

Grajzl, Peter; Murrell, Peter

Working Paper

A Structural Topic Model of the Features and the Cultural Origins of the Baconian Program

CESifo Working Paper, No. 6443

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Grajzl, Peter; Murrell, Peter (2017) : A Structural Topic Model of the Features and the Cultural Origins of the Baconian Program, CESifo Working Paper, No. 6443, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/161882>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A Structural Topic Model of the Features and the Cultural Origins of the Baconian Program

Peter Grajzl, Peter Murrell

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl

www.cesifo-group.org/wp

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: www.CESifo-group.org/wp

A Structural Topic Model of the Features and the Cultural Origins of the Baconian Program

Abstract

We use machine-learning methods to study the features and origins of the Baconian program, a cultural and methodological paradigm viewed as providing the intellectual roots for modern economic growth. After building a machine-readable corpus of Bacon's works, we estimate a structural topic model, a state-of-the-art technique for analysis of text corpora. The estimates uncover the dominant themes in Bacon's opus, clearly identifying two central to the Baconian program: the emphasis on probing for facts and the epistemology of deriving lessons from those facts. Examining the connectedness of the themes, clear evidence links Bacon's epistemology to his jurisprudence. The emphasis on seeking facts is linked to Bacon's epistemology but is less connected with other themes, and therefore much more *sui generis* with Bacon. The utilitarian promise of science, embraced by Bacon's followers, is not emphasized by Bacon. Finally, we demonstrate how Bacon's use of the different themes varies with the intended audience and his chosen medium.

JEL-Codes: B310, Z100, C550, N730, K100, P100.

Keywords: Baconian program, culture, law, knowledge, natural philosophy, politics, religion.

Peter Grajzl
Department of Economics
The Williams School of Commerce,
Economics, and Politics
Washington and Lee University
USA - 24450 Lexington VA
grajzlp@wlu.edu

Peter Murrell
Department of Economics
University of Maryland
USA - College Park, MD 20742
murrell@econ.umd.edu

April 1, 2017

With regard to authority, it is the greatest weakness to attribute infinite credit to particular authors, and to refuse his own prerogative to time, the author of all authors, and, therefore, of all authority. For truth is rightly named the daughter of time, not of authority.

Francis Bacon, *Novum Organum*

1. Introduction

We use recently developed machine-learning techniques to revisit the source of the ideas underlying the cultural and methodological paradigm that is often referred to as 'the Baconian program'. For some, Francis Bacon's (1561-1626) ideas were the very rock on which the modern world was built, encouraging the development of modern science and the application of technology to economic activities. While Bacon is one of the most studied thinkers ever, the vast scope of his learning and the enigmatic quality of his writing has led to a wide range of disparate ideas on the key elements of his thought and a variety of conjectures about the source of that thought. Drawing on novel analytical techniques designed for the purpose of analysis of large text corpora, we re-examine the features and origins of the Baconian program, a topic of enduring interest among scholars of Bacon's opus as well as economic historians (see, e.g., Peltonen 1992, 1996a; Vickers 1992, Mokyr 2005, 2016).

Culture has been awarded a prominent role in the recent literature on long-run economic development (see, e.g., Barro and McCleary 2003, Guiso et al. 2006, Tabellini 2008, Algan and Cahuc 2010, Bowles and Gintis 2011, Nunn 2012, Alesina and Giuliano 2015), with that literature often focusing on the deep historical roots of cultural norms (see, e.g., Nunn 2008, Becker et al. 2016, Guiso et al. 2016). In a recent important contribution, Mokyr (2016) argues that the key event in the initiation of modern economic development was the creation and spread of a "culture of growth". Specific to early modern Europe, this culture reoriented its adherents to the possibilities of scientific advance and invention and encouraged their application to economic activities, stimulating technological progress and economic growth. Mokyr identifies Bacon, the preeminent English philosopher, statesman, scientist, jurist, and author, as the principal "cultural entrepreneur" and the Baconian program and its emphasis on a new methodology of practical investigation as fundamental in the emergence and diffusion of economic-growth-facilitating cultural ideas and modes of thought (Mokyr 2005, 2006, 2010).

But from what spring did Bacon's thought and cultural perspective flow? Bacon, while possessing an avid interest in natural philosophy, was, after all, not a distinguished scientist (Drinker Bowen 1993: 9). Hence, the question of the source of the many dimensions of the Baconian program, and their connectedness, has inspired a lively literature in the history of science, political philosophy, and law (see, e.g., Vickers 1992, Peltonen 1992, Craig 2010, Kocher 1957, Matthews 2008). Many scholars, including Mokyr (2016), portray Bacon as *sui generis*, a one-of-a-kind genius alone as the 17th century began. Others suggest that the Baconian program was heavily influenced by Bacon's theological beliefs (e.g., Matthews 2008). Or,

notwithstanding Bacon's critique of Aristotelian epistemology, others see a strong influence from Bacon's early exposure to classical thought (see, e.g., Peltonen 2004).

Even though Bacon rose to the pinnacle of the legal profession and was a stout defender of the common-law (Peltonen 1996b: 290), the scholarly discussion of the genesis of Bacon's ideas has devoted remarkably little attention to his legal background and training.¹ Only a handful of scholars (e.g., Holdsworth 1927: 10; Shapiro 1968: 736; 2000: 108; Kocher 1957, Stone de Montpensier 1968: 454, Coquillette 1992, Martin 1992) have even raised the possibility that Bacon's methodology may have originated in his rigorous training in jurisprudential thought and the common-law system of case law. Kocher (1957), Martin (1992) and Shapiro (1968, 2000) are, to our knowledge, the only contributions that explicitly draw a connection between Bacon's scientific methodology and elements of his jurisprudential work.² Their ideas have not diffused broadly across the wider range of Bacon scholars. Peltonen (1996a: 22-23), in the introduction to what is perhaps the most comprehensive current overview of Bacon's contributions (Peltonen 1996c), only briefly mentions Bacon's legal opus and jurisprudential thought, and then only in the context of law reform. Neither Peltonen (1996a) nor any of the other chapters in Peltonen (1996c) explicitly link Bacon's legal thought to the core of Bacon's scientific method.³

Understanding the cultural and conceptual origins of the Baconian program is important for economists and social scientists more generally, for a number of reasons. First, the study of Bacon's opus holds potential to offer insight into how a set of cultural ideas and modes of thinking applicable in one societal domain may be related to, or even directly influenced by, methodological ideas and cultural paradigms prevalent in other domains. In the case of the Baconian program, we have the fundamental example of the potential connection between science and technology as a force for progress (see, e.g., Mokyr 2016) and ideas in politics, religion, and law. Second, Bacon was a very powerful political actor, jurist, and policy-maker in England at the beginning of a historical period that has reached iconic status in the story of the development of democracy and capitalism (see, e.g., North and Weingast 1989, Acemoglu and Robinson 2012, Murrell 2016). Understanding whether the Baconian program was *sui generis*

¹ On this point, see de Montmorency (1905: 263), Holdsworth (1927: 10), and Shapiro (1980: 333).

² Holdsworth (1927: 10) notes that "it is at least arguable" that Bacon's training in law "had some influence on his inductive system of experimental philosophy". Coquillette (1992) makes a conjecture about the connection between Bacon's law and methodology anecdotally in a study of Bacon's legal writings. Stone de Montpensier (1968) remarks that Bacon's logic of reasoning in *Maxims of the Law* resembles his scientific method. Wheeler (1999, reflecting ideas first proposed in the late 1940's) drew connections between Bacon's work in the common law and most other aspects of English culture: "Several things unique to 'the modern' in England were all 'unwritten' in origin: the common law, constitutionalism, scientific empiricism and cricket...The English did culture differently. One of the reasons they were able to produce an unusually strong and dynamic culture was because of a distinctive feature of their cultural template: the art of making the unwritten intelligible. The basis for this lay in their unique unwritten common law. The principles they developed for judging and discovering the dictates of the unwritten law were extended to nearly everything...."

³ In Peltonen (1996c) the only reference to Kocher (1957) is the following comment: "That certain statements of a general nature could be fruitfully applied to more than one case may well be a conviction deriving from his belief in the power of maxims in Jurisprudence" Kusakawa (1996: 54). This is not followed up by any other stronger statements about the origins of Bacon's views.

provides perspective on the question of how important the ideas and actions of influential individuals could be in shaping fundamental institutional and economic change, compared to deeper rooted forces (see, e.g., Jones and Olken 2005, Besley et al. 2011, Brender and Drazen 2013). Third, there is no doubt that the ideas associated with Bacon were adopted by large numbers of English followers as the 17th century progressed. But this could be simply because Bacon was a particularly insightful messenger for a larger cultural milieu. While this paper does not address the issue of why the ideas expressed by Bacon were adopted, we are able to show that they were indeed reflective of early 17th century English culture, particularly legal culture.

Departing from existing analyses of Bacon, we examine the cultural and conceptual origins of the Baconian program using topic modeling (see, e.g., Blei 2012), a recently developed unsupervised machine-learning technique for analysis of large text corpora. Topic modeling has become part of the economists' toolkit only very recently. The Latent Dirichlet Allocation (LDA) model (Blei et al. 2003, Blei 2012) in particular has been applied by scholars to understand, or develop new quantitative measures of, monetary policymaking (see, e.g., Fligstein et al. 2014, Hansen and McMahon 2016, Hansen et al. 2015), financial market performance (see, e.g., Larsen and Thorsrud 2015, Huang et al. 2016), and corporate behavior (e.g., Bandiera et al. 2016, Bellstam et al. 2016). We move beyond the application of LDA, and use the Structural Topic Model (Roberts et al. 2014, 2016a), a novel topic-modeling framework that incorporates document-level covariate information into the analysis.

Given the tremendous scope and breadth of Bacon's writings, the Baconian program lends itself as particularly suitable for topic modeling. This technique has a property that endows our analysis with a distinctive character. In contrast to human analysts, topic models use a semantically neutral 'bag-of-words' conceptualization of texts, enabling the researcher to produce results that are devoid of preconceived notions about the content of texts. (The interpretation of the results is, of course, another matter.) This is particularly useful in the study of such an icon as Bacon, where his own words have been overlain by centuries of commentary, criticism, interpretation, and reinterpretation. Therefore, avoiding this literature and going back directly to Bacon's own words has a strong appeal. But doing so as a 21st-century reader presents its own obstacles, not the least of which is understanding the peculiarities of the orthography, the grammar, the phrasing, and the word-usage of the English of Shakespeare's time. Machine-learning techniques provide a different direct path to Bacon. Nevertheless, we should emphasize very strongly that we are not making the argument that these techniques are superior to existing textual-analysis methods. Instead, machine-learning is one more, hitherto unexplored, route to obtaining insight into the nature and sources of the ideas of one of history's great intellects.

To study the Baconian program using topic modeling, we built a comprehensive, machine-readable corpus of Bacon's major writings, comprising nearly 900,000 words. We associated each document with the values of variables characterizing the form of Bacon's work (e.g. an essay, a letter etc.) and its targeted audience (e.g. philosophers, politicians etc.). As one influential scholar of Bacon has emphasized (Peltonen 1992, 1996a: 9-10), Bacon's positions and

ideas varied widely across audiences that he addressed and with the form of his writing. Thus, the inclusion of the text covariates is particularly pertinent in the context of our analysis.

We then specified and estimated a Structural Topic Model (STM). Our analysis first identifies sixteen key substantive themes, or topics, in the corpus—topics that would not necessarily be obvious for a reader of many disparate documents. The resulting topics provide novel interpretations of the ideas that spanned many of Bacon's works. By closely examining the specific words and the particular documents associated with each topic, we give each topic an evocative name suggestive of its contents. Thus our unsupervised machine-learning algorithm answers a question that it is still debated in the critical literature: 'What is in Bacon's work?' (see, e.g., Peltonen 1996a, Vickers 1992).

In particular, STM clearly pinpoints two central themes that the interpretative literature invariably places at the center of the Baconian program: the emphasis on probing for facts and the use of induction to generalize from the facts. The examination of the content of the estimated topics provides lessons on which topics arise from which areas of Bacon's many scholarly, political, and social activities, giving insights into the cultural and conceptual origins of Bacon's thought.

Second, drawing on STM's identification of the principal themes used by Bacon, we examine correlations across documents in the use of disparate topics and also commonalities between topics in the use of distinctive vocabulary. Most notably, our analysis demonstrates that Bacon's epistemological reasoning is strongly connected to his ideas about law and legal institutions, a result pointing to the jurisprudential origins of the epistemological core of the Baconian program. In contrast, we find the extent of connectedness between Bacon's emphasis on probing for facts and his legal ideas weaker. We interpret this as evidence that Bacon's emphasis on pursuing and ascertaining facts was more *sui generis* than was his common-law-inspired inductive epistemology.

Finally, we make direct use of the defining feature of STM, the integration of document-level (metadata) information into estimation, exploring whether Bacon's emphases on specific beliefs and ideas about a society's institutions and organizations varied with the intended audience and the form taken by his writing. Peltonen (1992, 1996a), for example, contends that various, seemingly conflicting, aspects of Bacon's work can only be understood when taking into account both the target audience and the chosen writing form. Using STM, we provide quantitative evidence that supports this argument. We find, for instance, that Bacon communicated his ideas and beliefs about probing for facts mostly to philosophers and expressed them primarily as axiomatic statements in the form of apothegms. In contrast, his epistemological reasoning was targeted to a wider audience, using book-length tomes that would reach beyond his aristocratic circles (see, e.g., Peltonen 1996a: 10). The wider audience, more interested in methodology, would include those most likely to put such ideas to socially productive use (Mokyr 2016). We also find evidence that Bacon relied on essays not only to communicate his ideas about civic life and, in particular, political strategies, as argued by some

Baconian scholars (Peltonen 1996b: 294), but also ideas about foreign affairs and religious matters. Interestingly, Bacon's insights about political strategies were targeted much more to philosophers than to politicians, a finding consistent with the prevailing view of Bacon as a highly engaged and very shrewd political actor and accomplished political philosopher (see, e.g., Peltonen 1996b). His thoughts about the intertwined nature of law and high politics, however, were expressed primarily in his (private) letters, a finding reflective of Bacon's own private political ambitions.

The rest of the paper is organized as follows. We first briefly lay out the nature of structural topic modeling, emphasizing its novel features. Then, we outline the steps we took in the construction of a large machine-readable corpus of Bacon's writings. We next describe the two variables, audience and writing form, that are associated with the documents. Using the STM estimates, we examine and name the sixteen topics that are found within Bacon's corpus. We draw lessons using distinctive features of STM's outputs—correlations of topic usage across documents, overlapping word usage between topics, and the variation of topic usage across audience and writing-form. Throughout the analysis we develop insights into the cultural and conceptual origins of Bacon's thought.

2. Structural Topic Modeling

The Structural Topic Model (STM), introduced by Roberts et al. (2014, 2016a), is a machine learning tool that belongs to the family of generative probabilistic models developed for purposes of analysis of large text corpora. Generative probability models require a researcher to postulate a model of a data generating process and then use the corpus data to find the most likely values for the parameters within the model. Topic models such as the STM, as well as Latent Dirichlet Allocation (LDA) (Blei et al. 2003, Blei 2012) and the related Correlated Topic Model (CTM) (Blei and Lafferty 2007), view texts as 'bags of words' and exploit the co-occurrence of words across many documents with the aim of identifying groups of words that tend to co-occur (Tingley 2017).

As the names suggest, the key output of topic models is a set of 'topics'. The topics are the themes or ideas of the corpus as identified by the unsupervised machine-learning algorithm. The name and interpretation associated with each topic are assigned by the researcher by examining the words that are most important for each specific topic and the particular texts in which a given topic is most used. However, the topics themselves are purely a product of the estimation of a model rather than being produced, as they would be in a supervised model, in order to match words and documents to concrete issues specified in advance by the researcher. Therefore, the estimated topics identify the substantive themes that might not be readily apparent to a reader of many disparate documents (Blei 2012). Importantly, the estimated topics cannot be a product of the many interpretations of Bacon that have overlain his works in the four centuries since his death.

The basic structure of the STM is as follows (see, e.g., Roberts et al. 2016a). There are D documents, indexed by d . The document generating process views a document, d , as beginning

with a collection of N_d empty positions, each of which is to be filled with a word. The process of filling a position first involves a choice of a topic from a fixed number available, indexed by $k \in \{1, \dots, K\}$. The 'bag-of-words' assumption implies that all positions in a document are interchangeable, meaning that the process determining the choice of topic for any empty position in a specific document is the same for all positions in that document. That process is characterized by a K -dimensional vector of the parameters of a distribution that generates one of the topics $k \in \{1, \dots, K\}$ for each position in the document. This vector is usually referred to as the topic-prevalence, or topic-attention, vector (Roberts et al. 2014, 2016b). A topic-prevalence vector, then, lists the probabilities that each of the K topics will be assigned to an empty position. Then, given the chosen topic, k , there is the choice of a word from a corpus-level vocabulary, the elements of which are indexed by $v \in \{1, \dots, V\}$. A topic, then, is a V -dimensional vector specifying the probabilities that each element of the vocabulary will be chosen to fill an empty position given that topic.⁴

Early versions of topic models, such as the LDA and CTM, viewed the topic-prevalence vector as constant across all documents. Thus, all documents were assumed to be generated by one and the same process, regardless of the document-level characteristics of the text (e.g. author, textual form, intended audience etc.). The key innovation of STM vis-à-vis earlier topic models is to allow the topic-prevalence vector to vary across documents with the document-level values of metavariables. STM thus allows the researcher to incorporate prior information into the model of the data-generating process for the corpus. Incorporation of the extra information embedded in the metadata improves the identification of topics (Roberts et al. 2014) and allows the researcher to estimate the effect of metadata covariates on the prevalence of specific topics.

The key features of the data-generating process are as follows (Roberts et al. 2014, 2016a, 2016b). The process of filling a word-position in a document begins with the generation of a document-specific topic-prevalence vector using metadata information. Let the document-level metadata be given by a matrix of covariates \mathbf{X} , each row (denoted \mathbf{x}_d) listing the values of all metavariables for document d . Then, the topic-prevalence vector for document d ($\boldsymbol{\theta}_d$) is drawn from a logistic-normal distribution whose parameters are a function of document metadata covariates:

$$\boldsymbol{\theta}_d \sim \text{LogisticNormal}(\boldsymbol{\Gamma}\mathbf{x}_d', \boldsymbol{\Sigma}). \quad (1)$$

$\boldsymbol{\Gamma}$ is a matrix of coefficients relating metavariables to topic-prevalence. The structure of $\boldsymbol{\Sigma}$ will imply that there could be correlations across documents in the topic-prevalence vector.

Now turn to the process of filling empty position $n \in \{1, \dots, N_d\}$ in document d . Given the topic-prevalence vector, one specific topic, denoted by z_{dn} , is associated with that position through the following process:

⁴ In some applications of STM, these probabilities are allowed to vary in a systematic way across documents. In our application, we do not use this feature of STM.

$$z_{dn} \sim \text{Multinomial}(\theta_d) \quad (2)$$

where the k^{th} element of z_{dn} is unity and all other elements are zero if topic k is chosen.

