural kinds or causation, including also a broad spectrum of epistemological questions about induction, confirmation, and the scientific method. More specifically, some 20 topics (16% of a total of 126 topics) relate to the philosophy of language, logic, and mathematics. Such topics include groups of terms that concern, for instance, questions of meaning, of linguistic expression, of synthetic or empirical truth (e.g., topics 86, 139, 185) but also various elements that appear formalized in logic and that relate, for instance, to predicate logic, to modal logic, or to questions about logical consistency, inferences, and axioms in logic (e.g., topics 20, 45, 140, 188). Other mathematics-related topics refer to mathematical variables, equations, and theorems, to geometry, or to notions of state phase and set theory (e.g., topics 26, 92, 123, 187, 196).

As could be expected, numerous topics also relate to the physical sciences (about 10% of the 126 topics), most notably quantum mechanics and relativity (topics 8, 19, 134, 138, 160) but also classical mechanics, thermodynamics, electromagnetism, and cosmology (topics 55, 67, 70, 73, 177, 190). We also identified a number of topics that more specifically concern models and simulations (topics 100, 116, 136) as well as others that can be related to complex systems, stability, and chaos (e.g., topics 80, 105, 117, 144). One topic was identified as being clearly related to chemistry (topic 183).

---

[9] Owing to a lack of space, only the topic labels are shown, not the words assigned to each topic or their conditional probability $\Pr(w|z)$.

[10] These categories are therefore not an outcome of the topic-modeling methods per se but result from our own interpretation of how topics best relate to one another. Our objective was to provide a convenient way of referring to groups of topics, given the high number of topics overall. Of course, in topic modeling, topics relate to topics in a multidimensional way (as can be noted in matrix Φ—see sec. 2). Hence, there are always multiple different ways of grouping them—so, for instance, we chose to assign the topic THEOREM-LOCALITY (184) to category C-Mathematics, although it is also directly relevant to category D-Physics and so forth. Other choices of category groupings than the ones we made are therefore possible. Yet, one should always remember that these category groupings are made for convenience: ultimately, it is only the topics that matter in topic-modeling analyses.

Interestingly, 11 topics were found to relate to the biological sciences (9% of the 126 topics). This is similar in size to the number of physics-related topics—and more than we would have anticipated.[11] Unsurprisingly, most of these topics concern evolutionary biology, in particular natural selection, adaptation, population genetics, and the famous species problem (topics 6, 12, 81, 93, 155, 173, 178). Other biological topics concern molecular biology and genetics as well as teleology and function (topics 31, 141, 194). Note also a fairly broad topic that concerns vital and mental phenomena (topic 28) and that somehow sits halfway between category H-Biology and category I-Mind. Topics in the philosophy of mind concern psychology and the neurosciences and cognitive sciences, with several topics about perception and learning as well as about intentionality (topics 9, 16, 56, 119, 195, 199). A few topics cover medicine, economics, and the social sciences (topics 133, 2, 22) as well as questions about values in science, the aims of science, science studies, intuition, or information that we grouped under category L-Varia (e.g., topics 35, 50, 118, 126, 170).

As could be expected, a large number of topics cover questions that are more directly epistemological (category M-Epistemology, with a total of 15 topics, hence about 12%) and that relate to evidence, beliefs, and justification, also including Bayesianism and the well-known problems of induction and confirmation (e.g., topics 10, 17, 39, 107, 127). A related category includes topics pertaining to scientific discovery and progress, with topics that specifically concern Thomas Kuhn's work but also research programs, heuristics, problem solving, and the scientific method (topics 33, 36, 62, 106, 193).

One of the largest categories is category O-Theory. It includes 17 topics (approximately 13% of the 126 topics) about the nature and grounding of scientific knowledge, from experiment and observation (topics 43, 57, 143) to accounts of scientific theories, theory discovery, and replacement (topics 47, 52, 54, 91, 113, 143), including as well scientific explanation, its Hempelian model, and the more recent causal, unificationist, and mechanistic accounts (topics 135, 158, 159, 163, 197).

Finally, we grouped under category P-Metaphysics a broader number of topics that concerned questions that could be framed as bearing upon "what there is" (16 topics, hence about 13%). These topics were expressed in the corpus with words that relate to space and time, to entities and kinds, and to properties, emergence, and supervenience (e.g., topics 3, 5, 40, 41, 82, 157). Other topics in that group related to realism and laws of nature and most notably to causation, including causal processes and causal relevance (topics 32, 38, 110, 182).

---

[11] That is to say, when comparing only with the topics that belong to category D-Physics (bearing in mind that topics about models, simulations, and complex systems do not just concern physics).

Table 3. The 126 Topics Identified in *Philosophy of Science* (1934–2015), Grouped into 16 Categories

Category	Topics (and Their ID Numbers)	Number of Topics	Average Probability (%)
A-Language	Definition (49); Term-meaning (86); Word-meaning (139); Concept (167); Truth-synthetic (185); Linguistic-expression (198)	6	5.0
B-Logic	Truth (11); Proposition-consistency (20); Modal-logic (45); Conclusion-argument (122); Sentence-predicate (140); Conditions (154); Logical-axiom (188)	7	5.7
C-Mathematics	Mathematical-equations (26); Mathematics-in-science (92); Variables (123); Geometry (152); Theorem-locality (184); State-space (187); Set-theory (196)	7	5.6
D-Physics	Quantum-mechanics (8); Relativity (19); Thermodynamics-Entropy (55); Electromagnetism (67); Celestial-mechanics (70); Thermodynamics-Perfect-gases (73); Physical-principles (115); Particle-physics (134); Quantum-measurement (138); Quantum-entanglement (160); Classical-mechanics-gravitation (177); Cosmology (190)	12	9.6
E-Models, simulation	Modeling (100); Models-and-representation (116); Computer-simulation (136)	3	2.3
F-Systems, complexity	Stability (80); Boundary-conditions (105); Complexity (117); Level-hierarchy (121); Chaos (144); Mereology (145)	6	4.7
G-Chemistry	Chemistry (183)	1	.8
H-Biology	Evolutionary-games (6); Adaptation (12); Life-consciousness-emotions (28); Teleology (31); Natural-selection (81); Species (93); Function (141); Evolution (155); Population-genetics (173); Group-selection (178); Genetics (194)	11	8.7
I-Mind	Neuroscience (9); Intentionality (16); Neurosciences (56); Therapy (101); Image-perception (119); Behavior (164); Learning (174); Psychology-cognitive (195); Perception (199)	9	7.1

J-Medicine	Disease-health (133)	1	.8
K-Social sciences	Economy (2); Social-science (22)	2	1.7
L-Varia	Philosophical-schools (21); Science-studies (35); Science-and-values (50); Patterns-structures (51); Intuition (118); Aims-of-science (126); Information (170); Perspective-pluralism (192)	8	6.3
M-Epistemology	Measure-probability-confirmation (0); Induction (10); Probabilities (17); Justification (39); Bets-odds (48); Error-risk (59); Test-hypothesis (60); Probability-chance (65); Evidence (83); Belief-degree (95); Justification-reasons (97); Color-Goodman (99); Confirmation (107); Knowledge (127); Rational-choice (168)	15	11.8
N-Scientific inquiry	Kuhn (33); Discovery-heuristics (36); Research-work-program (62); Scientific-method (106); Problem-solving (193)	5	4.0
O-Theory	Observation (43); Interpretation (47); Discovery-replacement (52); Assumption (54); Experiment (57); Prediction (72); Structure-model (91); Reductionism (112); Semantic-syntactic-accounts (113); DN-explanation (135); Data-techniques (143); Evaluation (153); Explanation-accounts (158); Explanation-description-prediction (159); Explanatory-power (163); Adequacy-success (186); Mechanism-explanation (197)	17	13.2
P-Metaphysics	Space-time (3); Realism (4); Natural-kinds (5); Laws-of-nature (13); Causal-process (32); Causal-relation (38); Type-token-kinds (40); Space-geometry (41); Time-event (82); Entity (89); Cause-effect (110); Physical-reality (129); Fact (131); Property-emergence-supervenience (157); Object-property-kind (169); Relevant-factor-causal (182)	16	12.7

Note.—The probability of each category is calculated by summing up the probabilities of occurrence of its topics in the entire corpus. Total number of topics = 126.