To emphasize the innovative element of STM, the choice of a specific vocabulary word, v , is modeled as a function of two parameters, one indicating the baseline importance of that word across all documents, m_v , and one indicating the importance of the word given the topic (k), κ_{kv} . Transforming the sum of these coefficients into probabilities for use in a multinomial distribution via a logistic transformation, one obtains:

$$\beta_{dkv}|z_{dn} \propto \exp(m_v + \kappa_{kv}) \quad (3)$$

where β_{dkv} is the probability of choosing vocabulary word v to fill a position in document d given topic k . Then a specific word, denoted w_{dn} , is chosen from the overall corpus vocabulary to fill position n in document d , using the following process:

$$w_{dn} \sim \text{Multinomial}(\beta_{dk1}, \dots, \beta_{dkV}). \quad (4)$$

The data to be used for estimation are all words in all documents, that is, w_{dn} , $n = 1, \dots, N_d$ and $d = 1, \dots, D$ and the metadata matrix, \mathbf{X} . Given these data, one estimates Γ , θ_d , m_v , κ_{kv} , and β_{dkv} by maximizing the posterior likelihood that the observed data were indeed generated by the above data-generating process. Computationally, the maximization problem is solved using an iterative approximation-based variational expectation-maximization algorithm available in R's `stm` package (see Roberts et al. 2016a, 2016b).

3. Data

3.1. The Corpus of Bacon's Works

The construction of the corpus began with a search for as many digital works of Bacon as could be found for which the quality of the text was sufficiently good that a large percentage of words could be machine read. The source of documents were established repositories that specialize in making available machine-readable forms of old documents, such as Project Gutenberg (n.d.), Hathi Trust (n.d.), Internet Archive (n.d.), and, most importantly, The Text Creation Partnership for Early English Books Online (2014).

The larger works were broken up into smaller documents in a manual process that used natural breaks in the text. The result was a corpus of 282 text documents of varying length, containing 898,582 words in total, a mean of 3,186 words per document. Table 1 lists the works included in the corpus. In order to make the documents suitable for a statistical routine that assumed standard orthography and common language for all corpus words, these documents were then processed in a number of stages that are listed below. All operations were carried out using Python programs written by the authors.

The chaotic orthography of late 16th and early 17th century English was converted into standard modern orthography using a program that absorbed the database available with Morph Adorner (2013). This database contains translations between spelling variants and standardized

spelling for words common in 16th and 17th century English. In the process of constructing our corpus, many corrections and additions were made to this database, resulting in translations being available for over 361,000 spelling variants.

Given standardized spelling, word inflections were then removed by converting all words to their lexical roots, again using databases available in Morph Adorner. Corrections and additions were made to these databases, resulting in the availability of translations between standard spellings and lexical roots for over 468,000 words.

The next steps required the use of a comprehensive list of standard English spellings, names, abbreviations, and acronyms. The construction of this list initially relied on databases from Moby Word Lists (2002). With additions made during the processing of the corpus, the resultant word list contained over 385,000 standard English words. If, at this stage, a word in the corpus matched a word in the English dictionary, then it was left in the form derived from the previous processing steps.

A very large proportion of the words that did not produce a match were Latin words. Bacon, as was universal among the educated at the turn of the 17th century, was fluent in Latin. But a peculiar variant of Latin was common at that time, the most distinctive peculiarity being the use of many, often idiosyncratic, accents on letters that do not appear in classical Latin. Therefore, accents were removed from all words that did not appear in the English dictionary, by replacing accented letters with their nearest ASCII equivalent.

All words remaining in the corpus that were not in the English word list were then matched against a word list of over 1 million Latin spellings (naturally including a large number of inflections of Latin root words). This Latin spelling list was obtained from Whitaker's Words (2006). Any word in the Latin spelling list, but not in the English word list, was stemmed using a version of the Schinke Latin stemming algorithm (Schinke et al. 1996) programmed in Python, stemming being a standard process by which the variant forms of a word are translated into standard forms. This process in Latin consists mainly of finding the inflection-less form of a word.

Once the Latin words in the corpus were in their standard forms, translation into English occurred using a Latin-English dictionary. This was created initially using the database available at Whitaker's Words (2006). Then, if a word in the corpus was simultaneously not in the English word list, within the Latin word list, and not in the Latin words in this initial Latin-English dictionary, a Python program searched for the word in the online Lewis and Short Latin Dictionary (Perseus n.d.). The Latin words and their online translations were then added to the original Latin-English dictionary. The resultant dictionary database contained over 59,000 Latin to English translations.

The resulting corpus was then imported into R using the *stm* package. To prepare the corpus for estimation, further text processing was implemented. Using R's *textProcessor* function all words were converted to lower case and the Porter stemming algorithm was applied. Standard

English stop words (natural language words which carry very little meaning, such as 'and', 'the', 'a', 'an'), numbers, and punctuation were removed. The resulting dataset consists of 282 text documents and 147,945 word tokens.

3.2. *The Metadata*

The last step in organizing the data was to associate the text documents with a set of metadata covariates. These variables provide information about each document, helping to improve the precision of overall estimation and facilitating tests of hypotheses concerning how the themes in an author's work vary across subsets of the corpus. In the case of Bacon this step—and hence the use of STM rather than LDA or CTM—is especially necessary given that his output was so broad and so varied in style. He was the consummate polymath and public intellectual, targeting many different types of readers and engaging a large variety of subjects (Peltonen 1992).

We coded two metavariables that capture two different characteristics of Bacon's varied opus, both highlighted as important by scholars of Bacon's opus (Peltonen 1992, 1996a: 9-10). First, there is the intended audience—historians, methodologists, lawyers, politicians, scientists, or philosophers. Second, there is the form that his finished work took—case reports, apothegms, letters, essays, book-length tomes, or speeches. Peltonen (1992, 1996a: 10) argues that Bacon purposefully targeted different audiences with different ideas and even strategically chose to disseminate different ideas in different forms. The coding of both of these metavariables was based on our judgment. The writing form of each text document was readily ascertained. The intended audience was coded after careful scrutiny of each document, simultaneously taking into account existing scholarship on Bacon's opus. Table 2 provides the frequencies of the number of documents within each category of the cross-classification of document form and intended audience.

In the estimations we describe below, we model topical prevalence (or importance), the distinguishing feature of STM, as a function of a linear combination of the dummy variables reflecting our two metadata covariates (see equation (1) above). The implication of this is that the K topics that we identify with Bacon are the same across all documents but the importance, or prevalence, of each topic can vary across document type (θ_d). However, the expected values of topic prevalences are not different for every document. For example, an essay written for historians will have different topic prevalence parameters in its probability generation function than an essay written for lawyers or an apothegm for historians, but it will have the same expected prevalence parameters as all other essays for historians.

4. Estimating and Interpreting the Themes in Bacon's Work

A key initial decision to be made concerns the choice of the (fixed) number of topics to be estimated. There exists no clear-cut or unified approach to this decision (Roberts et al. 2014, 2016b). We therefore proceeded by estimating a series of STM models, with the number of topics varying between from 5 to 20. We then examined the changes in the held-out likelihood

and performed residual analysis to assess how the model's goodness of fit varied with the number of topics (Wallach et al. 2009, Taddy 2012, Roberts et al. 2016b). We also inspected the coherence of the set of topics estimated in each of the models. This process resulted in the selection of an STM model with 16 topics as most appropriate for our Bacon corpus.⁵

Table 3 presents detailed results for the estimated 16-topic STM. The words assigned to each topic come directly from the STM output, and therefore comprise word stems, which in some cases are not English words themselves (e.g., 'statut'). The names assigned to each topic are our interpretations of the set of ideas and themes that underlie each topic. The usual procedure in assigning names is to examine the specific words that are most important within each topic (reflecting estimates of the β_{dkv} in (3) above) and the types of documents for which a topic is most important (reflecting estimates of the θ_d in (1) above). The following paragraphs use this procedure in justifying our interpretations of the topics. This is a crucial step in the analysis because many of the ultimate conclusions that we reach will be about which particular themes span the whole of Bacon's corpus, about how Bacon's usage of different themes varies across documents, and about how usage varies with both audience and document form. How words vary with topics—the basic output of STM—is not important in itself, but rather the crucial findings depend on how the ideas reflected in those words vary with topic.

Readers will note below that we are able to easily identify the underlying themes for all 16 topics and then assign non-controversial names for each. This is evidence for the internal validity of our results. As Mimno et al. (2011) remark, practical applications of topic modeling often result in some topics that seem nonsensical. This leads to the common procedure of presenting example topics and hiding the dubious ones, in order to bolster readers' confidence in the results. We present all estimated topics and provide evocative names that clearly capture topic content.

We present two lists of 30 words for each topic, each list corresponding to the most important words associated with the topic according to a specific criterion. The highest probability ('Highest prob') words are those that are most common (in terms of frequency) for a given topic, but are also non-exclusive, in the sense that they may be associated with any number of topics and might be the highest probability words in more than one topic (as is the case, for example, with 'king', 'law', 'man', 'will' in our corpus). In contrast, 'FREX' words are used more frequently in documents highly associated with a topic than in documents less highly associated with a topic.⁶ Because of their distinctiveness, the FREX words tend to provide more clues for naming the topics than do the highest probability words. But it is worth keeping in mind that the results of STM are driven by the frequency of occurrence of the highest probability words even

⁵ All of our substantive findings are robust to small variations in the number of estimated topics.

⁶ We do however restrict FREX words to those used with some frequency in order not to focus on, say, one word that just happens to be used in one document, and once only. Specifically, our choice of FREX words is characterized by a frequency to exclusivity ratio of 0.25 (see Roberts et al. 2016b).

though such words might seem mundane and provide little direct information about the themes underlying a topic (e.g. 'thing' is a highest probability word in 9 of the 16 topics).⁷

Table 3 begins with two topics that form the core of the Baconian program. Labeling the first of these topics Epistemology needs little justification, the words most highly associated with it being so unified in theme—'philosophy', 'knowledge', 'method', 'system', 'logic', 'inquiry', 'discovery', 'experiment', etc. This topic is most strongly associated with Bacon's two great works of methodology, *Novum Organum* and the *Advancement of Learning*. There can be no doubt about a topic whose most highly associated document is a section of the *Novum Organum* that urges that "We must not only search for, and procure a greater number of experiments, but also introduce a completely different method, order, and progress of continuing and promoting experience. For vague and arbitrary experience is (as we have observed), mere groping in the dark, and rather astonishes than instructs. But when experience shall proceed regularly and uninterruptedly by a determined rule, we may entertain better hopes of the sciences."

The documents associated with the second of the scientific method topics are a mixed bag. There are two of Bacon's most enigmatic works, *New Atlantis* and the *Apothegms* (nine pieces of text in all). There are three charges presented by Bacon the Attorney General. There are several of his essays. Reading through these works and looking for what connects them, one concludes that their common elements comprise questions being asked and answers being provided. In the charges, Bacon the prosecutor asks why a person is guilty and provides the evidence as answer. A large number of the *Apothegms*, a collection of widely disparate, cryptic aphorisms or amusements with obscure broader implications, pose a scene where one person asks and another answers, for example, "Mr. Popham (afterwards Lord chief Justice Popham) when he was Speaker; And the House of Commons had sat long and done, in effect nothing; coming one day to Queen Elizabeth, she said to him; Now Mr. Speaker; what hath passed in the Commons House? He answered, if it please your majesty, seven weeks."⁸ In the associated sections of the *New Atlantis* there are many similar inquisitions and responses, the Governor of Bensalem telling his visitors that "because he that knoweth least is fittest to ask questions, it is more reason, for the entertainment of the time, that ye ask me questions, than that I ask you." Notably of the five sections of *New Atlantis* that we include as separate items of our corpus, those three that are highly associated with this topic do not contain the famous descriptions of the infrastructure of science on Bensalem. There is no 'perfume house' for experiments on smell, no 'confiture house', no 'engine and instrument for all sort of motion', no 'mathematical house where be represent all instrument as well of geometry as astronomy'. This topic is not about organization of science, as some scholars have interpreted parts of *New Atlantis* (Sargent 1996).

⁷ The word 'thing' (or 'particular') was often used as indicating the more modern 'fact' during Bacon's time. It was only during the seventeenth century that the modern usage of 'fact' became much more common (Shapiro 2000). In Bacon's works, the most common use of 'fact' was indicating the deed of which someone was accused in a legal proceeding.

⁸ Modern spelling and slight editing by the authors.

Many of the words associated with this topic refer to particular situations and not broader phenomena—the poisoner (Somerset) and the victim, the chamber and the gallery on the island of Bensalem, and the name Nicholas because a number of the *Apothegms* are stories about the wisdom of Bacon's father, who, like his son, rose to the pinnacle of the legal profession. But there are two much more general words that appear under this topic and not under any other: 'ask' and 'answer'. Both of these words are used with a high probability within this topic, but also are highly specific to this topic (as indicated by FREX). This is consistent with the common thread that we find in the disparate documents most associated with this topic, and we therefore designate it as Probing for Facts.⁹ Thus, STM identifies a key aspect of Bacon's scientific method that has been highlighted by multiple scholars (see, e.g., Peltonen 1996a: 17; Mahlerbe 1996, Mokyr 2005: 289, 304). Notably, however, STM has found this theme in a different set of seemingly unrelated documents and expressed in a different way than in the works of Bacon that are usually identified in the prevailing critical literature.

Whereas the first two topics dealt with Bacon's avocation, the scientific method, the work for which he is most renowned, the next five center on his vocation, the law, for which he is less famous. The topic Land Law reflects Bacon's significant work for the Crown especially as Lord Chancellor, the chief judge of the Chancery court, which had a large role in property cases, many of which in England at that time involved the trusts connected with land and inheritance. Thus, one sees standard words such as 'legacy' and 'estate', but also words highly characteristic of English property relations; 'feoff', for example, is a word-stem connected with the transferring of land to another. The document most associated with this topic is Bacon's disquisition on the Statute of Uses, a 1535 law relevant to the inheritance of and the benefits from land. Many of the legal *Maxims* are also strongly associated with this topic because Bacon used examples from land law to elucidate the more general principles embodied in his maxims.

King, Court & Law reflects the experience of the law that was somewhat more personal to Bacon, arising in connection with official appointments: 'solicitor', 'attorney', 'chancellor' are all associated with this topic. 'Essex' was a patron of Bacon but Bacon was also later engaged in legal proceedings against his former patron. Other names appearing in the topic, 'Manuel' and 'Antonio', were connected with royal court intrigues that Bacon played a role in prosecuting. All of the documents most strongly associated with this topic are letters, but all are on law. For example, a letter to the King reports on Bacon's visit to a sickly Lord Chancellor but also comments on the "Sickness, of your Chancery Court, though, (by the Grace of God) that Cure will be much easier, than the other".

Law & Nation focuses on what constituted the nation and its citizens, and the legal status of its institutions. In the early 17th century, England had acquired a Scottish King, who was head of two separate nations. As Attorney General, Bacon pleaded the government's cause in the great

⁹ Our use of the word 'facts' might be somewhat anachronistic because the modern use of that word only gradually developed during the 17th century. However, as Shapiro (2000) relates, the modern usage had its origin in the law and Bacon was one of the authors who began to use it in the modern sense in many areas of human activity.

legal case in which questions of citizenship arose, that of the Postnati, or Calvin's case. Thus, the stem 'allegi' appears in only this topic as Bacon debates the legal nature of the origins of allegiance. Bacon's plea won and the judges extended the protections of the common law (within England) to all those born in Scotland after 1603. The four documents with which this topic is most highly associated all concern this issue. But there is also a more general essay on Judicature, in which Bacon reminds judges of their allegiance to the sovereign.

As many of the associated words indicate, the Religious Law topic captures legal issues connected with religious controversies and with the status of organizations closely connected with religious bodies. Religious words are numerically dominant in the word lists, but law is one of the highest probability words. In the documents most associated with this topic the religious words are almost invariably used in connection with the discussion of legal issues. Documents highly associated with this topic include those where Bacon suggests compiling, systematizing, and amending the English laws; for example, telling James I that "King Henry the Eighth...was authorized by Parliament to purge the Canon Law, and to make it agreeable to the Law of God, and the Law of the Land". Second highest in these documents is Bacon's discourse on Sutton's case, a landmark in the law concerning the status of organizations dependent on charities, a legal subject inevitably associated with religion in a country where education and medical care were usually provided by charitable religious entities.¹⁰ In another document strongly associated with this topic, Bacon suggests updating the laws relevant to the Church of England, a long-neglected subject in contrast to other laws: "But, if it be said to me, that there is a difference, between civil causes, and ecclesiastical, they may as well tell me, that churches, and chapels, need no reparations, though castles, and houses, do."

The Legal Theory topic is the one in which Bacon discusses themes that are broader than those on one single legal issue. Many of the documents associated with this topic are from his famous *Maxims*, in which Bacon tried to condense the principles of English laws into pithy statements. For example, the seventeenth maxim (the third most strongly associated with this topic) is that "The faith and duty of a judge are not subject to question, but it is otherwise of his knowledge either of law or fact", which contains an obvious resonance with Bacon's notions that factual and scientific claims should always be contestable. Also strongly associated with this topic is an essay on the uses of the law, a pedagogical discussion of the purposes of the law, drawing links between the law's objectives and the most important English legal institutions. Consistently, the words associated with this topic cover a broad range of the institutions of the English legal system, words such as 'assize', 'sergeant', 'leet' [court], 'petty' [sessions court], 'jury', 'escheat', 'attaint', etc.

The next three topics reflect Bacon the political animal and the historian, two deeply interwoven activities (see, e.g., Peltonen 1996b). Judging by words and documents, they are not

¹⁰ Sutton's charitable donation included the funds for a preacher, Coke, in his report on the case, referring to "this holy and charitable intention" including funds to "fit and to finish and furnish a Grammar-school...a learned Preacher to teach all in the word of God."

closely connected to either the law or the scientific method. The first of these three centers on foreign affairs, at a time when foreign affairs were always intertwined with religious issues. Thus, two of the documents most associated with this topic concern deliberation on the advisability of wars to propagate religion, in which Bacon advises "That Wars Defensive for Religion...are most just; Though Offensive Wars, for Religion, are seldom to be approved". The arguments used are not mainly legal, but rather moral and strategic. Consistently, the distinctive FREX words center on military, foreign, and religious terms, justifying our choice of the name Religion & Diplomacy for this topic.

Bacon wrote in the *Advancement of Learning* that "we are much beholden to Machiavelli and others, that write what men do, and not what they ought to do". This is the spirit that he imbues in the topic we call Political Strategy. It is about specific aspects of civic engagement, which Bacon was very interested in exploring from a philosophical standpoint (see, e.g., Peltonen 1996a: 7): the games that politicians play, the use of dissimulation and cunning, the role of virtue and malum, and the advice of Cicero, Tacitus, and Machiavelli, all of whom greatly influenced Bacon's views about civic life (see Benjamin 1965, Peltonen 1996b: 295-296). These words and names are among those associated with this topic. The documents highly associated with this topic are Bacon's essays, a set of works that provide more general ruminations rather than focusing on specific events or decisions. Thus one of the essays highly associated with this topic analyzes "three degrees of this hiding and veiling of a man's self: secrecy, when a man leaveth himself without observation what he is; dissimulation, when a man lets fall signs and arguments, that he is not, that he is; and simulation, when a man industriously and expressly feigns and pretends to be, that he is not."¹¹

Bacon was a historian (Tinkler 1996), a fact that is reflected in the topic we call Dynastic Politics. The documents most associated with this topic are Bacon's discussions of the reigns of the three great Tudor monarchs. Most important was the history of the reign of Henry VII, which naturally was concerned with the dynastic rivalries that led to a century of ongoing war—hence FREX words such as 'York' and 'Plantagenet'—and that continued into Henry's reign with later challenges by pretenders to the throne—hence 'Perkin'. Similarly, within the words associated with the topic, the mention of Maximilian and Ferdinando suggests broader European dynastic rivalries and alliances. Many references to nobility indicate similar concerns. But, these are not mere histories: it seems that Bacon's writing always had an ulterior purpose. Thus his history of the reign of Henry VII was dedicated to the young Prince Charles that he might learn about his ancestor "...a Wise Man, and an Excellent King...I have not flattered him, but took him to life as well as I could, sitting so far off, and having no better light...it is not amisse for You also to see one of these Ancient Pieces". If history is any judge, Charles did not read carefully enough.

Our name for the next topic, Classical Thought, needs little justification. Over one half of the FREX words are transparent references to Greek or Roman places, gods, people, or

¹¹ Edited by the authors to convey the notions more clearly to a modern audience.

mythological creatures. Notably, law is absent within this topic, although Bacon wrote much about government and law in Greek and Roman times. But those writings are associated with other topics. The documents most highly associated with this topic are sections of *The Wisdom of the Ancients*. These are a collection of short ancient tales but as always Bacon saw more than myth in them for "these Fables contain certain hidden and involved meanings".

Five topics remain, all related to Bacon's writings on the philosophical study of nature and the physical universe. The names of these topics all are self-evidently justifiable from both the highest probability and FREX words for each topic. We use Physics, Energy; Physics, Air and Sound; Botany; Life & Death; and Pharmacology.

The topic names we have chosen and the details in Table 3 illustrate the tremendous breadth of Bacon's opus, a subject much emphasized by contemporary scholars (see, e.g., Peltonen 1992). Figure 1 shows the relative importance of the topics in the corpus we assembled. For each topic, we display the assigned names and the top fifteen FREX words. The size of the bars to the left of each topic is proportional to the probability that a random word drawn from the whole corpus has been generated by that particular topic. The most prevalent topics in the corpus reflect legal and political subjects, a finding that is, on the one hand, a natural reflection of Bacon's legal background and, at the same time, consistent with Bacon's continuous engagement in high politics (see, e.g., Holdsworth 1927). To be sure, the relative frequencies in Figure 1 reflect the mix of documents that we have been able to assemble for the corpus. But the frequencies reflect something much more important than that. They reflect the fact that when one adds a document that nominally seems to be about something highly specific, such as religion, that document can contain themes that are resonant with other aspects of Bacon's work, such as the law or foreign affairs. We pursue this observation much more systematically in the next section, highlighting the associations between Bacon's use of different themes and how such use varied with audiences and form of document.

First, however, it is productive to summarize some of the overarching lessons derived from the exercise of topic labeling. After all, as the name suggests, one of the principal contributions of STM is to identify the topics that are easily identifiable in an author's corpus. The analysis above therefore gives a machine-learning statistical answer to the fundamental and still unsettled question 'What is in Bacon's work?' (see, e.g., Peltonen 1996a, Vickers 1992).

Two core elements of the Baconian program are revealed quite clearly in STM's identification of the two topics relating to scientific method and our choice of names for them, derived from STM's output. Probing for Facts corresponds to Bacon's emphasis on an intensive effort to find out about, and catalogue, the world (see, e.g., Malherbe 1996: 79, 83). Epistemology captures Bacon's path-breaking emphasis on the inductive logic of interpreting the world given the ascertained facts (see, e.g., Peltonen 1996a, Rossi 1996). These two topics together reveal the spirit of the Baconian emphasis on the appropriate research procedures and the scientific method (see, e.g., Jonas 1984, Sagasti 2000, Malherbe 1996).

We have emphasized so far those themes that Bacon's corpus contains. But the other side of this coin are the themes that are absent. Given its immense scope and diversity, it is hardly surprising that Bacon's work, including his scientific program, has been subject to many interpretations, attributing to Bacon many different ideas. For example, while recent scholarship argues that Bacon's emphasis on the methodology of science was driven primarily by quest for truth per se, and was thus distinctly non-utilitarian (Rossi 1996: 35-36), some view the Baconian program as setting the foundation for the later (18th-century) production of useful knowledge (e.g., Mokyr 2005, 2010). Others have further viewed Bacon as a visionary in the institutional and organizational arrangements necessary for implementation of a scientific program (see, e.g., Jonas 1984, Sargent 1996). Interestingly, our STM estimates do not reveal any emphasis in Bacon on these two latter themes, the utilitarian emphasis on useful knowledge and the elaborate organization of the scientific quest. The results of STM suggest that the emphasis on these might have arisen at least as much via later interpretation as through Bacon's emphases.

Furthermore, Bacon has hardly been studied as a legal scholar (see, e.g., Stone de Montpensier 1968, de Montmorency 1905: 263; Holdsworth 1927: 10; Shapiro 1980: 333; Vickers 1992, Peltonen 1996c). Yet the set of topics identified by our STM unmistakably reveal that Bacon was, above all, a lawyer.¹² Five of the sixteen topics are directly on law. Law, in addition, contributes significantly to the documents associated with Probing for Facts, and is present in Religion & Diplomacy. The prevalence of legal themes among the STM-identified topics in our corpus raises the question of the potential connection between Bacon's jurisprudential thought and his scientific method. In the following section, we turn to examining this question explicitly, drawing on STM-based quantitative evidence.

5. Evidence on the Origin and Features of the Baconian Program

5.1. Topic Correlations

Aside from providing the only framework that currently allows for inclusion of metadata covariates, one important advantage of STM over LDA is that STM allows topics to be correlated across documents.¹³ Upon estimating our 16-topic STM, we use this feature of STM to examine the interconnectedness at the document level of the various themes in Bacon's opus. Figure 2(a) provides a visualization of the links between topics, where those links capture how the importance of two topics covaries across documents (reflecting the estimate of $\theta_d'\theta_d$, which in turn reflects the metadata in Γ and the estimates of \mathbf{x}_d and Σ). The links summarize positive

¹² One might be tempted to conclude that this is simply a result of the selection of documents into the corpus. That is unlikely for two reasons. First, the legal aspect of Bacon's thought has received much less attention over time than the elements of his scientific program and therefore the works that have been made machine-readable by scholars are unlikely to overly emphasize law. Second, given our metadata, STM's routines produce a set of topics that is not conditional on the balance of different types of documents that is captured by the metadata.

¹³ Topic correlation is allowed also in CTM (Blei and Lafferty 2007).

correlations, indicating that when one linked topic is important for a document the other linked topic is more likely to be important for that document.¹⁴

A key empirical fact about Bacon's opus, evident from Figure 2(a), is that Epistemology, the central theme of the Baconian program, is directly correlated, on the one hand, with Physics, Energy and, on the other hand, with Religious Law, as well as indirectly linked to Law & Nation and Religion & Diplomacy. The link between Epistemology and Religious Law is slightly stronger than that between Epistemology and Physics, Energy. Bacon's work therefore combined epistemological reasoning with discussion of both natural philosophy and legal themes. Given Bacon's early training in law and a later interest in natural philosophy, this is one piece of evidence consistent with the hypothesis that the Baconian program emerged from Bacon's immersion in the epistemology of the common law. William Harvey's disparagement of Bacon—that he wrote philosophy like a Lord Chancellor (Clark 1898: 299)—was perhaps a more acute observation than Harvey had intended.

Next, we examine the topics that are inversely correlated across documents. Negative topic correlations are not typically highlighted in the topic modeling literature; however, as we demonstrate below, identifying negatively correlated topics provides further valuable insights about the features of Bacon's opus. Figure 2(b) shows the connections between topics arising from the set of correlations that are the most negative.¹⁵ The central feature of this diagram is the inverse correlation of King, Court & Law with five other topics. The negative correlation between King, Court & Law and Epistemology, for example, demonstrates that when Bacon was discussing matters of law relating to the monarchy and those close to the monarchy he was less likely to use his epistemology than elsewhere. As indicated in the previous section, the topic King, Court & Law reflects Bacon's experience of the law that was more personal for him, arising out of his positions serving the crown. Apparently, when Bacon was applying legal reasoning to issues arising in the most intimate circles around the monarch, he was not willing to use the types of methodological arguments that he advocated (Epistemology, Political Strategy) and used himself in the context of other legal domains (Land Law, Legal Theory). This is evidence in favor of the thesis that Bacon was more than willing to pursue his own personal and political goals, often by whatever means (see, e.g., Drinker Bowen 1993, Powell 1996).¹⁶

We note also that several topics are not connected, in either Figure 2(a) or 2(b). This finding is relevant to one disputed element of the scholarship on Bacon as captured by Peltonen

¹⁴ See Roberts et al. (2016b: 23) for more discussion of topic correlation. The links that appear in our diagram are those for which the correlation is positive. Note that if all data were random the generative model implies that the correlation between topics would be -0.0667. There are 120 topic correlations, implying that our figure captures the 6% of correlations that are the strongest positive ones.

¹⁵ The criterion for inclusion of links is analogous to that in Figure 2 in that we include all links for which the estimated correlation is more than 0.0667 below the expected value under the null hypothesis of random data of -0.0667. Figure 3 captures the 7% of correlations that are the strongest negative ones.

¹⁶ Figure 2(b) also confirms our previous conclusion about the relationship between Bacon's epistemology and law in general, under the assumption that the correlation relationships are transitive. Figure 2(b) shows negative correlations and therefore the connection from Epistemology to Land Law (or Legal Theory) via King, Court & Law indicates an indirect positive correlation. However, the assumption of transitivity of correlations is a strong one, which does not hold in general.

(1992: 279): "One of the central issues in recent scholarly work on Francis Bacon has been the unity of his different writings. It has often been insisted that Bacon's separate works are governed by a high degree of internal consistency and that they form a single great project." We certainly have found connections between some topics, but there are sufficient topics that are isolated, those on history (Dynastic Politics) and Classical Thought, for example, that we hardly find a single great project. Thus, we endorse the conclusion of Peltonen (1992: 281) that "Bacon's different writings did not form a consistent and coherent system."

5.2. Word Usage Across Topic Pairs

We next examine the use of common vocabulary in selected pairs of topics. The degree of overlap in vocabulary use for two topics is indicative of the degree to which distinct themes in Bacon's opus share a common semantic and methodological foundation. To examine this, in Figures 3-6, we present a series of plots of the top 80 FREX words within chosen topic pairs.¹⁷ We are especially interested in exploring how vocabulary use in the two core methodological aspects of the Baconian program, Epistemology and Probing for Facts, is common with that in use in legal and scientific topics. In order to keep the number of figures to a minimum, for each of law and science, we choose one topic that is directly connected to Epistemology in the topic correlation diagram (Figure 2) and one topic that is not connected. We use Religious Law and Legal Theory for law and Physics, Energy, and Botany for science.

In each of the plots, the size of any displayed word is proportional to the word's use within the combined topics. The position of a word along the horizontal axis measures the difference in the probabilities that associate a word with each topic normalized by the maximum difference in probabilities between the two topics for all 80 words (Roberts et al. 2016b, fn. 20).¹⁸ The position of each word along the horizontal axis is thus indicative of how common a word is in one topic versus the other topic (given that the words are among the 80 used most frequently by the combination of topics). Words located at, or close to, the vertical dashed line are the ones that are shared equally by both topics and therefore are very important in the analysis. They provide the vocabulary that is in common for the two topics.

Analyzing Figure 3, we focus on words characteristic of methodology such as 'reason', 'true', 'find', 'cause', 'therefore', 'yet', 'understand', and 'matter', all indicative of discussion that refers to justifying arguments and examining relationships between facts.¹⁹ These are words that

¹⁷ To relate these plots to our formal model in Section 2, had we chosen to use the 80 highest probability words (instead of FREX words), when relating topics k and j , we would have chosen the 80 words corresponding to the 80 v , for which estimates of the following was greatest: $\sum_d N_d(\beta_{dkv} + \beta_{djv})$. Use of FREX implies using estimates of parameters that we have not included in our formal model description.

¹⁸ The vertical configuration of the words is random.

¹⁹ To make judgments on word usage in Bacon's time, we used the Oxford English Dictionary (OED). For example the OED defines one use of 'yet' "as conjunctive adv. or conj. introducing an additional fact or circumstance which is adverse to, or the contrary of what would naturally be expected from, that just mentioned". This is a frequent usage in the law reports of that time, for example, "[T]he court answered, that all the justices of peace in England did so, and therefore, though they have not authority to do it in strictness of law, yet *communis error facit jus*" (1 Lord Raymond 42; English Reports 91: 925). The OED gives the following usage "1596, Spenser *Prothalamion* 117: As he would speake, but that he lackt a tong yeat did by signes his glad affection show." Our Python programs have yet as the modern spelling of yeat.

are central to inductive reasoning. For both Physics, Energy and Religious Law these words are positioned near the center line, meaning that both topics share these concepts equally with Epistemology. These concepts are at least as important in the overlap with Epistemology for Religious Law as for Physics, Energy, perhaps even more so. They are also at least as important (that is, used as much) in the combination of Epistemology with Religious Law as with its combination with Physics, Energy. But given the imprecision of the evidence, we do not venture the stronger claim. It is sufficient for our argument that the connection of Epistemology to Physics, Energy is obviously no stronger than the connection of Epistemology to Religious Law, an observation that endorses the conclusion reached when examining topic correlations. Exactly the same argument could be made for Figure 4, which examines word sharing between Epistemology and Legal Theory and Botany, respectively.

Words not shared by the contrasting topics are also important. Stems of 'knowledge' and 'philosophy' are no more shared by Physics, Energy (or Botany) and Epistemology than they are by Religious Law (or Legal Theory) and Epistemology. Indeed, the two parts of Figure 3 show very similar overall patterns, with 'god', 'law', and 'time' playing the same distinctive role for Religious Law that 'motion', 'bodi', and 'heat' play for Physics, Energy. The same general image is also present in Figure 4, this time with 'land' playing the same role for Legal Theory that 'tree' plays for Botany.

Given that Bacon was first and foremost a jurist, and not a scientist, this evidence on word usage points to the legal origins of Bacon's inductive reasoning. Bacon carried ideas absorbed from legal processes into the epistemology of science. This is very clearly captured by statements in two documents highly associated with two of our topics. The document that is fourth most associated with Land Law is the *Reading on the Statute of Uses*, which was prepared in 1599 for an audience of lawyers. It states that "The nature of a use is best discerned by considering what it is not, and then what it is, for it is the nature of all humane Science, and knowledge to proceed most safely by Negative and exclusive." This is Bacon's approach to induction that focuses on ruling out instances by negative example. It is virtually repeated for scientists in the *Novum Organum* in 1620, in a section that is the fifth most highly associated with our Epistemology topic: "...it is the peculiar and perpetual error of the human understanding to be more moved and excited by affirmatives than negatives, whereas it ought duly and regularly to be impartial; nay, in establishing any true axiom the negative instance is the most powerful."

Thus the first great element of the Baconian program has its genesis in the law. This is a conclusion that is almost absent from the vast interpretative literature on Bacon. Over half a century ago, Kocher (1957: 3) remarked that "The world sometimes forgets that Francis Bacon was a great jurist as well as a great philosopher of natural science" before then arguing that "...of prime importance to his total philosophy, he believed his inductive method valid in the one field as in the other". The results of our STM suggest that, at least for the field of law, Kocher's largely ignored argument should have been given more heed.

Figures 5 and 6 carry out the same exercise for the other great element of the Baconian program, Probing for Facts. Commonalities between word usage are few between the pairs of topics. The distinctiveness of shared words is less. 'Ask' and 'answer', the two words emblematic of Probing for Facts, appear in similar places on both figures, among words not shared to any significant degree with the legal or scientific topics.

One must conclude that these word-usage figures tell us little about the relationship between Probing for Facts and other Baconian topics. The word usage diagrams suggest Probing for Facts is much more *sui generis* than Epistemology. This conclusion about the origins of the two great elements of the Baconian program is consistent with the evidence in the correlations in Figure 2. But even though their origins might lie in different places, there is evidence of commonality between Probing for Facts and Epistemology, as indicated in Figure 7. The words 'cause', 'understand', 'reason', 'know', 'therefore', 'find', 'yet', 'ask', 'mean', 'think', 'matter', 'observe', 'like', and 'method' are all shared to some degree, but by no means evenly between the two. The two words ('nature' and 'say') that are distinct to each topic provide information about what is at the center of each topic: Epistemology is about nature and Probing for Facts is about dialogue. But 'man' stands at the center, shared by both processes, indeed at the center of the Baconian program.

The relationship between the ideas captured by Probing for Facts and Epistemology has been controversial in the literature. "The fact that so many modern historians of science have accepted the image of Bacon as a fact-collector who separated that process from an unworkable system of induction merely shows, I think, that received ideas constantly reproduce themselves if unchecked by recourse to the text or challenged by independent thought. Bacon insisted time and again on a reciprocal, symbiotic movement from 'works and experiments' to 'causes and axioms' and back again" Vickers (1992: 510). Indeed, we have checked this by recourse to the text using statistical methods and we find that there is something to say for Vickers' views. Figure 7 shows some connection, though not a strong one, between Probing for Facts and Epistemology.

5.3. *Effect of Metadata Covariates*

We now make direct use of the defining feature of STM—the inclusion of metadata covariates into topic model estimation—to analyze the effect of our two sets of covariates on topical prevalence. Specifically, after obtaining the parameter estimates for our 16-topic STM, we estimate the relationship between the proportion of each document about a topic (reflecting estimates of the θ_d in (1) above) and the values of a particular covariate for all documents (see Roberts et al. 2016b: 17-18), thereby obtaining estimates of the pertinent column of Γ in (1) above. We then plot either mean estimated topic proportions for different document-level covariate values (Figures 8-10, 12, and 18) or mean differences in estimated topic proportions for two different values of a document-level metadata covariate (Figures 11, 13-17).

We first examine the role of the intended audience. Peltonen (1992), for example, emphasizes that an understanding of Bacon's opus necessitates taking into account the fact that

Bacon, the master rhetorician, engaged with different audiences in distinct, even seemingly contradictory, ways. We explore how the prevalence of the two central topics associated with the Baconian program, Probing for Facts and Epistemology, varies with the intended audience. Figures 8 and 9 summarize the results.

Figures 8 and 9 show whom Bacon was targeting with his writing on the two separate aspects of his methodological program. Bacon's emphasis on questioning the world to establish facts (Probing for Facts) was addressed primarily to philosophers. This is consistent with his aim of challenging the standard Aristotelian paradigm of his day, which he viewed as jumping much too quickly to general principles and not focused enough on deriving intermediate generalizations from existing facts. Bacon emphasized inductive reasoning primarily when addressing methodologists. Thus, his efforts to communicate his Epistemology were primarily directed toward those social actors who would find them most useful. Interestingly, none of the core Baconian program seems to be targeted at either historians or politicians. It is difficult to derive from Figures 8 and 9 the oft-assumed conclusion that Bacon believed that his method was universally applicable in all domains of thought.

Was the heart of the Baconian program relatively favored in some writing forms over others? Kocher (1957: 6), for example, argues that Bacon's apothegms (i.e. concise sayings, or maxims), many of which are expressed in *The Elements of the Common Laws of England*, are "generalizations or axioms of that middle order which Bacon throughout his philosophical writings describes as most fruitful for works". In particular, "[i]n natural science the utility of the middle axiom is to state a rule applicable to new physical situations".

Figure 10 shows that when articulating his convictions about the necessity to question and investigate the world, Bacon preferred apothegms to other writing forms. Consistent with Kocher (1957), Figure 11 shows that Bacon relies on apothegms disproportionately more also when he expresses some of his legal ideas (Legal Theory; Land Law). Bacon, however, does not view apothegms as the form to convey anecdotes for politicians (despite the many political vignettes that appear in the *Apothegms*), reflections of classical thought (despite the many Greek and Roman vignettes that appear in the *Apothegms*), or, importantly, his epistemological beliefs.²⁰ Indeed, as Figure 12 illustrates, Bacon strove to develop his deep epistemological ideas primarily in book-length volumes and, to some extent, essays. This finding resonates with Peltonen's (1996a: 10) interpretation that Bacon, in order to reach new audiences, deliberately chose to disseminate his groundbreaking views about the genesis of knowledge in a broadly-available printed form—despite the fact that to publish, as opposed to circulate a manuscript privately, had at the time been viewed as "degrading" from the perspective of his aristocratic status.