4. How Did *Philosophy of Science* Change between 1934 and Today?

Topic significance in *Philosophy of Science* varies over time according to specific diachronic trends that our method can reveal. Some topics were very significant in the early periods of the journal and almost nonexistent some years later. Other topics emerged only recently. Still others maintained a relatively stable significance in the corpus through time. By analyzing the matrix **Ω**, it is possible to identify these trends and retrieve the most probable topics for every period from 1934 until 2015. One can also aggregate these probabilities per category so as to obtain a coarser-grained view at the category level.[12] These topics correspond to the words used by philosophers during these periods. They therefore reflect the type of questioning and research problems that received attention at different moments in the history of philosophy of science. Note, however, that, because they are embedded in one particular journal, they may also reflect the editorial policies of this journal and changes to these policies over time. It is known, for instance, that the founder and first editor of *Philosophy of Science*, from 1934 until 1947, William M. Malisoff, had a very encompassing and engaged view of philosophy of science (Howard 2003; Reisch 2005, chap. 5). The journal also struggled with quality issues, a problem that C. West Churchman found pressing when in charge from 1949 (after some interim arrangements following the sudden death of Malisoff) until 1959 (Malisoff 1944; Churchman 1984). Changes in editorial policy were then stirring. Implementing these changes would fall to Richard Rudner, who took the role of editor from 1959 till 1975, at a time characterized by the professionalization of the discipline and a reorganization of the governance of the Philosophy of Science Association; Rudner, together with a strengthened editorial board, turned the journal into the publication we know today (Howard 2003; Douglas 2010, 2019). As we will see, some changes in topic importance in the journal were contemporary to these editorial shifts. The evolution of the topics of philosophy of science is not only the result of the larger-scale intellectual and historical context but also of the smaller-scale editorial decisions (Giere 1996; Hardcastle and Richardson 2003; Reisch 2005).

A broad overview of the evolution of topics over time is depicted in figure 2, which represents the evolution of topic categories from 1934 until 2015. In order to get a better understanding of these trends, one can look at the level of

---

[12] To calculate the category probability per period, we added and renormalized the probabilities (given in Θ) of all sentences of all documents per topic. For every document, we then only kept the most probable topics by filtering out the topics whose probability was lower than a threshold of 0.012 (this value was found appropriate to cut the L-curve probability distributions slightly above their elbows). These probability values were then added and renormalized for all documents of the same period and then aggregated per topic category.

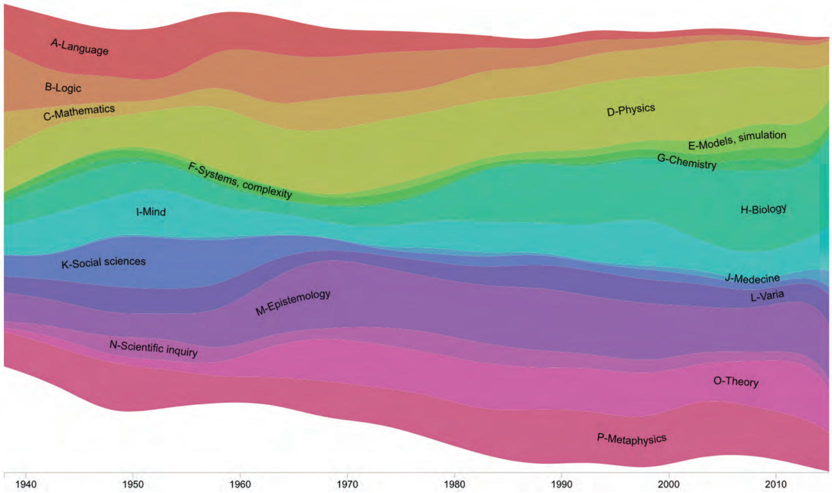


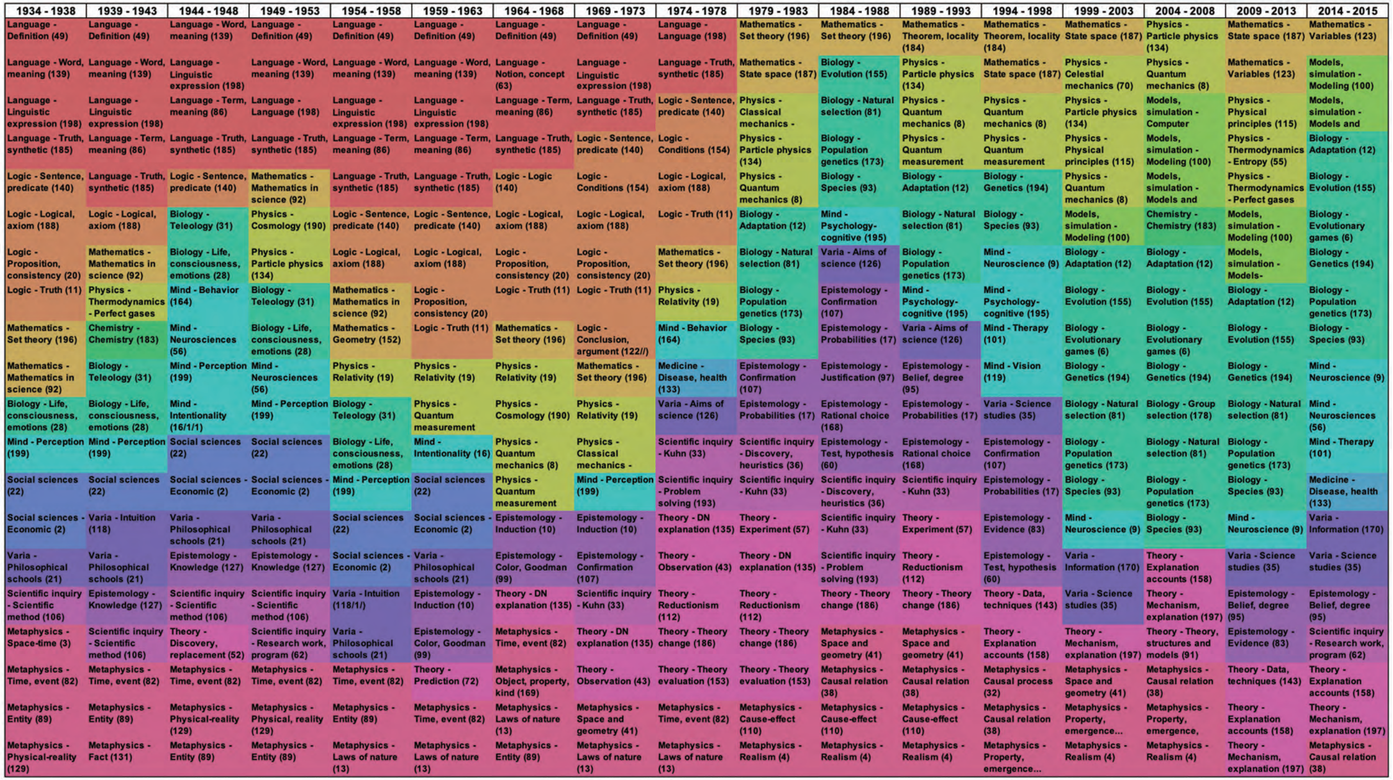
Figure 2. Diachronic evolution of topic categories in *Philosophy of Science*, 1934–2015. The width of each 'stream' is proportional to the probability of each category in the corpus. Image made with RAWGraphs (Mauri et al. 2017).

individual topics. This is what we did by focusing on the 20 most-probable topics for every period.[13] Of course, the absence of a topic from the top 20 does not mean this topic was not expressed in the corpus during that time period but only that it was not where the action was, so to speak. Out of the 126 topics of interest, we found that 88 made it at least once to the top list (approximately 70%). For every period, we grouped them by categories and color coded them accordingly (fig. 3).