What writing forms did Bacon favor for communication of other ideas and beliefs about society, and what audiences did he target with particular topics? Our STM analysis offers several

²⁰ We distinguish here between Bacon's apothegms, short pithy disconnected statements, and his use of aphorisms in the longer works *Novum Organum* and *The Advancement of Learning*. In those works, the aphorisms are connected with each other, and constitute sequential steps in arguments, rather than disconnected statements.

additional insights about Bacon's opus, the content of which has been much debated by modern-day scholars of his work (see, e.g., Paterson 1987, Vickers 1992; Peltonen 1992, 1996a). First, Peltonen (1996b: 294), for example, argues that "scholars have tended to see Bacon's *Essays* as an account of his civil knowledge. Of late, however, the best scholarship has questioned this conclusion". Our STM analysis provides quantitative evidence that sheds light on this ongoing debate. We find that Bacon indeed used essays (Figure 13) primarily to communicate his theories of Political Strategy, a particular facet of civil knowledge in which he was passionately interested (see Section 4). However, Bacon also relied on essays to express his views about foreign policy and religion (Religion & Diplomacy).

Second, consistent with our observations on Figures 8 and 9, when Bacon addressed historians (Figure 14), he did not emphasize the methodological core of his program. Instead, he focused on the history of England's dynasties (Dynastic Politics) and classical stories (Classical Thought).

Third, in Bacon's communications with politicians (Figure 15),²¹ his methodological ideas took a back seat, suggesting that his supposed unstinting advocacy of his great program did have limits. When he addressed politicians, Bacon emphasized religion, in connection with both law and diplomacy. Unsurprisingly, given his direct involvement in political intrigues, when communicating with politicians, Bacon also chose to express his thoughts about high politics and associated legal matters (King, Court & Law). Surprisingly, his deep insights about political tactics and the games politicians play (Political Strategy) were not targeted at all to politicians, but rather to philosophers (Figures 16, 18). Bacon's analysis of politics seems to have been more congruent with that of a modern-day rational-choice political-economy theorist than a practicing politician. Note that this communication strategy could not have arisen simply from a desire to hide his willingness to endorse political chicanery, since, in stark contrast to his communication of other themes, much of Bacon's communication with politicians came in the form of private communications, in letters (Figure 17). For Bacon apparently, teaching political strategy to politicians was hardly any more important than teaching it to scientists (Figure 18).

6. Conclusion

We have examined Bacon's works using the new methodological tool of structural topic modeling, relying almost wholly on machine-learning to uncover and to analyze the themes that are present in those works. Yet, for one of our most important conclusions we can do no better than turn to remarks made nearly a century ago by the venerable Holdsworth (1927): "Of all Bacon's claims to greatness, his claims as a lawyer are the least of all known. It is strange that this should be so. The law was his profession....I think that those who study his philosophy should remember that Bacon was trained as a student and practitioner in a system of case law, which taught him to construct the rules and principles of law from concrete cases. It is at least arguable that this had some influence on his inductive system of experimental philosophy."

²¹ In our coding of the metadata, monarchs are politicians.

Whereas this conjecture has lain dormant and largely ignored in the massive critical literature on Bacon produced since Holdsworth made it, STM raises it to the fore.

Thus the results of our paper lead to the broader question of the degree to which sets of ideas attributed to individuals are really a product of those individuals rather than of the prevailing culture inculcated in them. Interestingly Bacon had a clear view on this, and applied it to himself, viewing his insights "the offspring of time rather than wit", a remark that itself has clear origins in the way that 17th-century common-law lawyers thought about the development of law (Grajzl and Murrell 2016).²² However, our conclusions cannot stretch quite as far as this. In the patterns detected by STM, the Probing for Facts topic is less strongly related to the other topics than is Epistemology, the one implied in Holdsworth's remarks. Although there is evidence that the emphasis on fact-finding grew out of the practice of law and spread much more generally (Shapiro 2000), there are only hints of this in our results, rather than compelling conclusions. Thus, we cannot go as far as Keynes (1967) in concluding that Bacon "...was the human mouth-piece of that impalpable thing, the *Zeitgeist*, a presence which would have made itself felt even if Bacon had never been born".

Importantly, the techniques used here not only identify what a corpus contains, but also identify what is not in it, or at least what is not prominent. STM offers the possibility of checking whether themes and ideas attributed to a particular individual by a later interpretative literature can be readily identified in a statistical analysis of the person's works. We have found that Bacon did not place any emphasis on the utilitarian argument that the development of science is important because of the promise that technological knowledge holds for improving the human condition. To be sure, one can find statements in Bacon that hint at that point. But one can also find statements with the opposite tone. In Book 1 of *The Advancement of Learning*, Bacon remarked that "...the greatest error of all is mistaking the ultimate end of knowledge...the end ought to be...to separate and reject vain and empty speculations, and preserve and increase all that is solid and fruitful."²³ Our inability to find a utilitarian theme in Bacon's thought using machine-learning techniques is therefore evidence in support of those who have emphasized that truth provided its own utility for Bacon, and that there was no ultimate goal of improving life's condition (Rossi 1996: 35-36).

Nevertheless, the promise of science as a means to improve the human condition was a central part of the Baconian program as advanced in the centuries after Bacon (Mokyr 2005, 2010, 2016). But what is in the Baconian program, the intellectual paradigm embraced by Bacon's followers, and what is in Bacon's own work are different questions. Our results indicate that the utilitarian aspect of the Baconian program cannot be characterized as a direct product of

²² Bacon uses this phrase twice in the *Novum Organum*. The two separate sections are the documents that are the second and the fourth most highly associated with the topic Epistemology.

²³ This appears in a section that is the thirteenth most highly associated document in the Epistemology topic and the sixteenth in Religious Law.

Bacon's works, and therefore we hypothesize that it was an emphasis generated by those who were inspired in a more general way by Bacon.

Our application of topic modeling has therefore helped to distinguish between the elements of the Baconian program that were a product of the culture of Bacon's own time, the elements that were probably Bacon's own, and the elements that were a product of cultural developments after Bacon's death. We hope that these findings help to provide some perspective on the source of the culture that has been characterized as so important in the early development of the modern world.

References

- Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Publishers.
- Alesina, Alberto and Paola Giuliano. 2015. "Culture and Institutions." *Journal of Economic Literature*, 53:4, 898-944.
- Algan, Yann and Pierre Cahuc. 2010. "Inherited Trust and Growth." *American Economic Review*, 100:5, 2060-2092.
- Bandiera, Oriana, Stephen Hansen, Andrea Prat, and Raffaella Sadun. 2016. "CEO Behavior and Firm Performance." Unpublished manuscript.
- Barro, Robert J. and Rachel M. McCleary. 2003. "Religion and Economic Growth Across Countries." *American Sociological Review*, 68:5, 760-781.
- Becker, Sascha, Katrin Boeckh, Christa Hainz, and Ludger Woessmann. 2016. "The Empire Is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy." *Economic Journal*, 126:590, 40-74.
- Bellstam, Gustaf, Sanjai Bhagat, and J. Anthony Cookson. 2016. "A Text-Based Analysis of Corporate Innovation." SSRN Working Paper No. 2803232.
- Benjamin, Edwin B. 1965. "Bacon and Tacitus." *Classical Philology*, 60:2, 102-110.
- Besley, Timothy, Jose G. Montalvo, and Marta Reynal-Querol. 2011. "Do Educated Leaders Matter?" *Economic Journal*, 121:554, F205-227.
- Blei, David M. 2012. "Probabilistic Topic Models." *Communications of the ACM*, 55:4, 77-84.
- Blei, David M. and John D. Lafferty. 2007. "A Correlated Topic Model of Science." *The Annals of Applied Statistics*, 1:1, 17-35.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research*, 3, 993-1022.
- Bowles, Samuel and Herbert Gintis. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, NJ: Princeton University Press.
- Brender, Adi and Allan Drazen. 2013. "Elections, Leaders, and the Composition of Government Spending." *Journal of Public Economics*, 97:C, 18-31.
- Clark, Andrew. 1898. *'Brief lives', Chiefly of Contemporaries, Set Down by John Aubrey, between the Years 1669 & 1696*. Oxford, UK: At the Clarendon Press.
- Coquillette, Daniel R. 1992. *Francis Bacon*. Stanford, CA: Stanford University Press.
- Craig, Tobin L. 2010. "On the Significance of the Literary Character of Francis Bacon's *New Atlantis* for an Understanding of His Political Thought." *The Review of Politics*, 72, 213-239.
- de Montmorency, J.E.G. 1905. "Francis Bacon." *Journal of the Society of Comparative Legislation*, 6:2, 263-283.
- Drinker Bowen, Catherine. 1993. *Francis Bacon, The Temper of a Man*. New York, NY: Fordham University Press.

- Fligstein, Neil, Jonah Stuart Brundage, and Michael Schultz. 2014. "Why the Federal Reserve Failed to See the Financial Crisis of 2008: The Role of 'Macroeconomics' as a Sense Making and Cultural Frame." Institute for Research on Labor and Employment Working Paper No. 111-14.
- Grajzl, Peter and Peter Murrell. 2016. "A Darwinian Theory of Institutional Development Two Centuries Before Darwin." *Journal of Economic Behavior and Organization*, 131(A) November: 346-372.
- Guiso, Luigi, Paola Sapienza and Luigi Zingales. 2006. "Does Culture Affect Economic Outcomes?" *Journal of Economic Perspectives*, 20:2, 23-48.
- Guiso, Luigi, Paola Sapienza, Paola, and Luigi Zingales. 2016. "Long Term Persistence." *Journal of the European Economic Association*, 14:6, 1401-1436.
- Hansen, Stephen and Michael McMahon. 2016. "Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication." *Journal of International Economics*, 99, S114–S133.
- Hansen, Stephen, Michael McMahon, and Andrea Prat. 2015. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." Unpublished manuscript.
- Hathi Trust. n.d. <https://www.hathitrust.org>. Accessed January-May 2016.
- Holdsworth, W. S. 1927. "Bacon." *The Cambridge Law Journal*, 3:1, 10-12.
- Huang, Allen, Reuven Lehavy, Amy Zang, and Rong Zheng. 2016. "Analyst Information Discovery and Interpretation Roles: A Topic Modeling Approach." University of Michigan Ross School of Business Working Paper No. 1229.
- Internet Archive. n.d. <https://archive.org>. Accessed January-May 2016.
- Jonas, Hans. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago, IL: University of Chicago Press.
- Jones, Benjamin F. and Benjamin A. Olken. 2005. "Do Leaders Matter? National Leadership and Growth Since World War II." *Quarterly Journal of Economics*, 120:3, 835-864.
- Keynes, Geoffrey. 1967. "The Wilkins Lecture, 1967: Bacon, Harvey and the Originators of the Royal Society." *Proceedings of the Royal Society of London. Series B, Biological Sciences* Vol. 169, No. 1014 (Dec. 19, 1967), pp. 1-16.
- Kocher, Paul H. 1957. "Francis Bacon on the Science of Jurisprudence." *Journal of the History of Ideas*, 18:1, 3-26.
- Kusukawa, Sachiko. 1996. "Bacon's Classification of Knowledge." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 47-74.
- Larsen, Vegard H. and Leif Anders Thorsrud. 2015. "The Value .of News." Unpublished manuscript.
- Mahlerbe, Michel. 1996. "Bacon's Method of Science." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 75-98.

- Martin, Julian. 1992. *Francis Bacon, the State, and the Reform of Natural Philosophy*. Cambridge, UK: Cambridge University Press.
- Matthews, Steven. 2008. *Theology and Science in the Thought of Francis Bacon*. Burlington, VT: Ashgate.
- Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. "Optimizing Semantic Coherence in Topic Models" *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*: 262-272.
- Moby Word Lists. 2002. <http://www.gutenberg.org/ebooks/3201>. Accessed April 28, 2016.
- Mokyr, Joel. 2016. *A Culture of Growth: The Origins of the Modern Economy*. Princeton, NJ: Princeton University Press.
- Mokyr, Joel. 2010. "The Contribution of Economic History To The Study of Innovation and Technical Change: 1750-1914." In: Bronwyn H. Halland and Nathan Rosenberg (Eds.), *Handbook of the Economics of Innovation, Vol. 1*, Burlington, MA: Academic Press, 11-50.
- Mokyr, Joel. 2006. "The Great Synergy: the European Enlightenment as a Factor in Modern Economic Growth." In: Dolfma, Wilfred and Luc Soete (Eds.), *Understanding the Dynamics of a Knowledge Economy*. Cheltenham, UK: Edward Elgar, Chapter 1.
- Mokyr, Joel. 2005. "The Intellectual Origins of Modern Economic Growth." *Journal of Economic History*, 65:2, 285-381.
- Morph Adorner. 2013. <http://morphadorner.northwestern.edu>. Accessed April 26, 2016.
- Murrell, Peter. 2016. "Design and Evolution in Institutional Development: The Insignificance of the English Bill of Rights." *Journal of Comparative Economics*, forthcoming.
- North, Douglass and Barry Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *Journal of Economic History*, 49:4, 803-832.
- Nunn, Nathan. 2008. "The Long-term Effects of Africa's Slave Trades." *Quarterly Journal of Economics*, 123:1, 139-176.
- Nunn, Nathan. 2012. "Culture and the Historical Process." *Economic History of Developing Regions*, 27:S1, 108-126.
- Paterson, Timothy H. 1987. "On the Role of Christianity in the Political Philosophy of Francis Bacon." *Polity*, 19:3, 419-442.
- Peltonen, Markku. 1992. "Politics and Science: Francis Bacon and the True Greatness of States." *The Historical Journal*, 35:2, 279-305.
- Peltonen, Markku. 1996a. "Introduction." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 1-24.
- Peltonen, Markku. 1996b. "Bacon's Political Philosophy." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 283-310.
- Peltonen, Markku (ed.). 1996c. *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press.

- Peltonen, Markku. 2004. "Bacon; Francis, Viscount St Alban (1561-1626)." *Oxford Dictionary of National Biography*. Oxford: Oxford University Press.
- Perseus. n.d. <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059>. Accessed May-December 2016.
- Powell, Damian X. 1996. "Why Was Sir Francis Bacon Impeached? The Common Lawyers and the Chancery Revisited:1621." *History*, 81:264, 511-526.
- Project Gutenberg. n.d. <https://www.gutenberg.org>. Accessed January-May 2016.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2016b. "stm: R Package for Structural Topic Models." *Journal of Statistical Software*, forthcoming.
- Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airolti. 2016a. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association*, 111:515, 988-1003.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science*, 58:4, 1064-1082.
- Rossi, Paolo. 1996. "Bacon's Idea of Science." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 25-46.
- Sagasti, Francisco. 2000. "The Twilight of the Baconian Age and the Future of Humanity." *Futures*, 32, 595-602.
- Sargent, Rose-Mary. 1996. "Bacon as an Advocate for Cooperative Scientific Research." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 146-171.
- Schinke, Robyn, Mark Greengrass, Alexander M. Robertson, and Peter Willett. 1996. "A Stemming Algorithm for Latin Text Databases." *Journal of Documentation*, 52, 172-187.
- Shapiro, Barbara J. 1968. "Law and Science in Seventeenth-Century England." *Stanford Law Review*, 21, 727-766.
- Shapiro, Barbara J. 1980. "Sir Francis Bacon and the Mid-Seventeenth Century Movement for Law Reform." *The American Journal of Legal History*, 24:4, 331-362.
- Shapiro, Barbara J. 2000. *A Culture of Fact: England, 1550-1720*. Ithaca NY: Cornell University Press.
- Stone de Montpensier, Roy L. 1968. "Bacon as a Lawyer and a Jurist." *Archives for Philosophy of Law and Social Philosophy*, 54:4, 449-483.
- Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association*, 6:2-3, 255-294.
- Taddy, Matthew A. 2012. "On Estimation and Selection for Topic Models." *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*, 1184-1193.
- The Text Creation Partnership for Early English Books Online. 2014. <http://www.textcreationpartnership.org>. Accessed January-May 2016.

- Tingley, Dustin. 2017. "Rising Power on the Mind." *International Organization*, forthcoming.
- Tinkler, John F. 1996. "Bacon and History." In: Peltonen, Markku (ed.), *The Cambridge Companion to Bacon*. Cambridge, UK: Cambridge University Press, 232-259.
- Vickers, Brian. 1992. "Francis Bacon and the Progress of Knowledge." *Journal of the History of Ideas*, 53:3, 495-518.
- Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno. 2009. "Evaluation Methods for Topic Models." *ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning*, 1105-1112.
- Wheeler, Harvey. 1999. "Francis Bacon's 'Verulamium'. The Common-Law Template of the Modern in English Science and Culture." *Angelaki: Journal of the Theoretical Humanities* 4:1, 7-26.
- Whitaker's Words. 2006. <http://archives.nd.edu/whitaker/wordsoth.htm> and <http://archives.nd.edu/whitaker/dictpage.htm>. Accessed May 15, 2016.

Table 1: Works Included in the Corpus

Name or type of work	No. of text documents (chunks)
<i>New Atlantis</i>	5
<i>Novum Organum</i>	17
<i>The Advancement of Learning</i>	12
<i>The Use of the Law</i>	5
<i>A Collection of Apothegms, New and Old</i>	6
<i>History Natural and Experimental of Life and Death</i>	9
<i>Sylva Sylvarum</i>	10
<i>The History of The Reign of King Henry the Seventh</i>	6
<i>The Natural and Experimental History of Winds</i>	7
<i>The Wisdom of the Ancients</i>	5
<i>The Elements of the Common Laws of England</i>	28
Other writings (various essays, letters, speeches, case reports)	172
Total	282

Notes: Some names edited for rendering in modern English.

Table 2: Text Documents by Writing Form and Audience

Audience	Writing Form						Total
	Apothegm	Case report	Essay	Letter	Speech	Tome	
Historians	0	0	2	0	0	12	14
Lawyers	25	1	12	22	10	4	74
Methodologists	1	0	3	1	0	34	39
Philosophers	6	0	42	2	0	0	50
Politicians	2	1	16	36	3	0	58
Scientists	0	0	20	0	1	26	47
Total	34	2	95	61	14	76	282

Table 3: Topics and Top Words for the Estimated 16-Topic STM

Note that the 'words' listed below are those used by STM after reducing all original text words to their stemmed form. Thus, for example, 'theolog' could reflect an original usage of theology, theologies, theologian, theologians, theologise, theologised, theologises, theologising, etc, etc.