The results in figure 2 show interesting diachronic patterns, some of which corroborate known episodes in the philosophy of science but also less obvious ones. Maybe one of the most striking patterns is the dominance of language- and logic-related topics from the start of *Philosophy of Science* in 1934 up until the 1970s, with language consistently accounting for about 20% of the top topics and logic coming in two bursts, one before the 1950s and a stronger one from the 1950s to the 1970s. These episodes were then followed by a sharp decline and a total disappearance of language and logic within top-20 topics from the 1980s onward. Topics related to mathematics, in contrast, were present more or less continuously across the whole existence of the journal, with an

---

[13] When looking at the probability distribution of the 126 significant topics for any given period, we found that this distribution usually has an S-shape, with a head comprising some 15 to 30 highly probable topics, followed by a flattening of the curve for 60 to 80 topics and a tail of some 15 to 30 topics with low probability. The number of 20 topics therefore appeared as a reasonable cut-off point for singling out the most probable topics.



1934 - 1938	1939 - 1943	1944 - 1948	1949 - 1953	1954 - 1958	1959 - 1963	1964 - 1968	1969 - 1973	1974 - 1978	1979 - 1983	1984 - 1988	1989 - 1993	1994 - 1998	1999 - 2003	2004 - 2008	2009 - 2013	2014 - 2015
Language - Definition (49)	Language - Definition (49)	Language - Word, meaning (139)	Language - Definition (49)	Language - Definition (49)	Language - Definition (49)	Language - Definition (49)	Language - Definition (49)	Language - Language (198)	Mathematics - Set theory (196)	Mathematics - Set theory (196)	Mathematics - Theorem, locality (184)	Mathematics - Theorem, locality (184)	Mathematics - State space (187)	Physics - Particle physics (134)	Mathematics - State space (187)	Mathematics - Variables (123)
Language - Word, meaning (139)	Language - Word, meaning (139)	Language - Linguistic expression (198)	Language - Word, meaning (139)	Language - Word, meaning (139)	Language - Word, meaning (139)	Language - Notion, concept (63)	Language - Linguistic expression (198)	Language - Truth, synthetic (185)	Mathematics - State space (187)	Biology - Evolution (155)	Physics - Particle physics (134)	Mathematics - State space (187)	Physics - Celestial mechanics (70)	Physics - Quantum mechanics (8)	Mathematics - Variables (123)	Models, simulation - Modeling (100)
Language - Linguistic expression (198)	Language - Linguistic expression (198)	Language - Term, meaning (86)	Language - Language (198)	Language - Linguistic expression (198)	Language - Linguistic expression (198)	Language - Term, meaning (86)	Language - Truth, synthetic (185)	Logic - Sentence, predicate (140)	Physics - Classical mechanics -	Biology - Natural selection (81)	Physics - Quantum mechanics (8)	Physics - Quantum mechanics (8)	Physics - Particle physics (134)	Models, simulation - Computer	Physics - Physical principles (115)	Models, simulation - Models and
Language - Truth, synthetic (185)	Language - Term, meaning (86)	Language - Truth, synthetic (185)	Language - Truth, synthetic (185)	Language - Term, meaning (86)	Language - Term, meaning (86)	Language - Truth, synthetic (185)	Logic - Sentence, predicate (140)	Logic - Conditions (154)	Physics - Particle physics (134)	Biology - Population genetics (173)	Physics - Quantum measurement	Physics - Quantum measurement	Physics - Physical principles (115)	Models, simulation - Modeling (100)	Physics - Thermodynamics - Entropy (55)	Biology - Adaptation (12)
Logic - Sentence, predicate (140)	Language - Truth, synthetic (185)	Logic - Sentence, predicate (140)	Mathematics - Mathematics in science (92)	Language - Truth, synthetic (185)	Language - Truth, synthetic (185)	Logic - Logic (140)	Logic - Conditions (154)	Logic - Logical, axiom (188)	Physics - Quantum mechanics (8)	Biology - Species (93)	Biology - Adaptation (12)	Biology - Genetics (194)	Physics - Quantum mechanics (8)	Models, simulation - Models and	Physics - Thermodynamics - Perfect gases	Biology - Evolution (155)
Logic - Logical, axiom (188)	Logic - Logical, axiom (188)	Biology - Teleology (31)	Physics - Cosmology (190)	Logic - Sentence, predicate (140)	Logic - Sentence, predicate (140)	Logic - Logical, axiom (188)	Logic - Logical, axiom (188)	Logic - Truth (11)	Biology - Adaptation (12)	Mind - Psychology-cognitive (195)	Biology - Natural selection (81)	Biology - Species (93)	Models, simulation - Modeling (100)	Chemistry - Chemistry (183)	Models, simulation - Modeling (100)	Biology - Evolutionary games (6)
Logic - Proposition, consistency (20)	Mathematics - Mathematics in science (92)	Biology - Life, consciousness, emotions (28)	Physics - Particle physics (134)	Logic - Logical, axiom (188)	Logic - Logical, axiom (188)	Logic - Proposition, consistency (20)	Logic - Proposition, consistency (20)	Mathematics - Set theory (196)	Biology - Natural selection (81)	Varia - Aims of science (126)	Biology - Population genetics (173)	Mind - Neuroscience (9)	Biology - Adaptation (12)	Biology - Adaptation (12)	Models, simulation - Models-	Biology - Genetics (194)
Logic - Truth (11)	Physics - Thermodynamics - Perfect gases	Mind - Behavior (164)	Biology - Teleology (31)	Mathematics - Mathematics in science (92)	Logic - Proposition, consistency (20)	Logic - Truth (11)	Logic - Truth (11)	Physics - Relativity (19)	Biology - Population genetics (173)	Epistemology - Confirmation (107)	Mind - Psychology-cognitive (195)	Mind - Psychology-cognitive (195)	Biology - Evolution (155)	Biology - Evolution (155)	Biology - Adaptation (12)	Biology - Population genetics (173)
Mathematics - Set theory (196)	Chemistry - Chemistry (183)	Mind - Neurosciences (56)	Biology - Life, consciousness, emotions (28)	Mathematics - Geometry (152)	Logic - Truth (11)	Mathematics - Set theory (196)	Logic - Conclusion, argument (122//)	Mind - Behavior (164)	Biology - Species (93)	Epistemology - Probabilities (17)	Varia - Aims of science (126)	Mind - Therapy (101)	Biology - Evolutionary games (6)	Biology - Evolutionary games (6)	Biology - Evolution (155)	Biology - Species (93)
Mathematics - Mathematics in science (92)	Biology - Teleology (31)	Mind - Perception (199)	Mind - Neurosciences (56)	Physics - Relativity (19)	Physics - Relativity (19)	Physics - Relativity (19)	Mathematics - Set theory (196)	Medicine - Disease, health (133)	Epistemology - Confirmation (107)	Epistemology - Justification (97)	Epistemology - Belief, degree (95)	Mind - Vision (119)	Biology - Genetics (194)	Biology - Genetics (194)	Biology - Genetics (194)	Mind - Neuroscience (9)
Biology - Life, consciousness, emotions (28)	Biology - Life, consciousness, emotions (28)	Mind - Intentionality (16/1/1)	Mind - Perception (199)	Biology - Teleology (31)	Physics - Quantum measurement	Physics - Cosmology (190)	Physics - Relativity (19)	Varia - Aims of science (126)	Epistemology - Probabilities (17)	Epistemology - Rational choice (168)	Epistemology - Probabilities (17)	Varia - Science studies (35)	Biology - Natural selection (81)	Biology - Group selection (178)	Biology - Natural selection (81)	Mind - Neurosciences (56)
Mind - Perception (199)	Mind - Perception (199)	Social sciences (22)	Social sciences (22)	Biology - Life, consciousness, emotions (28)	Mind - Intentionality (16)	Physics - Quantum mechanics (8)	Physics - Classical mechanics -	Scientific inquiry - Kuhn (33)	Scientific inquiry - Discovery, heuristics (36)	Epistemology - Test, hypothesis (60)	Epistemology - Rational choice (168)	Epistemology - Confirmation (107)	Biology - Population genetics (173)	Biology - Natural selection (81)	Biology - Population genetics (173)	Mind - Therapy (101)
Social sciences (22)	Social sciences (22)	Social sciences - Economic (2)	Social sciences - Economic (2)	Mind - Perception (199)	Social sciences (22)	Physics - Quantum measurement	Mind - Perception (199)	Scientific inquiry - Problem solving (193)	Scientific inquiry - Kuhn (33)	Scientific inquiry - Discovery, heuristics (36)	Scientific inquiry - Kuhn (33)	Epistemology - Probabilities (17)	Biology - Species (93)	Biology - Population genetics (173)	Biology - Species (93)	Medicine - Disease, health (133)
Social sciences - Economic (2)	Varia - Intuition (118)	Varia - Philosophical schools (21)	Varia - Philosophical schools (21)	Social sciences (22)	Social sciences - Economic (2)	Epistemology - Induction (10)	Epistemology - Induction (10)	Theory - DN explanation (135)	Theory - Experiment (57)	Scientific inquiry - Kuhn (33)	Theory - Experiment (57)	Epistemology - Evidence (83)	Mind - Neuroscience (9)	Biology - Species (93)	Mind - Neuroscience (9)	Varia - Information (170)
Varia - Philosophical schools (21)	Varia - Philosophical schools (21)	Epistemology - Knowledge (127)	Epistemology - Knowledge (127)	Social sciences - Economic (2)	Varia - Philosophical schools (21)	Epistemology - Color, Goodman (99)	Epistemology - Confirmation (107)	Theory - Observation (43)	Theory - DN explanation (135)	Scientific inquiry - Problem solving (193)	Theory - Reductionism (112)	Epistemology - Test, hypothesis (60)	Varia - Information (170)	Theory - Explanation accounts (158)	Varia - Science studies (35)	Varia - Science studies (35)
Scientific inquiry - Scientific method (106)	Epistemology - Knowledge (127)	Scientific inquiry - Scientific method (106)	Scientific inquiry - Scientific method (106)	Varia - Intuition (118/1/)	Epistemology - Induction (10)	Theory - DN explanation (135)	Scientific inquiry - Kuhn (33)	Theory - Reductionism (112)	Theory - Reductionism (112)	Theory - Theory change (186)	Theory - Theory change (186)	Theory - Data, techniques (143)	Varia - Science studies (35)	Theory - Mechanism, explanation (197)	Epistemology - Belief, degree (95)	Epistemology - Belief, degree (95)
Metaphysics - Space-time (3)	Scientific inquiry - Scientific method (106)	Theory - Discovery, replacement (52)	Scientific inquiry - Research work, program (62)	Varia - Philosophical schools (21)	Epistemology - Color, Goodman (99)	Metaphysics - Time, event (82)	Theory - DN explanation (135)	Theory - Theory change (186)	Theory - Theory change (186)	Metaphysics - Space and geometry (41)	Metaphysics - Space and geometry (41)	Theory - Explanation accounts (158)	Theory - Mechanism, explanation (197)	Theory - Theory, structures and models (91)	Epistemology - Evidence (83)	Scientific inquiry - Research work, program (62)
Metaphysics - Time, event (82)	Metaphysics - Time, event (82)	Metaphysics - Time, event (82)	Metaphysics - Time, event (82)	Metaphysics - Time, event (82)	Theory - Prediction (72)	Metaphysics - Object, property, kind (169)	Theory - Observation (43)	Theory - Theory evaluation (153)	Theory - Theory evaluation (153)	Metaphysics - Causal relation (38)	Metaphysics - Causal relation (38)	Metaphysics - Causal process (32)	Metaphysics - Space and geometry (41)	Metaphysics - Causal relation (38)	Theory - Data, techniques (143)	Theory - Explanation accounts (158)
Metaphysics - Entity (89)	Metaphysics - Entity (89)	Metaphysics - Physical-reality (129)	Metaphysics - Physical, reality (129)	Metaphysics - Entity (89)	Metaphysics - Time, event (82)	Metaphysics - Laws of nature (13)	Metaphysics - Space and geometry (41)	Metaphysics - Time, event (82)	Metaphysics - Cause-effect (110)	Metaphysics - Cause-effect (110)	Metaphysics - Cause-effect (110)	Metaphysics - Causal relation (38)	Metaphysics - Property, emergence...	Metaphysics - Property, emergence,	Theory - Explanation accounts (158)	Theory - Mechanism, explanation (197)
Metaphysics - Physical-reality (129)	Metaphysics - Fact (131)	Metaphysics - Entity (89)	Metaphysics - Entity (89)	Metaphysics - Laws of nature (13)	Metaphysics - Laws of nature (13)	Metaphysics - Entity (89)	Metaphysics - Laws of nature (13)	Metaphysics - Laws of nature (13)	Metaphysics - Realism (4)	Metaphysics - Realism (4)	Metaphysics - Realism (4)	Metaphysics - Property, emergence...	Metaphysics - Realism (4)	Metaphysics - Realism (4)	Theory - Mechanism, explanation (197)	Metaphysics - Causal relation (38)