1. Epistemology

Highest Prob: natur, man, will, thing, philosophi, may, histori, knowledg, scienc, part, use, one, experi, can, mind, great, upon, shall, make, therefor, yet, first, matter, particular, understand, caus, much, invent, observ, true

FREX: philosophi, method, defici, logic, system, scienc, axiom, histori, metaphys, idol, discoveri, everyth, aristotl, invent, inquiri, theori, abstract, mathemat, rhetor, notion, mankind, deduc, knowledg, induct, fallaci, theolog, inventor, principl, poesi, contempl

2. Probing for Facts

Highest Prob: say, will, man, one, shall, upon, come, great, answer, lord, make, take, give, king, may, think, day, tell, know, ask, hous, can, sir, see, time, like, use, much, mani, friend

FREX: overburi, impoison, somerset, ask, weston, diogen, galleri, bensalem, tell, poison, answer, pompey, nichola, vespasian, dinner, aristippus, gentleman, injunct, sir, front, chamber, yes, forgiv, tower, philip, friend, chair, ladi, feast, boat

3. Land Law

Highest Prob: use, law, shall, statut, make, feoff, upon, case, will, word, take, land, yet, therefor, may, seiz, estat, heir, first, can, act, reason, say, good, one, time, life, give, right, grant

FREX: feoff, cesti, feme, lesse, remaind, disseise, statut, aver, seiz, clausula, rend, duress, usuri, claus, leas, purchas, advowson, remitt, commonlaw, fee, proviso, tail, parol, demis, atturn, void, dissent, deed, bargain, enrol

4. King, Court & Law

Highest Prob: majesti, will, shall, may, good, lord, upon, time, lordship, king, make, think, self, man, can, know, great, say, take, matter, letter, give, much, well, thing, god, now, mean, one, first

FREX: lordship, plantat, majesti, humbl, undertak, letter, squir, essex, solicitor, attorney, gracious, lord, project, pleas, queen, secretari, coke, chancellor, servic, busi, con, honest, advertis, pray, manuel, antonio, wish, care, protest, beseech

5. Law & Nation

Highest Prob: law, king, shall, may, england, will, natur, majesti, one, subject, case, say, make, time, first, kingdom, man, therefor, can, scotland, person, upon, court, question, sever, point, whether, parliament, take, yet

FREX: scotland, england, duel, allegi, born, union, parliament, kingdom, britain, challeng, alien, conquest, gascoign, subject, question, vote, nation, offenc, royal, law, normandi, post, preced, crown, submiss, prerog, duchi, infer, court, style

6. Religious Law

Highest Prob: god, man, time, learn, law, may, great, make, church, will, shall, upon, good, work, one, say, yet, thing, see, majesti, well, give, first, natur, matter, part, can, use, mani, person

FREX: learn, church, christ, preach, preacher, atheism, benefic, scriptur, dedic, controversi, baptism, liturgi, hospit, endow, alexand, god, theme, holi, reveal, prayer, ecclesiast, bishop, institut, ministri, minist, callisthen, censur, pastor, polici, compil

7. Legal Theory

Highest Prob: land, man, shall, king, good, feloni, may, make, upon, court, treason, call, take, justic, heir, parti, give, one, can, peac, will, law, offic, constabl, everi, yet, life, lord, grant, person

FREX: feloni, constabl, sheriff, counti, debt, executor, forfeit, escheat, juri, tenur, legaci, attaint, treason, petti, assiz, shire, felon, leet, attaind, gaol, owner, chattel, chap, indict, viii, clerk, writ, clergi, outlawri, praemunir

Table 3 continued

8. Religion & Diplomacy

Highest Prob: great, upon, war, will, man, state, make, spain, time, say, shall, yet, may, peopl, part, one, king, take, come, now, england, mani, law, much, like, natur, nation, true, never, princ

FREX: spain, invas, spaniard, speaker, libel, cathol, war, spanish, enterpris, pollio, christendom, turk, sundri, heresi, leagu, palatin, navi, indi, defens, portug, papist, religion, nation, invad, fleet, confeder, armi, worship, lowcountri, nobil

9. Political Strategy

Highest Prob: man, good, will, make, great, say, thing, upon, one, may, natur, can, time, much, shall, mind, like, yet, see, well, mani, come, part, virtu, take, therefor, use, person, first, fortun

FREX: envi, anger, dissimul, malum, felic, bewar, faction, evil, bold, tacitus, reprehens, cun, solomon, lover, secreci, fortun, gradus, discontent, reprehend, simul, virtu, precept, tiberius, friendship, demosthen, quoth, cicero, busi, machiavelli, eam

10. Dynastic Politics

Highest Prob: king, upon, make, great, time, will, shall, come, man, may, part, take, person, one, peopl, also, yet, unto, war, well, princ, england, think, good, lord, much, can, duke, give, peac

FREX: perkin, maximilian, duke, earl, plantagenet, ferdinando, york, french, flander, margaret, ambassador, treati, richard, castill, charl, britain, ladi, rebel, ambassag, king, fillip, henri, lovel, london, john, duchess, bruce, stanley, tower, castl

11. Classical Thought

Highest Prob: thing, man, may, natur, also, shall, say, seem, god, see, will, one, yet, unto, make, matter, can, time, great, like, take, come, mani, first, kind, well, fabl, much, jupit, upon

FREX: prometheus, jupit, proserpina, page, orpheus, miss, pan, cupid, siren, typhon, pentheus, sphinx, parabl, albeit, perseus, moreov, fabl, nemesi, giant, palla, bacchus, cere, muse, allegori, atalanta, monster, hercul, proteus, hell, icarus

12. Physics, Energy

Highest Prob: motion, heat, bodi, natur, instanc, water, will, air, also, must, may, fire, place, can, part, let, one, cold, first, like, flame, yet, substanc, great, spirit, differ, observ, thing, power, appear

FREX: magnet, instanc, expans, similar, class, howev, anim, rapid, latent, liquid, heat, predomin, ignit, exist, ebb, homogen, heterogen, migrat, perpendicular, negat, expand, conspicu, investig, ray, excit, concret, motion, sphere, flame, exclus

13. Physics, Air & Sound

Highest Prob: wind, sound, air, will, make, one, may, blow, thing, great, motion, water, part, bodi, upon, see, like, likewis, also, much, sail, caus, place, sea, man, come, shall, yet, south, two

FREX: wind, string, sound, brass, sail, rain, tone, blow, south, echo, percuss, bell, mast, east, pipe, nois, north, presag, articul, concav, audibl, dram, lute, west, cloud, nurseri, gale, loud, engend, tin

14. Botany

Highest Prob: tree, will, upon, plant, fruit, make, may, earth, water, also, put, come, ground, grow, root, caus, forth, like, great, herb, seed, flower, set, leav, see, kind, bear, one, therefor, much

FREX: sap, plum, cherri, tree, moss, oak, pear, herb, bough, holli, stalk, peach, dung, radish, fig, graft, blossom, cucumb, plant, lettuc, rosemary, mushroom, wheat, vine, escul, colewort, compost, mistleto, turnip, fruit

15. Life & Death

Highest Prob: man, life, year, long, live, age, thing, old, also, hundr, spirit, bodi, may, yet, great, time, death, one, part, much, now, young, creatur, will, last, touch, shall, good, mani, without

FREX: diet, repar, eighti, nineti, canon, liver, aliment, hundr, repair, age, live, prolong, etern, statur, thi, youth, vital, life, desicc, seventi, old, long, short, year, fifti, function, consubstanti, ghost, jesus, consumpt

16. Pharmacology

Highest Prob: bodi, spirit, will, water, part, may, make, upon, see, caus, also, heat, air, thing, great, therefor, like, take, much, use, man, natur, time, motion, cold, put, littl, good, one, come

FREX: liquor, tooth, oil, opiat, broth, explic, amber, stomach, beer, nit, bottl, putrefact, opium, saffron, refriger, sugar, wine, milk, intener, astring, malaciss, purger, indur, smell, sweat, vinegar, infus, drink, candl, spirit

Figure 1: Expected Values for Topic Proportions in the Bacon Corpus

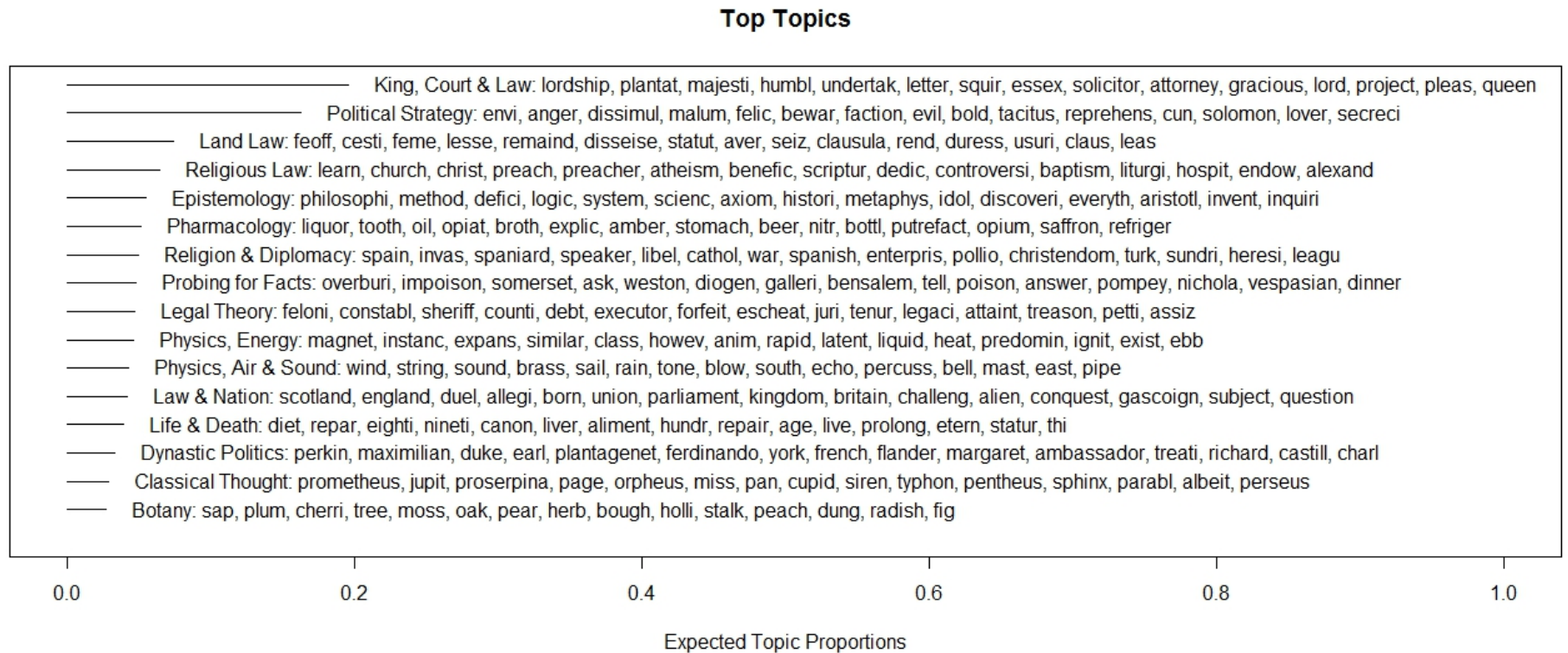
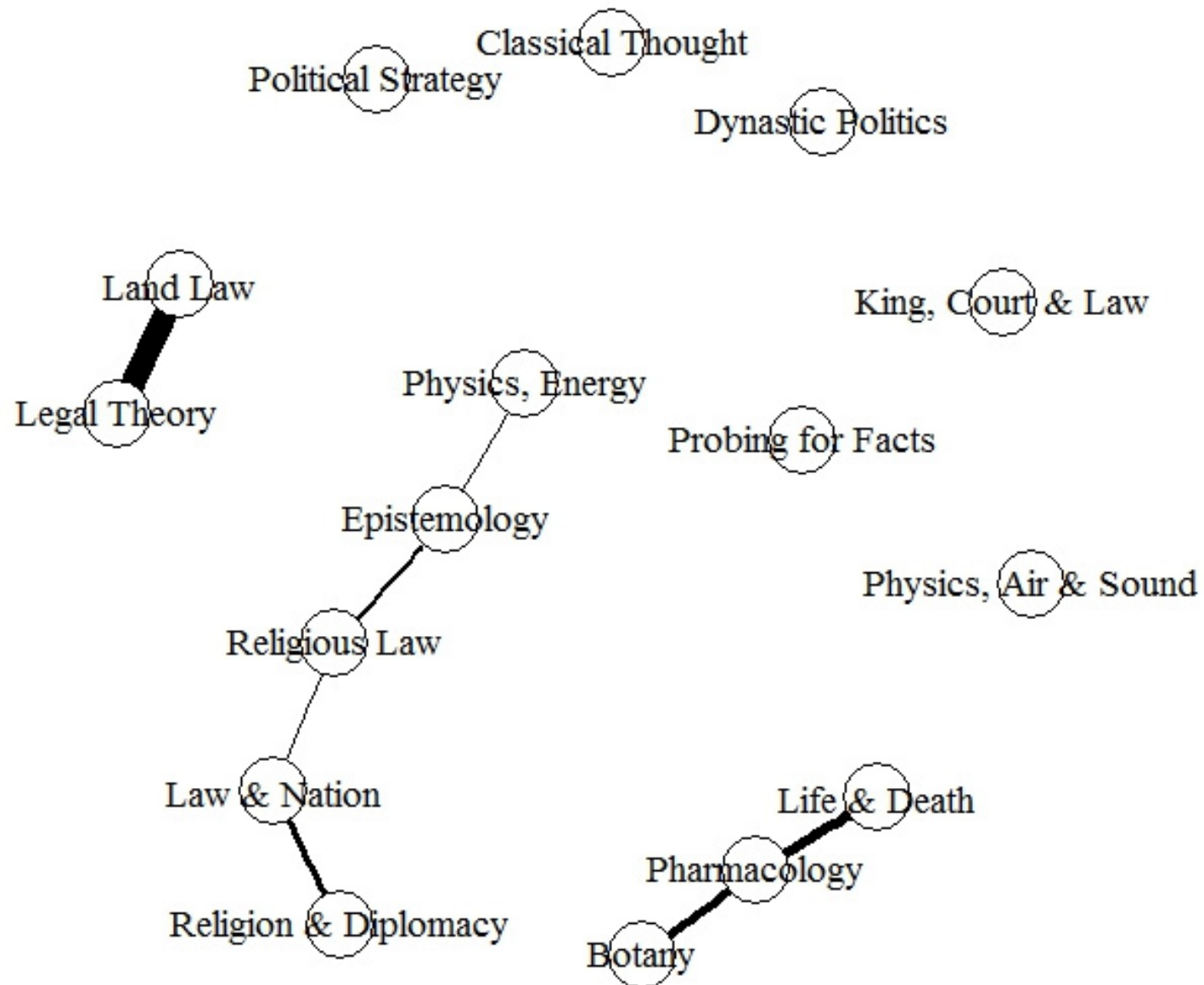
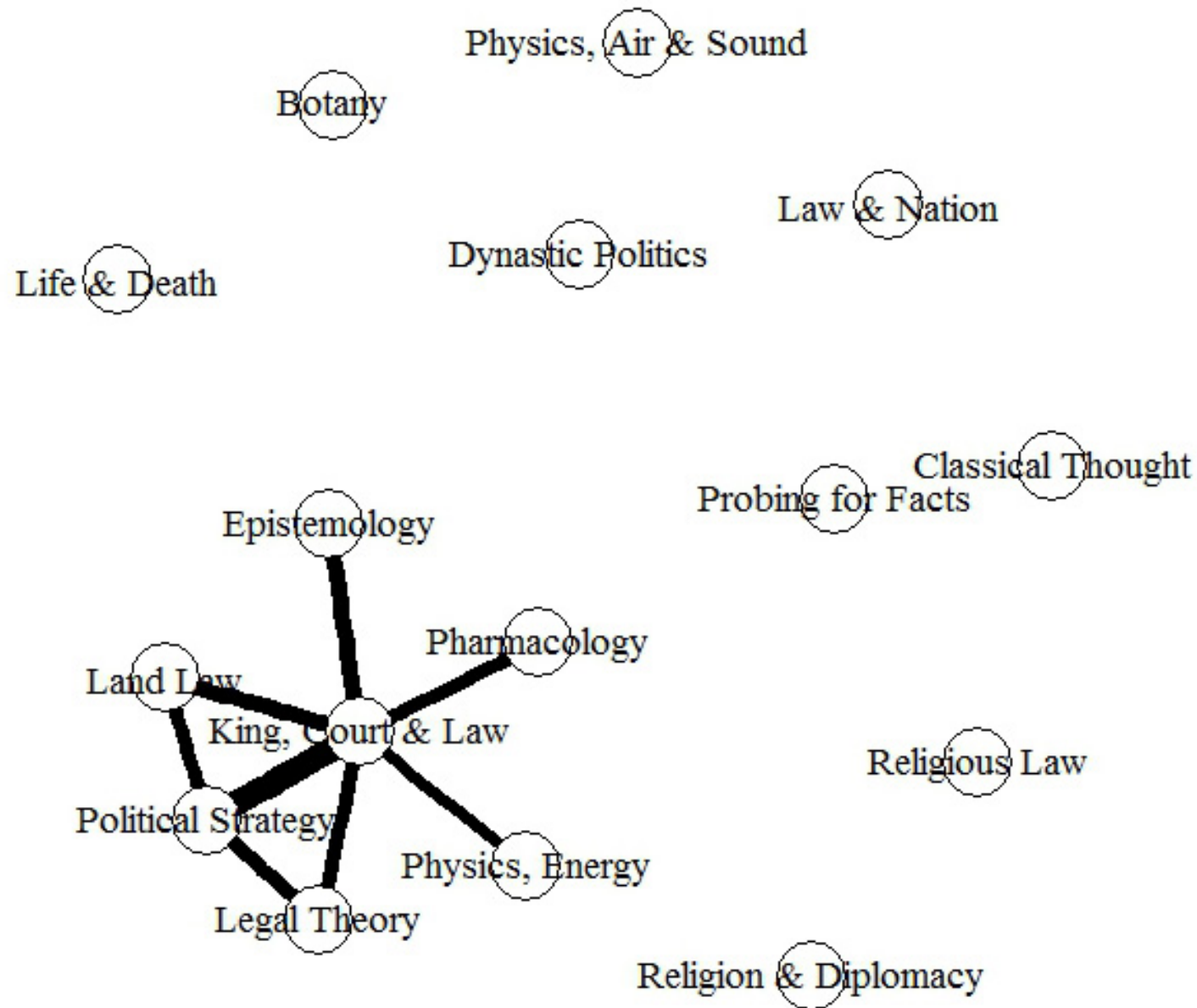


Figure 2(a): Positive Topic Correlations



Note: The relative thickness of links reflects the magnitude of (positive) correlations between respective topic pairs.

Figure 2(b): Negative Topic Correlations



Note: The relative thickness of links reflects the magnitude of (negative) correlations between respective topic pairs.