Figure 3. Diachronic evolution of the 20 most-probable topics in *Philosophy of Science*, 1934–2015, grouped and color coded by categories

average of about one topic per period. Because several of these mathematics-centered topics include terms that will be mobilized by any sufficiently formalized scientific theory (e.g., *variables*, *state-space*, or *theorem*), one possibility could be that they somehow piggyback on a number of physics-related topics.

Our results show physics topics to be reasonably well represented in *Philosophy of Science*, although not in any dominant fashion (about 10% of top topics on average). Note how they appear within the top-20 topics in mainly two periods: a first period from the 1950s to the 1970s with a simultaneous interest in relativity (topic 19) and quantum mechanics (topics 8, 138) and a second period from the 1990s up till the 2010s, much more focused on quantum mechanics and particle physics (topics 8, 138, 115, 134). Other physics-related topics include classical mechanics, thermodynamics, and cosmology. Interestingly, one notices the emergence in the first decade of the 2000s of several topics that are somehow connected to the physical sciences (although not exclusively) and that concern models and simulations (topics 100, 116, 136). These models-related topics could correspond to a relative shift of interest away from the traditional topics of the physical sciences to modeling and simulation, which are topics that are not restricted to physics but also concern other scientific disciplines.

Another striking pattern is that of the biology-related topics, for which there clearly is a distinction between before and after the 1960s and 1970s. From the journal's start in 1934 until the 1960s, one could find only two biology-related topics: one that concerned the question of teleology, with related problems of organismal purposiveness and goal orientation (topic 31), and another that was quite broad in scope and concerned all vital and mental processes (topic 28—hence also relevant to mind-related topics). These topics, however, testify to the presence of philosophy of biology during the formative years of philosophy of science (Byron 2007; Nicholson and Gawne 2015). The 20 years that followed were characterized by a relative eclipse of philosophy of biology topics, which is concomitant with the change in editorial policy of the journal, following Rudner's nomination as editor in 1959 (Howard 2003; Reisch 2005). Toward the very end of the 1970s to the early 1980s, one witnessed a very strong development of biology-related topics, especially topics on natural selection, population genetics, and the concept of adaptation, with an even stronger presence from the first decade of the 2000s onward (topics 12, 81, 155, 173). Note how the topic on the concept of species (topic 93) was also very much present in all periods from the 1980s up until now, alongside topics that concerned evolutionary games and group selection (topics 6 and 178). Note also how the topic about genetics and molecular biology only made it among the top-20 topics in the 1990s (topic 194) yet still remained marginal compared to evolutionary biology topics. In total, with an average of five topics per period, biology-related topics accounted for more topics from the 1980s until the 2010s (25% of top topics) than philosophy of language from the 1930s until the 1970s.

Topics related to philosophy of mind broadly construed—including perception, vision, behavior, intentionality, and the neurosciences as well as psychology and therapy (topics 9, 16, 56, 101, 119, 195, 199)—were found to be irregularly present throughout the whole period of existence of the journal. Three bubbles could be identified, however, each about a decade long: the first during the 1950s, the second during the 1990s, and the third during the 2010s.

Besides these marked trends, interesting insights can be gained by looking at some more marginal topics. In particular, it is striking to see how topics related to the social sciences and economics (topics 2, 22) were consistently present from the 1930s up until the 1960s and then totally vanished from the top-20 topics. A topic about philosophical schools (topic 21, with such terms as *empiricism, doctrine, philosophy, kant, metaphysical, positivism, tradition, materialism*) was also present during the same periods, its subsequent disappearance possibly denoting a change in the writing style of *Philosophy of Science* articles. These findings are consistent with the editorial life of the journal, notably the change in editorial policies that took place at the end of the 1950s, in particular with Rudner's role as editor (Howard 2003). After the 1960s, some other marginal trends included topics about medicine, health, and disease (topic 133), information (topic 170), and maybe more significantly, topics about the aims of science and science studies (topics 35, 126).

Analysis of figure 2 also reveals another significant diachronic pattern that concerns the rise of epistemological topics, understood in a broad sense, from the 1960s until 2015, with a prominent surge in the 1980s. Some of these topics focused very much on formal epistemology, with questions related to induction, confirmation, evidence, or justification, especially up through the 1970s (topics 10, 99, 107, 127), and turning more specifically into questions about probabilities and Bayesianism from the 1990s onward (topics 17, 95). Other topics that were grouped under the category O-Theory concerned questions about the nature of scientific theories, experiment, and observation, as well as scientific explanation (topics 43, 57, 91, 135, 158, 197). These topics were quite strongly represented in the 1970s and 1980s and again since the first decade of the 2000s. Finally, a third group of epistemological topics included topics related to scientific inquiry, in particular topics about scientific methods (topic 106) from the 1930s to the 1950s, then shifting to topics about discovery, heuristics, Kuhn, and problem solving in the 1970s and 1980s (topics 33, 36, 193).

Finally, figure 2 reveals a last group of topics that made it to the top 20 and that included topics of a more metaphysical nature, ranging over classical philosophical questions about time and space (topics 3, 41, 82—fairly consistently from the 1930s until the 1990s), entities, properties, and kinds (topics 89, 157, 169—at different time periods), laws of nature (topic 13—from the 1950s until the 1970s), causation (topics 32, 38, 110—esp. throughout the 1980s and 1990s), and physical description and realism (topics 4, 129).

5. Zooming In on Specific Topic Trends

In addition to identifying large trends in topic evolution, the methodology makes it possible to zoom in on specific topics and to follow their diachronic patterns. In what follows, we have chosen to focus on three sets of topics: (1) topics that concern the notion of scientific explanation, which are easy to connect to well-known episodes in the development of the field; (2) topics that concern biology and its significant rise from the 1980s onward; and (3) topics in modeling that illustrate a more recent trend in philosophy of science.

The notion of scientific explanation is mostly captured by three topics: DN-EXPLANATION (135), EXPLANATION-ACCOUNTS (158), and MECHANISM-EXPLANATION (197). Their time evolution can be seen in figure 4. For each period, the most strongly associated articles with each topic have also been retrieved (see table A1). DN-EXPLANATION is strongly connected to the DN-model of scientific explanation, to questions about the extension of this model to statistical explanation, and to some of the criticisms it faced. Following Hempel and Oppenheim's paper in 1948, the topic peaked significantly in the 1970s and then slowly decreased up until now. In contrast, EXPLANATION-ACCOUNTS remained marginal in the 1970s, although slowly increasing, with two bumps, one in the 1990s and another in the 2010s; the topic can be linked to other philosophical accounts of scientific explanation, in particular unification and causal accounts, and more recently to questions about understanding. Finally, MECHANISM-EXPLANATION, which is clearly related to mechanistic explanation, showed a strong increase in

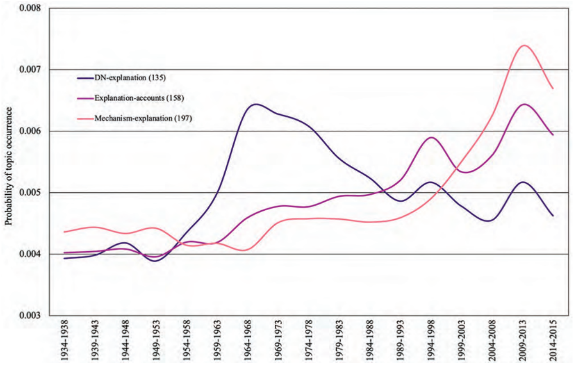


Figure 4. Evolution of the probability of occurrence of topics related to "scientific explanation"

the first decade of the 2000s up until a recent slight decrease. The diachronic evolution of these three topics mirrors known episodes in the development of philosophical accounts of scientific explanation and—we argue—corroborate the algorithmic methodology we used.

As noted above, topics related to biology surged in *Philosophy of Science* starting in the 1980s. Out of the 11 biology-related topics that emerged through the topic modeling of the whole corpus, we chose to focus below on topics 81 (Natural-selection), 93 (Species), and 194 (Genetics). As can be seen in figure 5, all three topics showed a very steep increase in the late 1970s and early 1980s, even more markedly for Species and Natural-selection. After a slightly slower increase in the 1980s, Genetics overtook Species and Natural-selection in the 1990s. Despite fluctuations in the last decade, all three topics remained quite strong in their probability of occurrence, with, however, a recent decreasing trend for Species and Natural-selection and an increasing one for Genetics, possibly because of a shift of interest toward genetics and molecular biology (as can be seen in the list of most strongly associated articles—see table A2).

The topics we chose to represent in figure 6 are all topics that concern models, simulation, and representation. All three topics can be seen to have increased in the first decade of the 2000s, although topics 100 (Modeling) and 116 (Models-and-representation)—which respectively relate to issues of models and modeling across the sciences and to issues of models as representations—did so much more significantly than topic 136, Computer-simulation, which more specifically related to simulation and computation (see also the list of most strongly associated articles in table A3). In any case, the recent trend is one of still sharp increase. This class of topics would therefore be expected to be of continued interest in the 2010s onward.

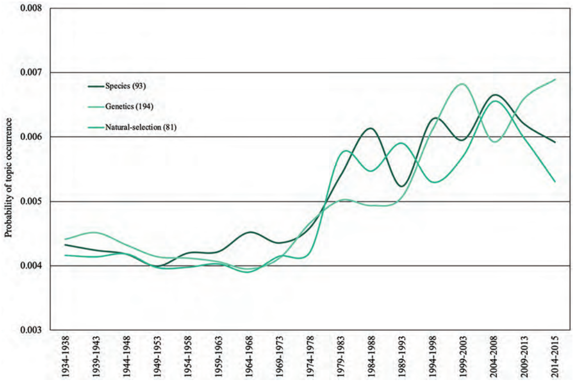


Figure 5. Evolution of the probability of occurrence of topics related to “biology”

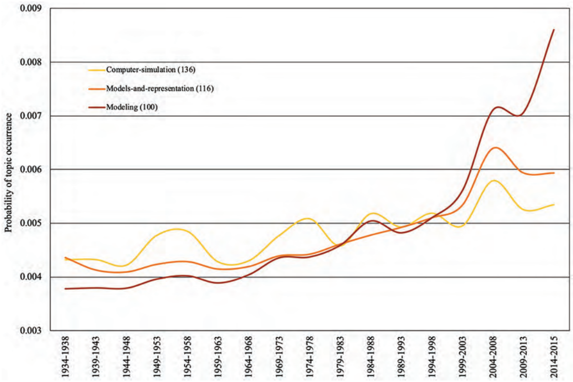


Figure 6. Evolution of the probability of occurrence of topics related to "models and simulations"

Of course, much more could be said about the detailed diachronic trends followed by many different topics. With these three sets of topics, we at least hope to have shown the types of analyses that can be conducted.

6. Conclusion

Looking back at 80 years of *Philosophy of Science* through the lenses of unsupervised topic-modeling algorithms provides a quantitative, data-driven perspective on known and lesser-known episodes in the development of the discipline, from the rise and fall of logical empiricism from the 1930s to the 1970s to the emergence of research in philosophy of biology in the 1980s, including numerous other topics about scientific theories and scientific explanation, models and simulations, causation, and realism, to mention just a few. Text-mining methods provide a wealth of novel avenues for analyzing the full-text content of large corpuses. We have shown here the first results of the type of synchronic and diachronic analyses that are now feasible. But much more is possible, from more detailed diachronic studies of specific topics up to the identification of clusters of authors and genealogies of topics, including the classification of topic temporal patterns and even the implementation of predictive modeling. Much has happened in 80 years of *Philosophy of Science*, and more is to come.