Figure 3: Vocabulary Use Across Topic Pairs: Epistemology vs. Physics, Energy and Religious Law

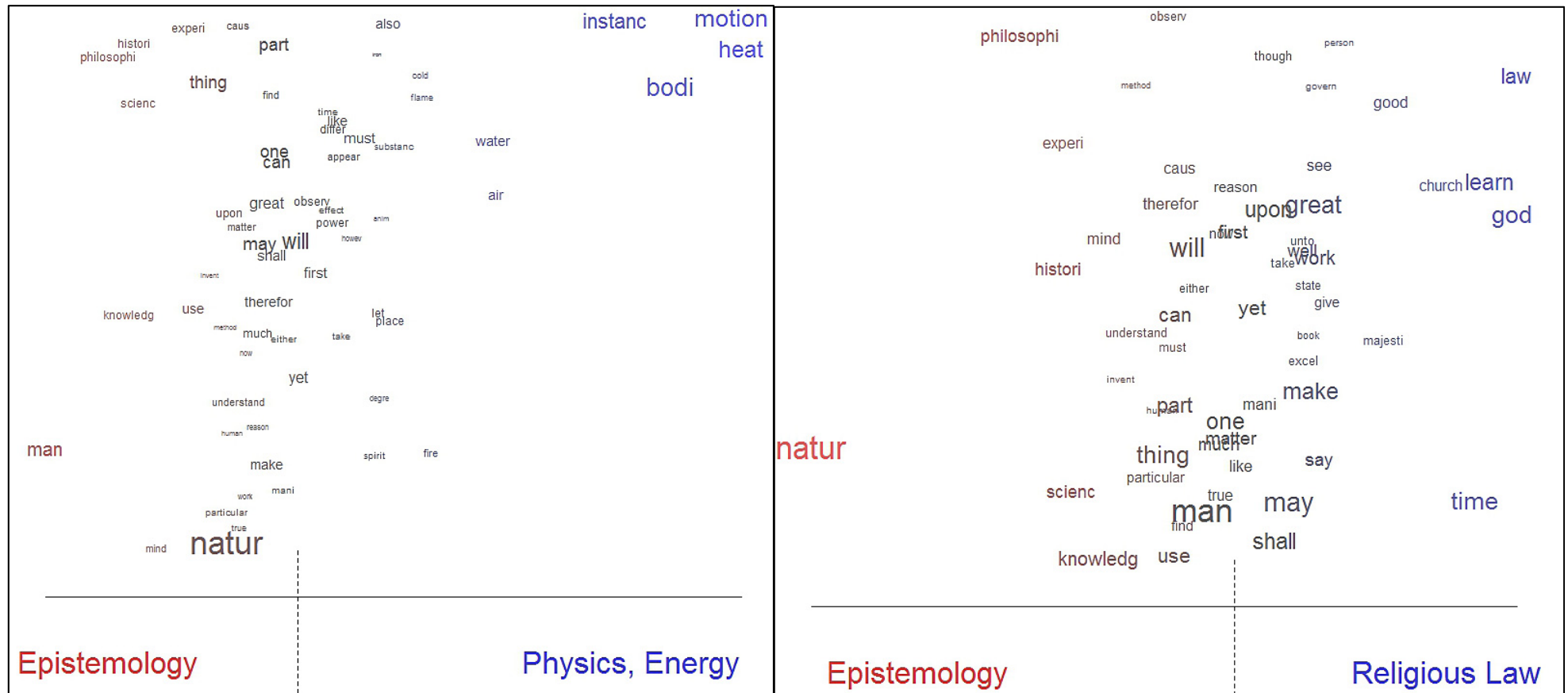


Figure 4: Vocabulary Use Across Topic Pairs: Epistemology vs. Botany and Legal Theory

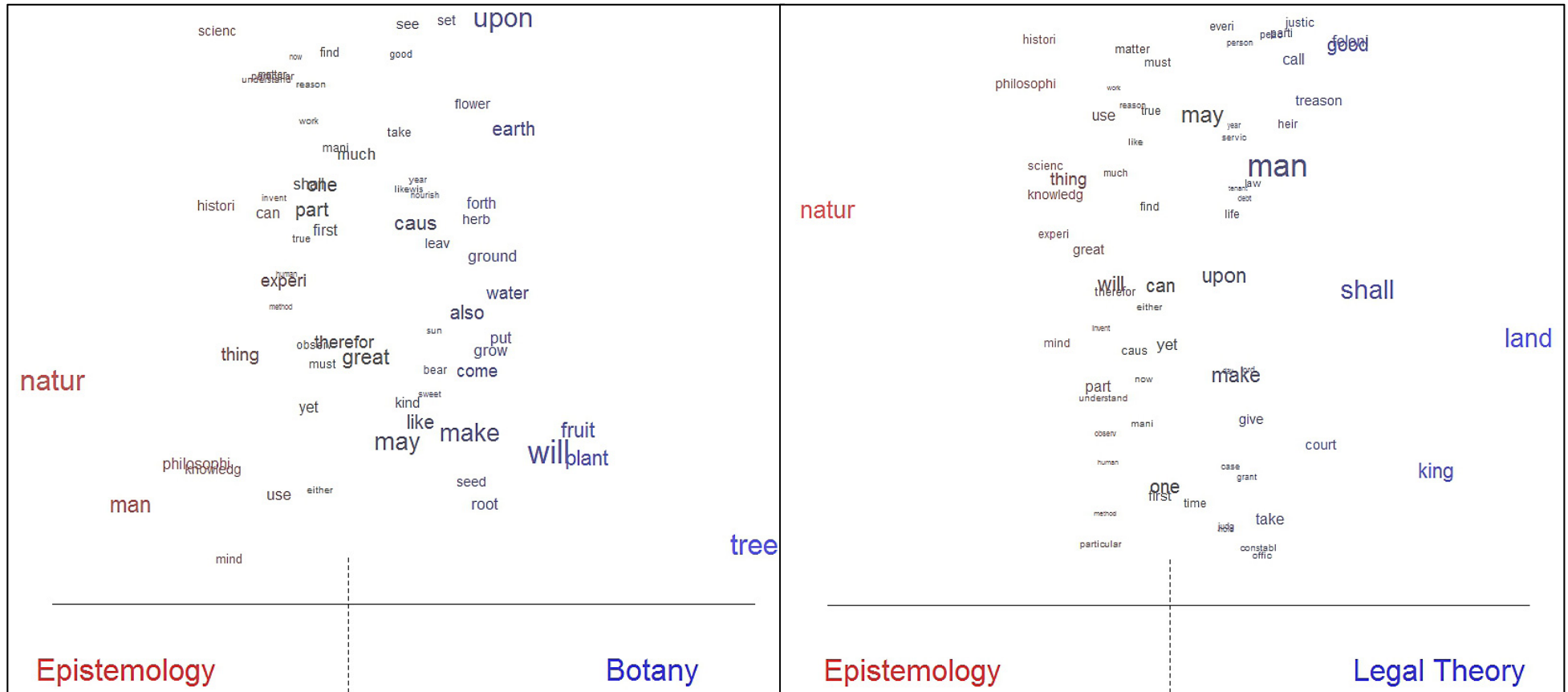


Figure 5: Vocabulary Use Across Topic Pairs: Probing for Facts vs. Physics, Energy and Religious Law

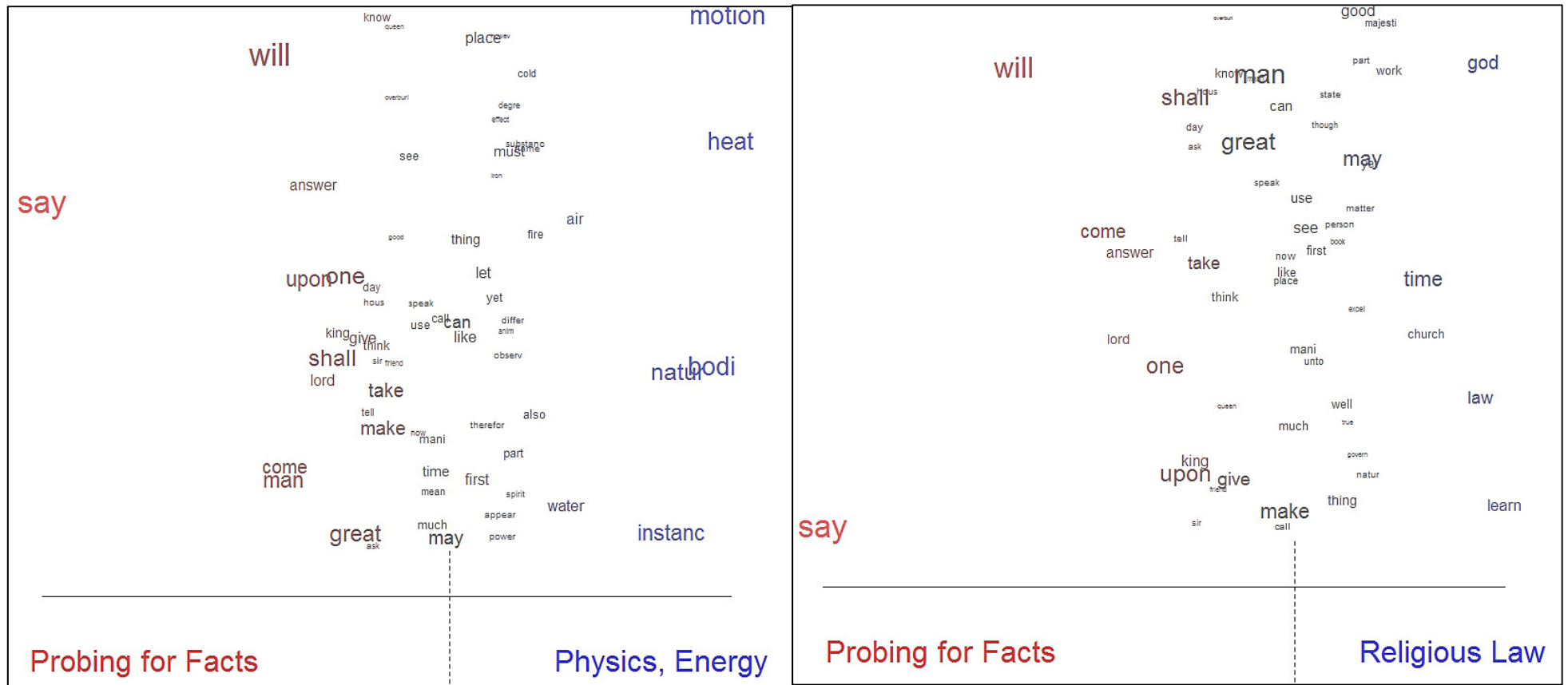


Figure 6: Vocabulary Use Across Topic Pairs: Probing for Facts vs. Botany and Legal Theory

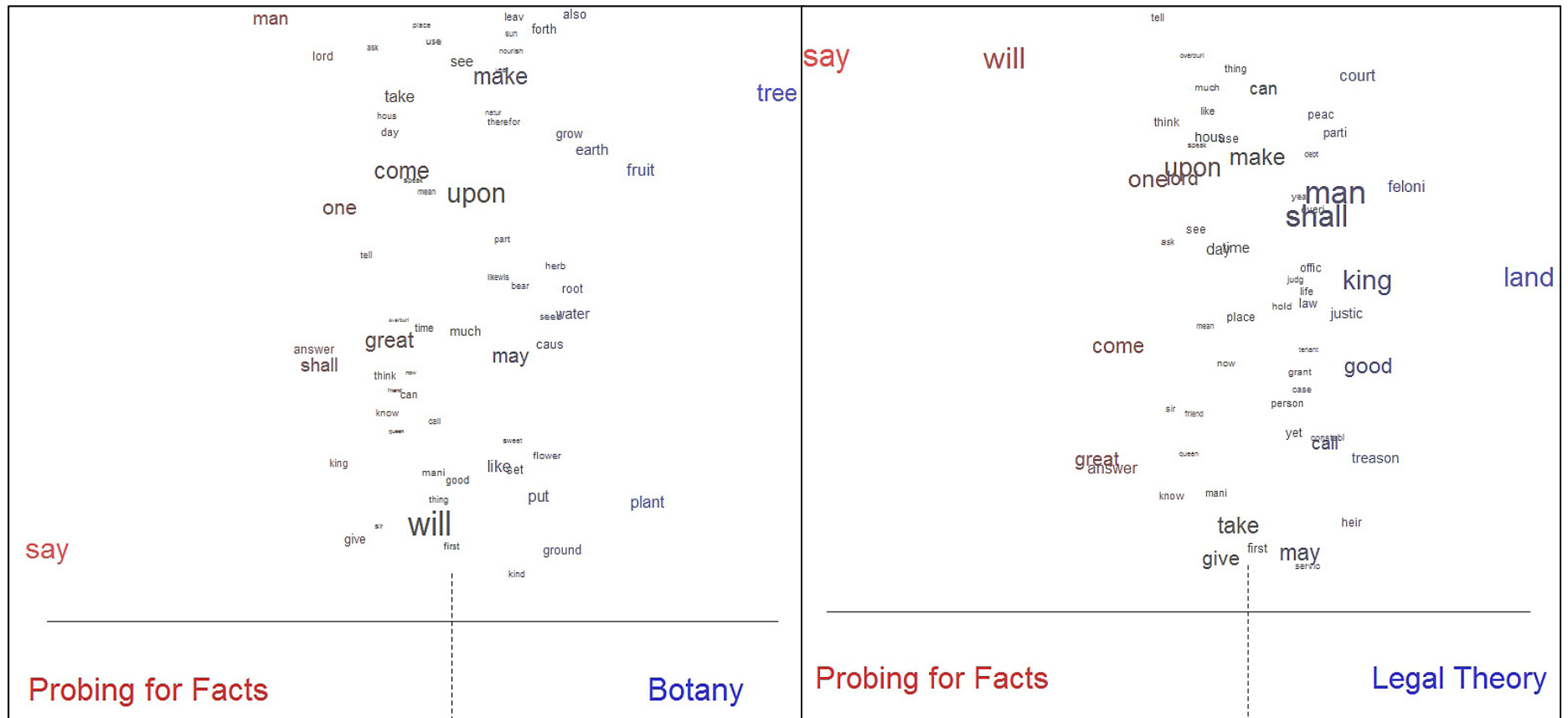


Figure 7: Vocabulary Use Across Topic Pairs: Epistemology vs. Probing for Facts

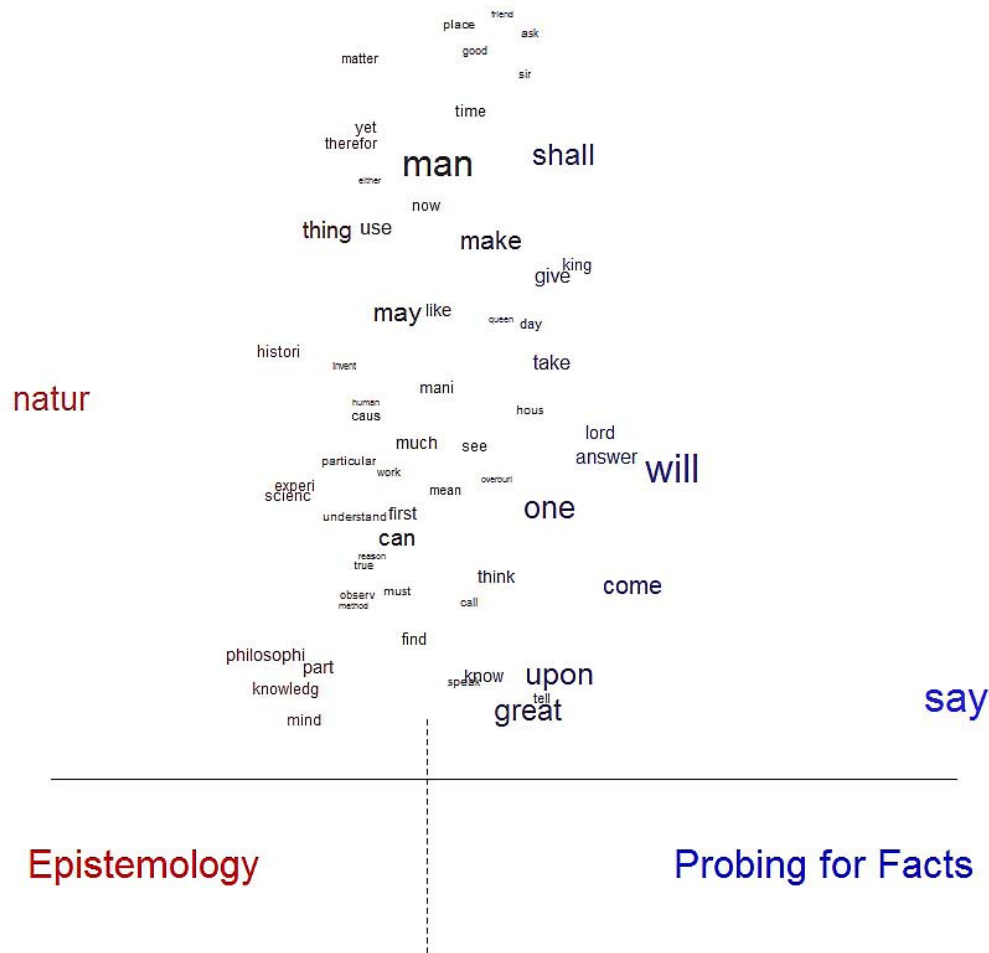
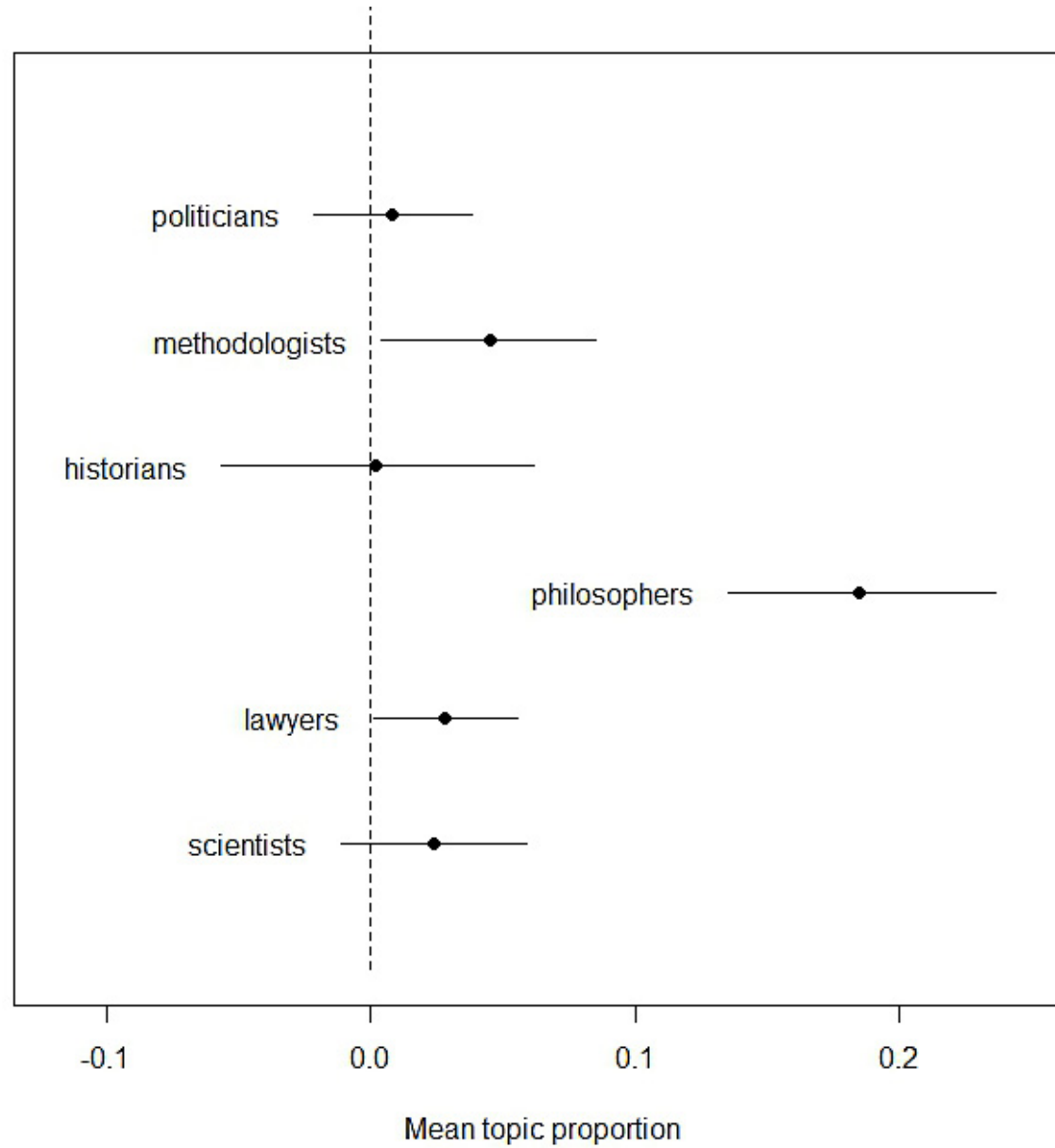
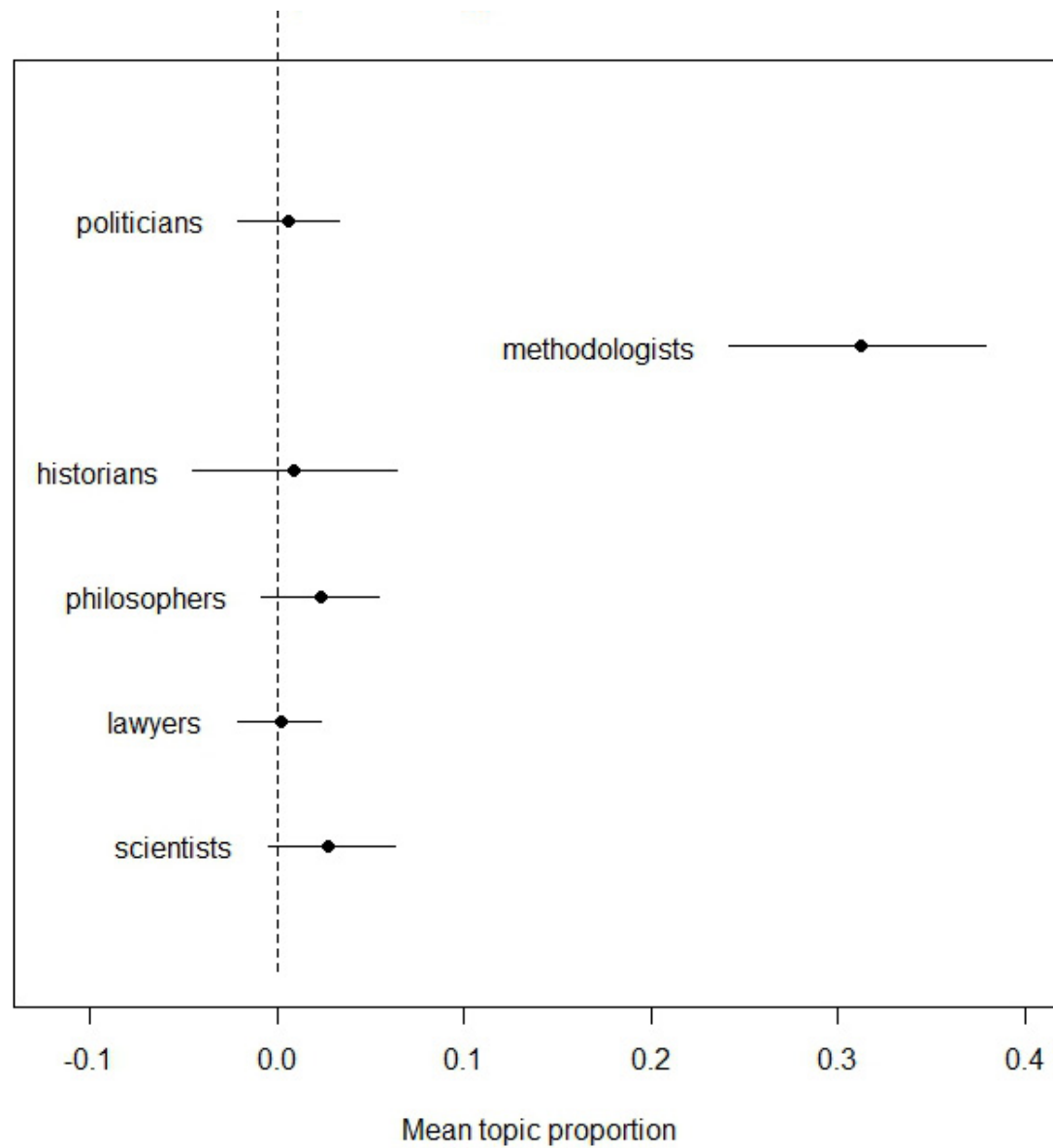


Figure 8: The Effect of Audience on Prevalence of Probing for Facts



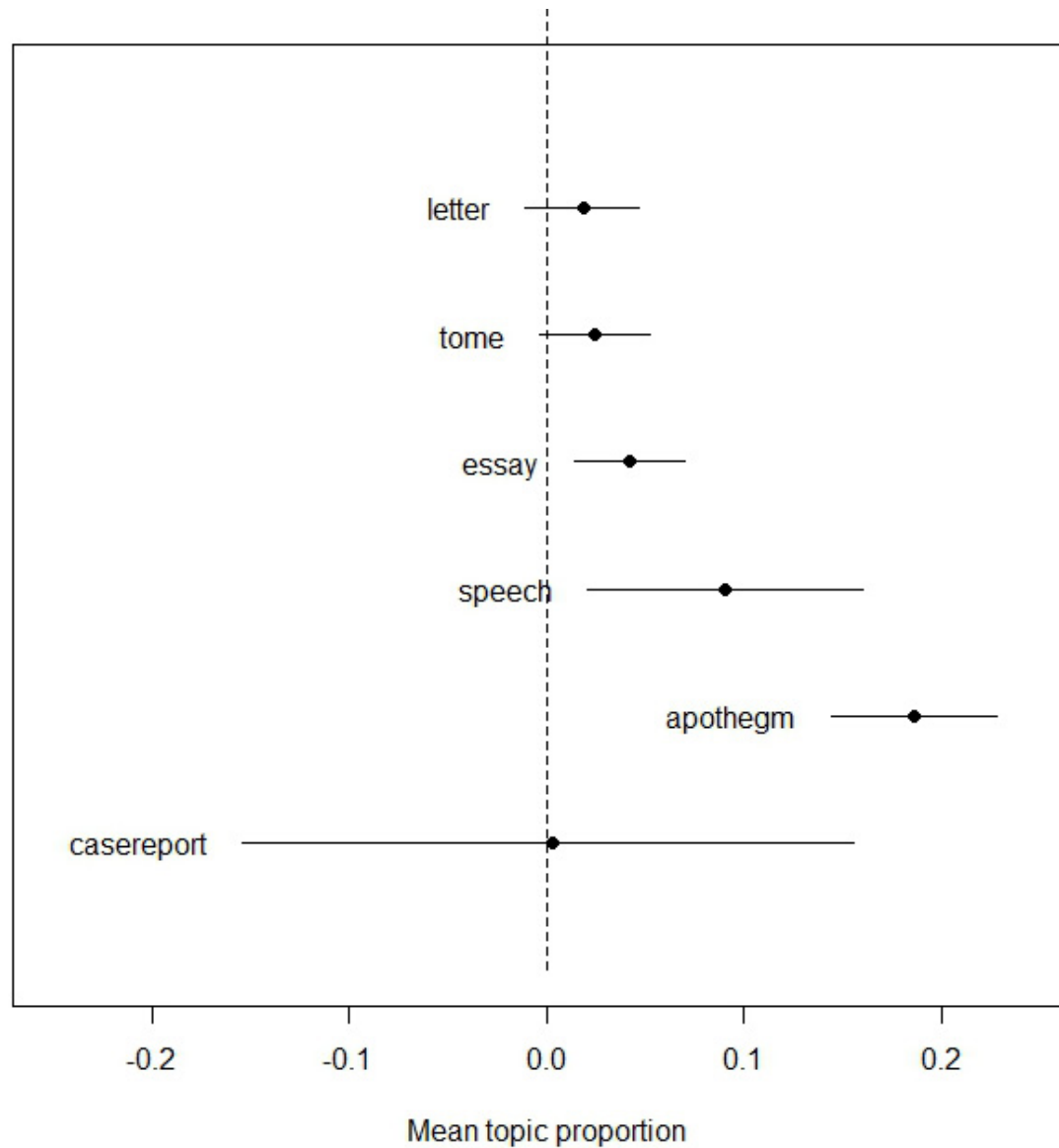
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Probing for Facts for each type of audience.

Figure 9: The Effect of Audience on Prevalence of Epistemology



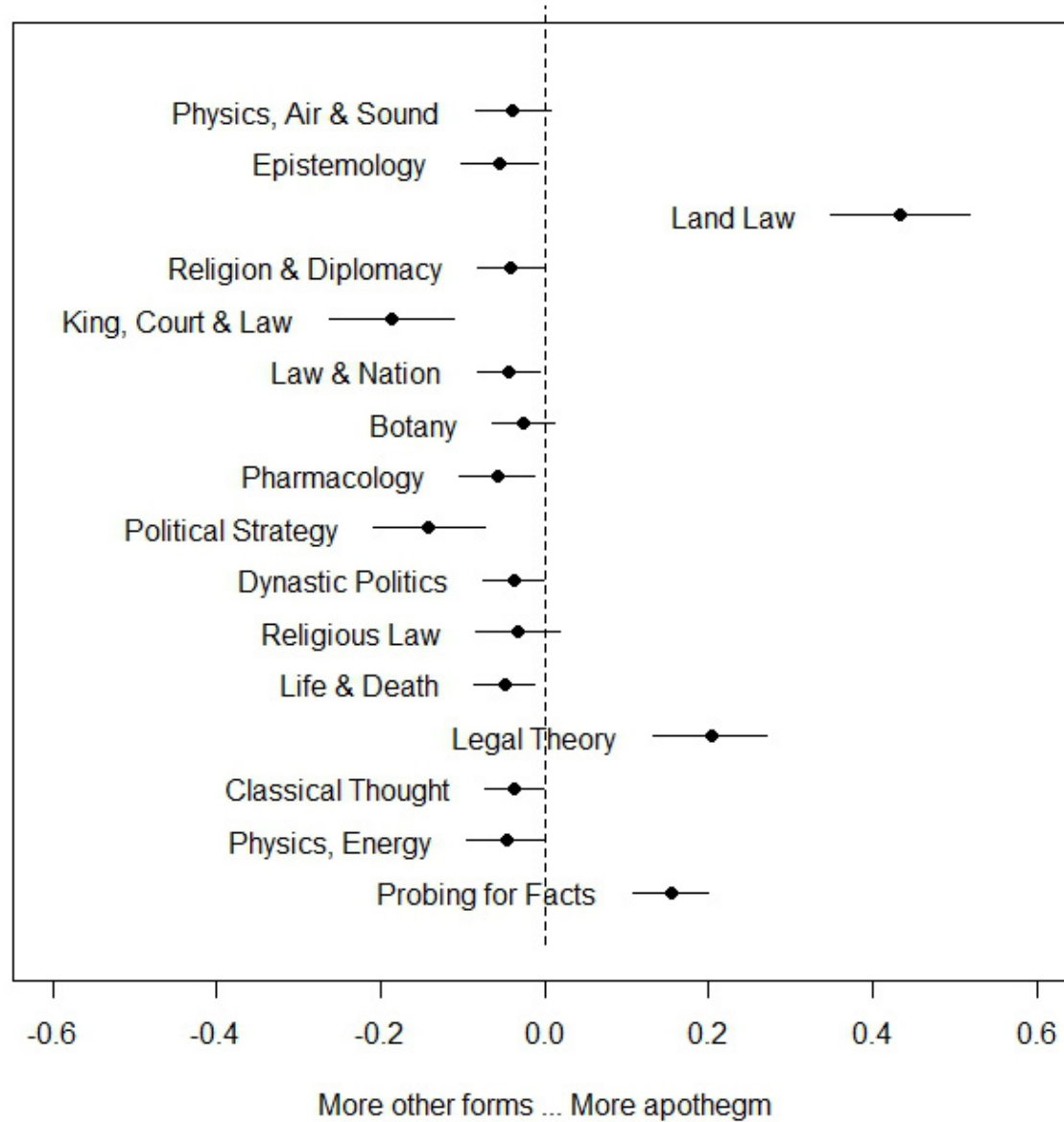
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Epistemology for each type of audience.

Figure 10: The Effect of Writing Form on Prevalence of Probing for Facts



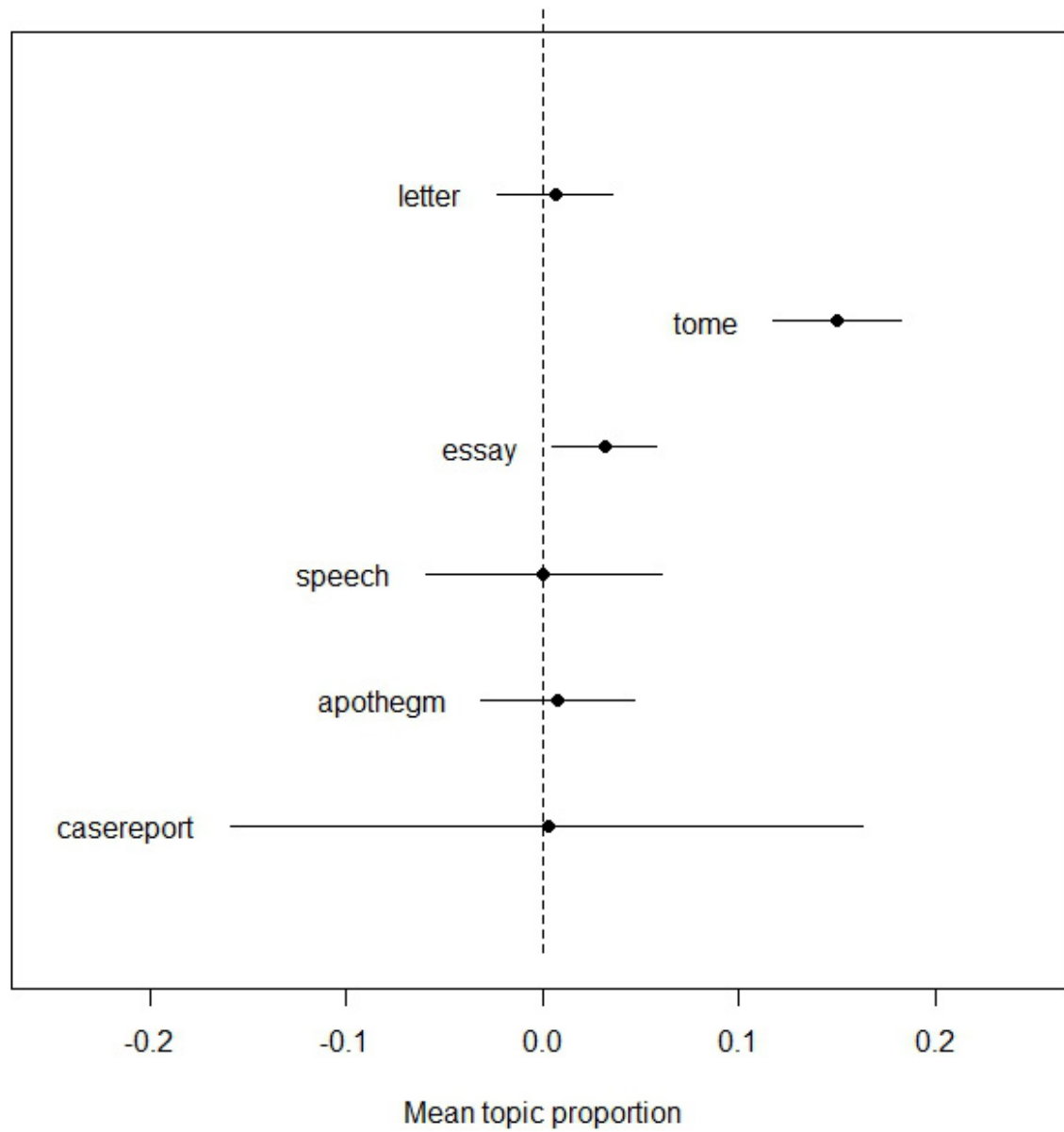
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Probing for Facts for each type of writing form.

Figure 11: The Effect of Writing Form, Apothegm



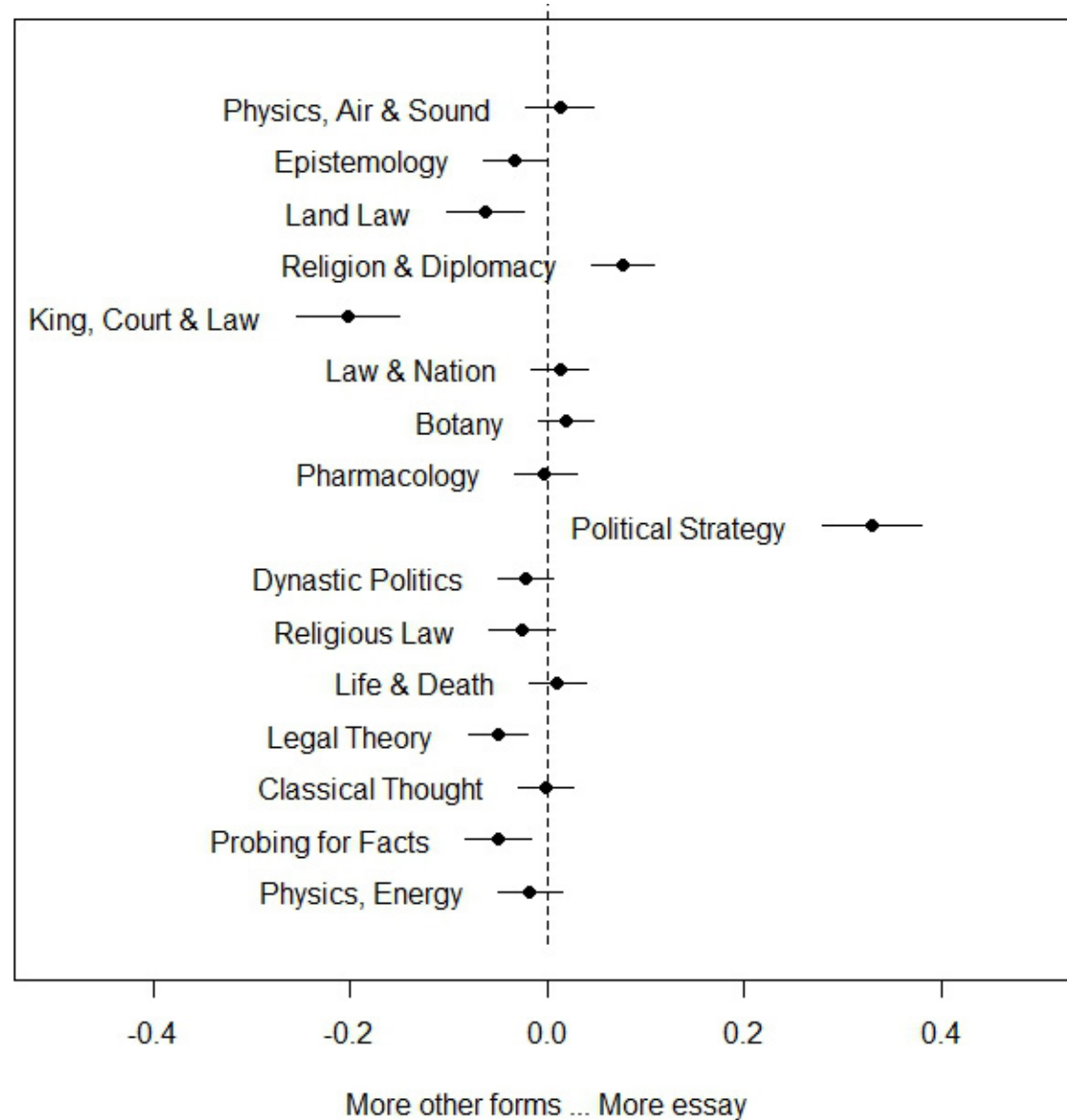
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for apothegms compared to all other document types.