Appendix

Table A1. Top-20 Articles Most Strongly Associated with Selected Explanation-Related Topics

Period	DN-EXPLANATION (135)	EXPLANATION-ACCOUNTS (158)	MECHANISM-EXPLANATION (197)
1934–38			
1939–43			
1944–48	Hempel, Carl G., Oppenheim, Paul (1948) “Studies in the Logic of Explanation”		
1949–53			
1954–58			
1959–63	Grünbaum, Adolf (1962) “Temporally-Asymmetric Principles, Parity between Explanation and Prediction, and Mechanism versus Teleology”		
1964–68	Hempel, Carl G. (1968) “Maximal Specificity and Lawlikeness in Probabilistic Explanation”		
1969–73	Nickles, Thomas (1971) “Covering Law Explanation” Omer, I. A. (1970) “On the D-N Model of Scientific Explanation” Fetzer, James H. (1972) “Statistical Explanations”		
1974–78	Cupples, Brian (1977) “Three Types of Explanation” Railton, Peter (1978) “A Deductive-Nomological Model of Probabilistic Explanation” Hanna, Joseph F. (1978) “On Transmitted Information as a Measure of Explanatory Power” Thorpe, D. A. (1974) “The Quartercentenary Model of D-N Explanation”		

	Barr, William F. (1974) “A Pragmatic Analysis of Idealizations in Physics”		
1979–83	Forge, John (1980) “The Structure of Physical Explanation” Gardenfors, Peter (1980) “A Pragmatic Approach to Explanations”	Kitcher, Philip (1981) “Explanatory Unification”	
1984–88	Woodward, James (1984) “Explanatory Asymmetries” Jobe, Evan K. (1985) “Explanation, Causality, and Counterfactuals” Achinstein, Peter (1984) “The Pragmatic Character of Explanation”		
1989–93	Fetzer, James H. (1992) “What’s Wrong with Salmon’s History: The Third Decade”	Schwartz, Justin (1993) “Functional Explanation and Metaphysical Individualism” Barnes, Eric (1992) “Explanatory Unification and Scientific Understanding”	
1994–98	Glymour, Bruce (1998) “Contrastive, Non-probabilistic Statistical Explanations” Rescher, Nicholas (1997) “H2O: Hempel-Helmer-Oppenheim, an Episode in the History of Scientific Philosophy in the 20th Century”	Steel, Daniel (1998) “Warfare and Western Manufactures: A Case Study of Explanation in Anthropology” Berger, Ruth (1998) “Understanding Science: Why Causes Are Not Enough” Jones, Todd (1998) “Unification, Deduction, and History: A Reply to Steel” Wilson, Robert A. (1994) “Causal Depth, Theoretical Appropriateness, and Individualism in Psychology” Rappaport, Steven (1996) “Inference to the Best Explanation: Is It Really Different from Mill’s Methods?”	Glennan, Stuart S. (1997) “Capacities, Universality, and Singularity”
1999–2003	Strevens, Michael (2000) “Do Large Probabilities Explain Better?”	Rosenberg, Alex (2001) “Reductionism in a Historical Science”	Craver, Carl F. (2001) “Role Functions, Mechanisms, and Hierarchy”

Table A1 (*Continued*)

Period	DN-Explanation (135)	Explanation-Accounts (158)	Mechanism-Explanation (197)
		Trout, J. D. (2002) “Scientific Explanation and the Sense of Understanding”	Machamer, Peter, Darden, Lindley, Craver, Carl F. (2000) “Thinking about Mechanisms”
			Darden, Lindley (2002) “Strategies for Discovering Mechanisms: Schema Instantiation, Modular Subassembly, Forward/Backward Chaining”
2004–8		Gijsbers, Victor (2007) “Why Unification Is Neither Necessary Nor Sufficient for Explanation”	Tabery, James G. (2004) “Synthesizing Activities and Interactions in the Concept of a Mechanism”
		de Regt, Henk W. (2004) “Discussion Note: Making Sense of Understanding”	Darden, Lindley (2008) “Thinking Again about Biological Mechanisms”
		Trout, J. D. (2005) “Paying the Price for a Theory of Explanation: De Regt’s Discussion of Trout”	Datteri, Edoardo, Tamburrini, Guglielmo (2007) “Biorobotic Experiments for the Discovery of Biological Mechanisms”
			Piccinini, Gualtiero (2007) “Computing Mechanisms”
			Barros, D. Benjamin (2008) “Natural Selection as a Mechanism”

2009–13	Douglas, Heather E. (2009) “Reintroducing Prediction to Explanation”	Khalifa, Kareem (2012) “Inaugurating Understanding or Repackaging Explanation?”	Bonnie Fagan, Melinda (2012) “The Joint Account of Mechanistic Explanation”
		Weslake, Brad (2010) “Explanatory Depth”	DesAutels, Lane (2011) “Against Regular and Irregular Characterizations of Mechanisms”
		Douglas, Heather E. (2009) “Reintroducing Prediction to Explanation”	Garson, Justin (2013) “The Functional Sense of Mechanism”
		Bonnie Fagan, Melinda (2012) “The Joint Account of Mechanistic Explanation”	Bechtel, William (2011) “Mechanism and Biological Explanation”
		Potochnik, Angela (2010) “Levels of Explanation Reconceived”	Overton, James A. (2011) “Mechanisms, Types, and Abstractions”
			Leuridan, Bert (2010) “Can Mechanisms Really Replace Laws of Nature?”
			Zednik, Carlos (2011) “The Nature of Dynamical Explanation”
2014–15		Reutlinger, Alexander (2014) “Why Is There Universal Macrobehavior? Renormalization Group Explanation as Noncausal Explanation”	Skillings, Derek John (2015) “Mechanistic Explanation of Biological Processes”
		Jansson, Lina (2014) “Causal Theories of Explanation and the Challenge of Explanatory Disagreement”	Baetu, Tudor M. (2015) “The Completeness of Mechanistic Explanations”

Table A2. Top-20 Articles Most Strongly Associated with Selected Biology Topics

Period	NATURAL SELECTION (81)	SPECIES (93)	GENETICS (194)
1934–38			
1939–43			
1944–48			
1949–53			
1954–58			
1959–63			
1964–68		Lehman, Hugh (1967) “Are Biological Species Real?”	
1969–73			
1974–78		Hull, David L. (1978) “A Matter of Individuality”	Darden, Lindley, Maull, Nancy (1977) “Interfield Theories”
1979–83	Wimsatt, William C. (1980) “The Units of Selection and the Structure of the Multi-Level Genome”		Kimbrough, Steven Orla (1979) “On the Reduction of Genetics to Molecular Biology”
	Sober, Elliott (1980) “Holism, Individualism, and the Units of Selection”		
	Sober, Elliott, Lewontin, Richard C. (1982) “Artifact, Cause, and Genic Selection”		
1984–1988	Mitchell, Sandra D. (1987) “Competing Units of Selection? A Case of Symbiosis”	Kitcher, Philip (1984) “Species”	Bechtel, William (1984) “Reconceptualizations and Interfield Connections: The Discovery of the Link between Vitamins and Coenzymes”
	Beatty, John (1984) “Chance and Natural Selection”	de Queiroz, Kevin (1988) “Systematics and the Darwinian Revolution”	Schank, Jeffrey C., Wimsatt, William C. (1986) “Generative Entrenchment and Evolution”
1989–93	Waters, C. Kenneth (1991) “Tempered Realism about the Force of Selection”	Ereshefsky, Marc (1992) “Eliminative Pluralism”	Moss, Lenny (1992) “A Kernel of Truth? On the Reality of the Genetic Program”

	Walton, David (1991) “The Units of Selection and the Bases of Selection”	Ereshefsky, Marc (1991) “Species, Higher Taxa, and the Units of Evolution”	Kauffman, Stuart A. (1990) “The Sciences of Complexity and ‘Origins of Order’”
	Lloyd, Elisabeth A. (1989) “A Structural Approach to Defining Units of Selection”		
	Shanahan, Timothy (1990) “Evolution, Phenotypic Selection, and the Units of Selection”		
1994–98	Millstein, Roberta L. (1996) “Random Drift and the Omniscient Viewpoint”	Ereshefsky, Marc (1994) “Some Problems with the Linnaean Hierarchy”	Waters, C. Kenneth (1994) “Genes Made Molecular”
	Sober, Elliott, Wilson, David Sloan (1994) “A Critical Review of Philosophical Work on the Units of Selection Problem”	Ereshefsky, Marc (1998) “Species Pluralism and Anti-realism”	Schaffner, Kenneth F. (1998) “Genes, Behavior, and Developmental Emergentism: One Process, Indivisible?”
		Mayr, Ernst (1996) “What Is a Species, and What Is Not?”	Culp, Sylvia (1995) “Objectivity in Experimental Inquiry: Breaking Data-Technique Circles”
		Stanford, P. Kyle (1995) “For Pluralism and against Realism about Species”	Darden, Lindley, Cook, Michael (1994) “Reasoning Strategies in Molecular Biology: Abstractions, Scans, and Anomalies”
		LaPorte, Joseph (1997) “Essential Membership”	Schaffner, Kenneth F. (1994) “Interactions among Theory, Experiment, and Technology in Molecular Biology”
1999–2003	Walsh, Denis M. (2002) “The Trials of Life: Natural Selection and Random Drift”	Ereshefsky, Marc (2002) “Linnaean Ranks: Vestiges of a Bygone Era”	Godfrey-Smith, Peter (2000) “On the Theoretical Role of ‘Genetic Coding’”
		Coleman, Keith A., Wiley, E. O. (2001) “On Species Individualism: A New Defense of the Species-as-Individuals Hypothesis”	Thagard, Paul (2003) “Pathways to Biomedical Discovery”
			Hardcastle, Valerie Gray (1999) “Scientific Papers Have Various Structures”
			Smith, John Maynard (2000) “The Concept of Information in Biology”

Table A2 (*Continued*)

Period	NATURAL SELECTION (81)	SPECIES (93)	GENETICS (194)
2004–8	Lloyd, Elisabeth A. (2005) "Why the Gene Will Not Return"	Devitt, Michael (2008) "Resurrecting Biological Essentialism"	Stegmann, Ulrich E. (2005) "Genetic Information as Instructional Content"
	Stephens, Christopher (2004) "Selection, Drift, and the 'Forces' of Evolution"	Ereshefsky, Marc, Matthen, Mohan (2005) "Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory"	Stotz, Karola (2006) "With 'Genes' Like That, Who Needs an Environment? Postgenomics's Argument for the 'Ontogeny of Information'"
	Brandon, Robert N., Nijhout, H. Frederik (2006) "The Empirical Nonequivalence of Genic and Genotypic Models of Selection: A (Decisive) Refutation of Genic Selectionism and Pluralistic Genic Selectionism"	Crane, Judith K. (2004) "On the Metaphysics of Species" Franklin, L. R. (2007) "Bacteria, Sex, and Systematics"	Franklin, L. R. (2007) "Bacteria, Sex, and Systematics"
	Waters, C. Kenneth (2005) "Why Genic and Multilevel Selection Theories Are Here to Stay"	Haber, Matthew H., Hamilton, Andrew (2005) "Coherence, Consistency, and Cohesion: Clade Selection in Okasha and Beyond"	
2009–13	Matthen, Mohan, Ariew, André (2009) "Selection and Causation"	Ereshefsky, Marc (2010) "What's Wrong with the New Biological Essentialism"	Perini, Laura (2011) "Sequence Matters: Genomic Research and the Gene Concept"
	Matthen, Mohan (2009) "Drift and 'Statistically Abstractive Explanation'"	Ramsey, Grant, Siebels Peterson, Anne (2012) "Sameness in Biology"	Baetu, Tudor M. (2011) "A Defense of Syntax-Based Gene Concepts in Postgenomics: Genes as Modular Subroutines in the Master Genomic Program"
			Fagan, Melinda Bonnie (2013) "The Stem Cell Uncertainty Principle"
2014–15	Gouvêa, Devin Y. (2015) "Explanation and the Evolutionary First Law(s)"	Ereshefsky, Marc (2014) "Species, Historicity, and Path Dependency"	Griffiths, Paul E., Pocheville, Arnaud, Calcott, Brett, Stotz, Karola, Kim, Hyunju, Knight, Rob (2015) "Measuring Causal Specificity"

Table A3. Top-20 Articles Most Strongly Associated with Selected Models-and-Simulations Topics

Period	Modeling (100)	Models-and-Representation (116)	Computer-Simulation (136)
1934–38			
1939–43			
1944–48			
1949–53			Elsasser, Walter M. (1951) “Quantum Mechanics, Amplifying Processes, and Living Matter”
1954–58			Fitch, Frederic B. (1958) “Representation of Sequential Circuits in Combinatory Logic”
1959–63			
1964–68			
1969–73			
1974–78			Nelson, R. J. (1976) “On Mechanical Recognition” Cummins, Robert (1977) “Programs in the Explanation of Behavior” Heffernan, James D. (1978) “Some Doubts about ‘Turing Machine Arguments’” Laing, Richard (1974) “Maxwell’s Demon and Computation”
1979–83	Beatty, John (1980) “Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology”	Schwartz, Robert (1980) “Imagery—There’s More to It Than Meets the Eye”	
1984–88		Sterelny, Kim (1986) “The Imagery Debate”	Nelson, R. J. (1987) “Machine Models for Cognitive Science”

Table A3 (*Continued*)

Period	Modeling (100)	Models-and-Representation (116)	Computer-Simulation (136)
			Stabler, Edward P. (1988) “Learning Simple Things: A Connectionist Learning Problem from Various Perspectives”
1989–93			Kauffman, Stuart A. (1990) “The Sciences of Complexity and ‘Origins of Order’”
			Wallis, Charles S. (1990) “Stich, Content, Prediction, and Explanation in Cognitive Science”
1994–98	Koperski, Jeffrey (1998) “Models, Confirmation, and Chaos”	Sargent, Pauline (1996) “On the Use of Visualizations in the Practice of Science”	Trenholme, Russell (1994) “Analog Simulation”
			Hogarth, Mark (1994) “Non-Turing Computers and Non-Turing Computability”
1999–2003	Smith, Sheldon R. (2001) “Models and the Unity of Classical Physics: Nancy Cartwright’s Dappled World”	Bokulich, Alisa (2003) “Horizontal Models: From Bakers to Cats”	Parker, Matthew W. (2003) “Undecidability in Rn: Riddled Basins, the KAM Tori, and the Stability of the Solar System”
		Shelley, Cameron (1999) “Multiple Analogies in Archaeology”	
	Odenbaugh, Jay (2003) “Complex Systems, Trade-Offs, and Theoretical Population Biology: Richard Levin’s ‘Strategy of Model Building in Population Biology’ Revisited”	Chemero, Anthony (2000) “Anti-representationalism and the Dynamical Stance”	
		Sismondo, Sergio, Chrisman, Nicholas (2001) “Deflationary Metaphysics and the Natures of Maps”	

2004–8	Weisberg, Michael, Reisman, Kenneth (2008) “The Robust Volterra Principle”	Contessa, Gabriele (2007) “Scientific Representation, Interpretation, and Surrogative Reasoning”	Piccinini, Gualtiero (2007) “Computing Mechanisms”
	Lehtinen, Aki, Kuorikoski, Jaakko (2007) “Computing the Perfect Model: Why Do Economists Shun Simulation?”	Perini, Laura (2005) “The Truth in Pictures”	Bub, Jeffrey (2008) “Quantum Computation and Pseudotelepathic Games”
	Morrison, Margaret (2007) “Where Have All the Theories Gone?”	Morrison, Margaret (2007) “Where Have All the Theories Gone?”	Datteri, Edoardo, Tamburrini, Guglielmo (2007) “Biorobotic Experiments for the Discovery of Biological Mechanisms”
	Alexandrova, Anna (2008) “Making Models Count”	Suárez, Mauricio (2004) “An Inferential Conception of Scientific Representation”	
	Weisberg, Michael (2006) “Robustness Analysis”	Knuuttila, Tarja (2005) “Models, Representation, and Mediation”	
		Perini, Laura (2005) “Visual Representations and Confirmation”	
		Bueno, Otávio (2006) “Representation at the Nanoscale”	
2009–13	Rohwer, Yasha, Rice, Collin (2013) “Hypothetical Pattern Idealization and Explanatory Models”	Wojtach, William T. (2009) “Reconsidering Perceptual Content”	Piccinini, Gualtiero (2010) “The Resilience of Computationalism”
	Houkes, Wybo, Vaesen, Krist (2012) “Robust! Handle with Care”	Rohwer, Yasha, Rice, Collin (2013) “Hypothetical Pattern Idealization and Explanatory Models”	Hagar, Amit (2009) “Active Fault-Tolerant Quantum Error Correction: The Curse of the Open System”
	Lloyd, Elisabeth A. (2010) “Confirmation and Robustness of Climate Models”	Kulvicki, John (2010) “Knowing with Images: Medium and Message”	
	Justus, James (2012) “The Elusive Basis of Inferential Robustness”	Shech, Elay (2013) “What Is the Paradox of Phase Transitions?”	
	Parker, Wendy S. (2011) “When Climate Models Agree: The Significance of Robust Model Predictions”	Batterman, Robert W., Rice, Collin C. (2014) “Minimal Model Explanations”	