Figure 12: The Effect of Writing Form on Prevalence of Epistemology



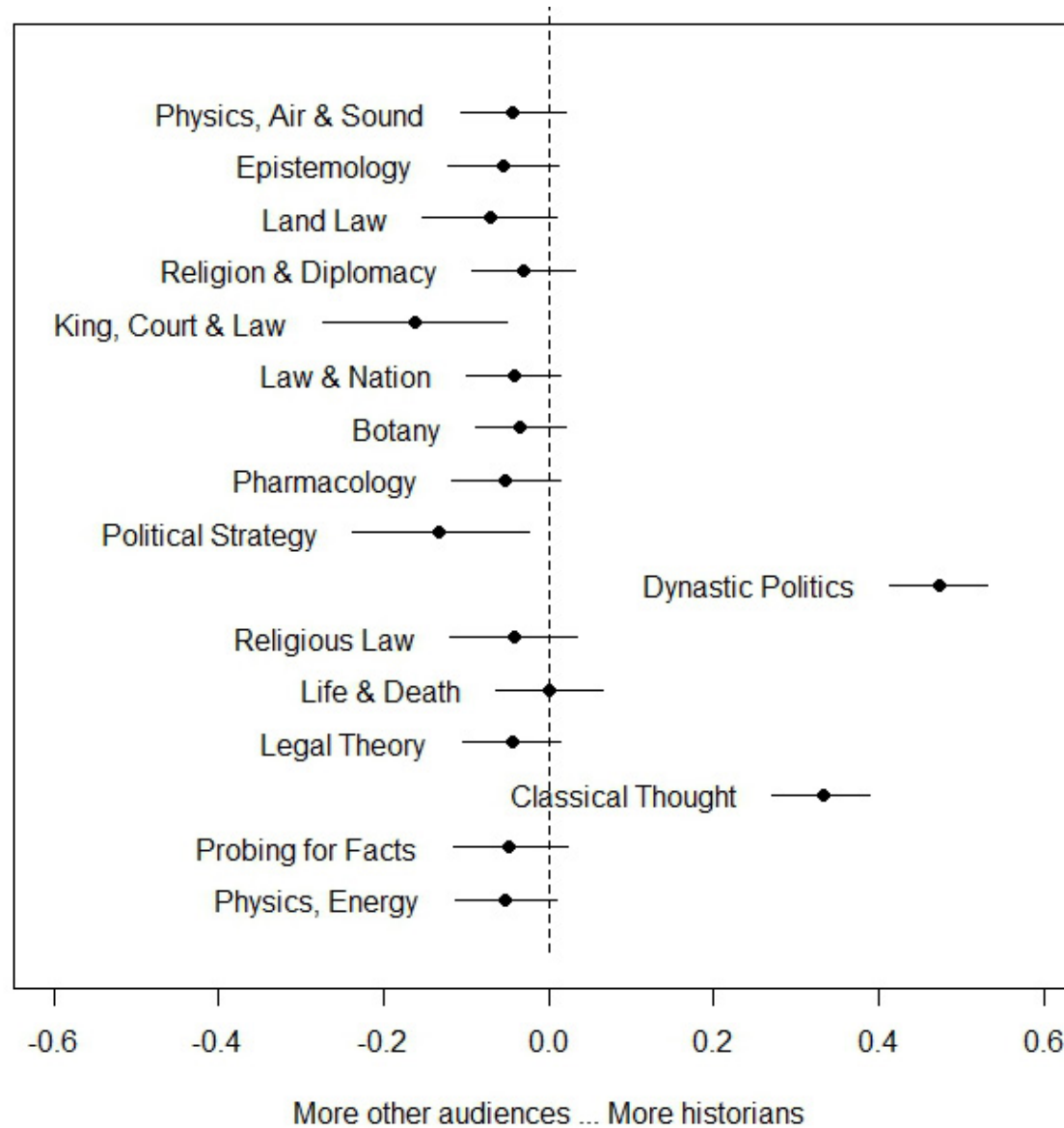
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Epistemology for each type of writing form.

Figure 13: The Effect of Writing Form, Essay



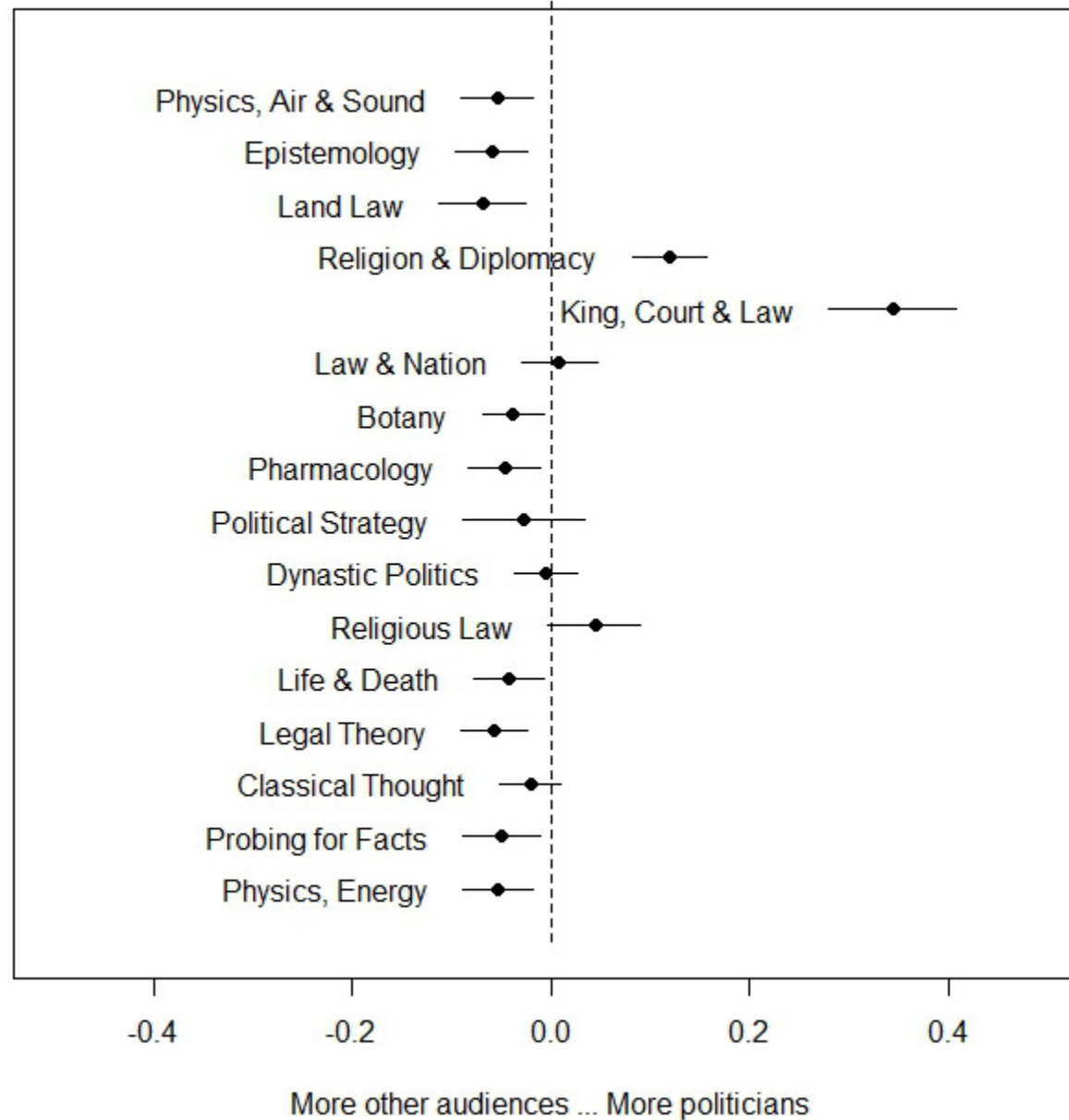
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for essays compared to all other document types.

Figure 14: The Effect of Audience, Historians



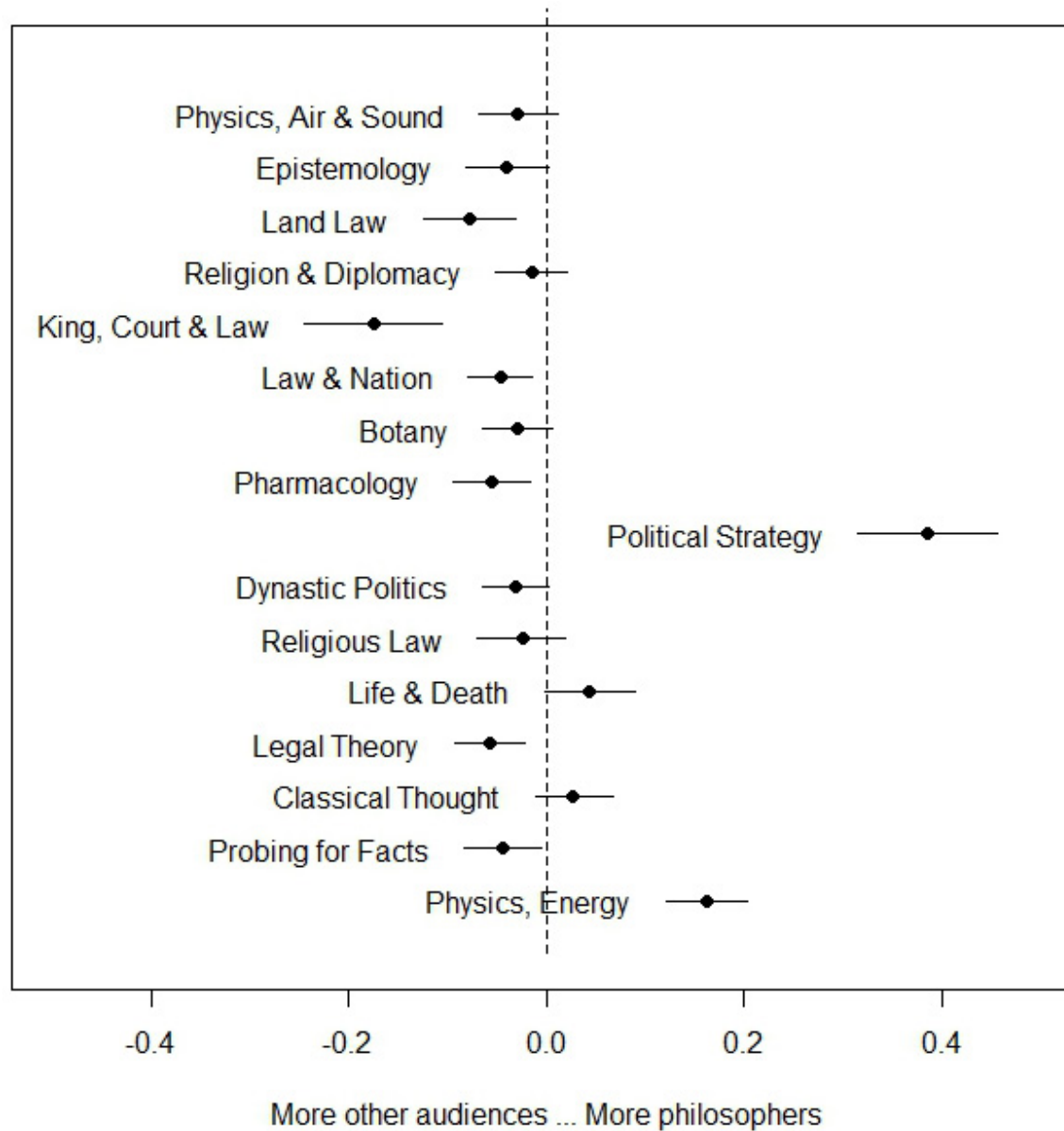
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which historians are the intended audience compared to works addressed to all other audiences.

Figure 15: The Effect of Audience, Politicians



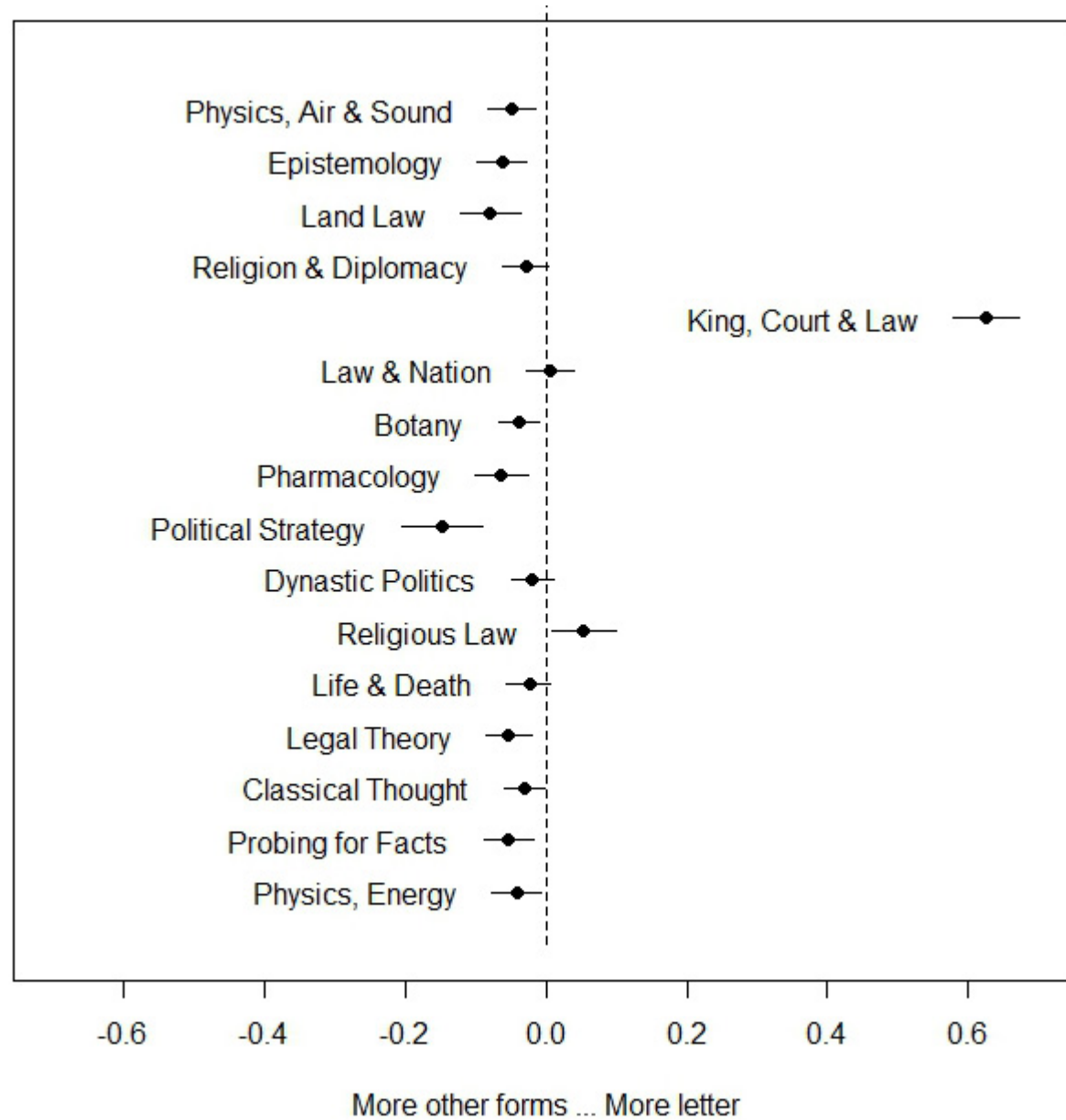
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which politicians are the intended audience compared to works addressed to all other audiences.

Figure 16: The Effect of Audience, Philosophers



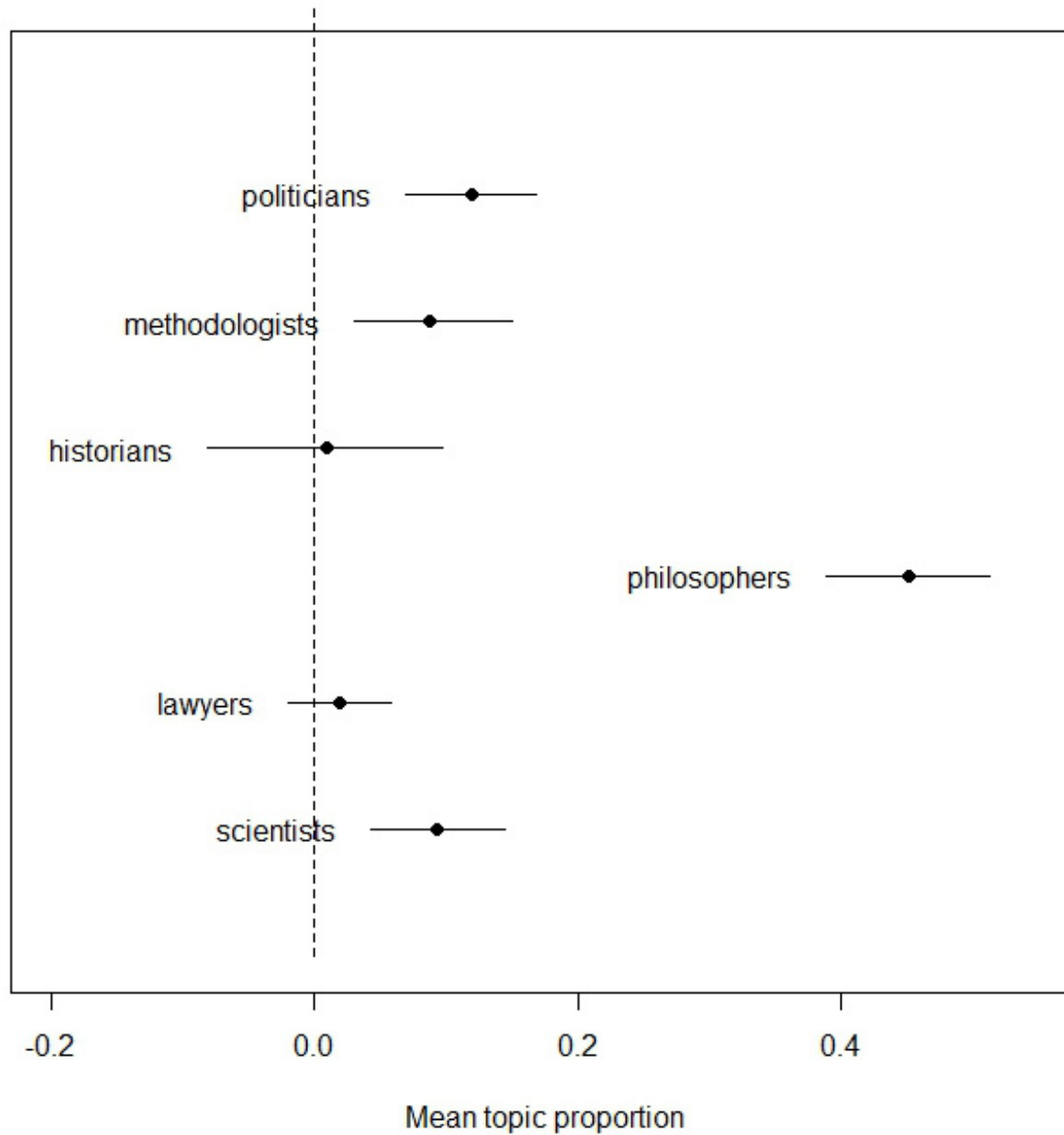
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which philosophers are the intended audience compared to works addressed to all other audiences.

Figure 17: The Effect of Writing Form, Letter



Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for letters compared to all other document types.

Figure 18: The Effect of Audience on Prevalence of Political Strategy



Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Political Strategy for each type of audience.