Table A3 (*Continued*)

Period	Modeling (100)	Models-and-Representation (116)	Computer-Simulation (136)
2014–15	Fumagalli, Roberto (2015) “No Learning from Minimal Models” Bedau, Mark A. (2014) “Testing Bottom-Up Models of Complex Citation Networks” Batterman, Robert W., Rice, Collin C. (2014) “Minimal Model Explanations” Ross, Lauren N. (2015) “Dynamical Models and Explanation in Neuroscience” Lange, Marc (2015) “On ‘Minimal Model Explanations’: A Reply to Batterman and Rice”		

REFERENCES

Aggarwal, Charu C., and ChengXiang Zhai. 2012. *Mining Text Data*. New York: Springer.

Barron, Alexander T. J., Jenny Huang, Rebecca L. Spang, and Simon DeDeo. 2018. "Individuals, Institutions, and Innovation in the Debates of the French Revolution." *Proceedings of the National Academy of Sciences* 115 (18): 4607–12.

Blei, David M., and John D. Lafferty. 2009. "Topic Models." In *Text Mining: Classification, Clustering, and Applications*, ed. Ashok N. Srivastava and Mehran Sahami, 71–94. Boca Raton, FL: CRC.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Buckner, Cameron, Mathias Niepert, and Colin Allen. 2011. "From Encyclopedia to Ontology: Toward Dynamic Representation of the Discipline of Philosophy." *Synthese* 182 (2): 205–33.

Byron, Jason M. 2007. "Whence Philosophy of Biology?" *British Journal for the Philosophy of Science* 58 (3): 409–22.

Chartier, Jean-François, and Jean-Guy Meunier. 2011. "Text Mining Methods for Social Representation Analysis in Large Corpora." *Papers on Social Representations* 20 (37): 1–47.

Churchman, C. West. 1984. "Early Years of the Philosophy of Science Association." *Philosophy of Science* 51 (1): 20–22.

Cohen Priva, Uriel, and Joseph L. Austerweil. 2015. "Analyzing the History of Cognition Using Topic Models." *Cognition* 135 (February): 4–9.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41 (6): 570–606.

Douglas, Heather. 2010. "Engagement for Progress: Applied Philosophy of Science in Context." *Synthese* 177 (3): 317–35.

———. 2019. "A History of the PSA before 1970." Philosophy of Science Association website. https://philsci.org/about-the-psa/history-of-the-association.html.

Evans, James A., and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42:21–50.

Firth, John R. 1957. "A Synopsis of Linguistic Theory, 1930–1955." In *Studies in Linguistic Analysis*, 1–32. Oxford: Blackwell.

Giere, Ronald N. 1996. "From Wissenschaftliche Philosophie to Philosophy of Science." In *Origins of Logical Empiricism*, ed. Ronald N. Giere and Alan W. Richardson, 335–54. Minnesota Studies in the Philosophy of Science 16. Minneapolis: University of Minnesota Press.

Griffiths, Thomas L., and Mark Steyvers. 2004. "Finding Scientific Topics." *Proceedings of the National Academy of Sciences* 101 (sup. 1): 5228–35. https://doi.org/10.1073/pnas.0307752101.

Griffiths, Thomas L., Mark Steyvers, and Joshua B. Tenenbaum. 2007. "Topics in Semantic Representation." *Psychological Review* 114 (2): 211–44.

Hardcastle, Gary L., and Alan W. Richardson. 2003. "Introduction: Logical Empiricism in North America." In *Logical Empiricism in North America*, ed. Gary L. Hardcastle and

Alan W. Richardson, vii–xxix. Minnesota Studies in the Philosophy of Science 18. Minneapolis: University of Minnesota Press.

Hicks, Daniel J., and Evelyn Brister. 2018. "Contributions of Women to 20th Century Philosophy of Science." Poster presented at the 2018 PSA Biennial Meeting, Seattle, November.

Howard, Don. 2003. "Two Left Turns Make a Right: On the Curious Political Career of North American Philosophy of Science at Midcentury." In *Logical Empiricism in North America*, edited by Gary L. Hardcastle and Alan W. Richardson, 25–93. Minnesota Studies in the Philosophy of Science 18. Minneapolis: University of Minnesota Press.

Hu, Yuening, Jordan Boyd-Graber, Brianna Satinoff, and Alison Smith. 2014. "Interactive Topic Modeling." *Machine Learning* 95 (3): 423–69.

Lynam, Timothy. 2016. "Exploring Social Representations of Adapting to Climate Change Using Topic Modeling and Bayesian Networks." *Ecology and Society* 21 (4). https://doi.org/10.5751/ES-08778-210416.

Malisoff, William M. 1944. "Editorial." *Philosophy of Science* 11 (1): 1–2.

Marcus, Mitchell P., Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. "Building a Large Annotated Corpus of English: The Penn Treebank." *Computational Linguistics* 19 (2): 313–30.

Mauri, Michele, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi. 2017. "RAWGraphs: A Visualisation Platform to Create Open Outputs." In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter—CHItaly '17*, 1–5. Cagliari, Italy: ACM. https://doi.org/10.1145/3125571.3125585.

Mimno, David. 2012. "Computational Historiography: Data Mining in a Century of Classics Journals." *Journal on Computing and Cultural Heritage* 5 (1). https://dl.acm.org/citation.cfm?doid=2160165.2160168.

Murdock, Jaimie, Colin Allen, and Simon DeDeo. 2017. "Exploration and Exploitation of Victorian Science in Darwin's Reading Notebooks." *Cognition* 159 (February): 117–26.

Nicholson, Daniel J., and Richard Gawne. 2015. "Neither Logical Empiricism Nor Vitalism, but Organicism: What the Philosophy of Biology Was." *History and Philosophy of the Life Sciences* 37 (4): 345–81.

Nikolenko, Sergey I., Sergei Koltcov, and Olessia Koltsova. 2017. "Topic Modelling for Qualitative Studies." *Journal of Information Science* 43 (1): 88–102.

Peirson, B. R. Erick, Erin Bottino, Julia L. Damerow, and Manfred D. Laubichler. 2017. "Quantitative Perspectives on Fifty Years of the *Journal of the History of Biology*." *Journal of the History of Biology* 50 (4): 695–751.

Peldszus, Andreas, and Manfred Stede. 2013. "From Argument Diagrams to Argumentation Mining in Texts: A Survey." *International Journal of Cognitive Informatics and Natural Intelligence* 7 (1): 1–31.

Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155 (2): 945–59.

Ramsey, Grant, and Charles H. Pence. 2016. "EvoText: A New Tool for Analyzing the Biological Sciences." *Studies in History and Philosophy of Science* C: *Studies in History and Philosophy of Biological and Biomedical Sciences* 57 (June): 83–87.

Reisch, George A. 2005. *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic*. Cambridge: Cambridge University Press.

Schmid, Helmut. 2013. “Probabilistic Part-of-Speech Tagging Using Decision Trees.” In *New Methods in Language Processing*, ed. D. B. Jones and H. L. Somers. London: Routledge.

Srivastava, Ashok N., and Mehran Sahami. 2009. *Text Mining: Classification, Clustering, and Applications*. Boca Raton, FL: CRC.

Swanson, Reid, Brian Ecker, and Marilyn Walker. 2015. “Argument Mining: Extracting Arguments from Online Dialogue.” In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 217–26. https://www.aclweb.org/anthology/W15-4600.

Turney, Peter D., and Patrick Pantel. 2010. “From Frequency to Meaning: Vector Space Models of Semantics.” *Journal of Artificial Intelligence Research* 37:141–88.

Widdows, Dominic. 2004. *Geometry and Meaning*. Stanford, CA: CSLI.