A Structural Topic Model of the Features and the Cultural Origins of Bacon's Ideas*

Peter Grajzl† Peter Murrell‡

March 9, 2018

Abstract

We use machine-learning methods to study the features and origins of the ideas of Francis Bacon, a key figure who provided the intellectual roots for a cultural paradigm that spurred modern economic development. We estimate a structural topic model, a state-of-the-art methodology for analysis of text corpora. The estimates uncover sixteen topics prominent in Bacon's opus. Two are key elements of the ideas usually associated with Bacon—inductive epistemology and fact-seeking. We provide the first quantitative evidence that the genesis of Bacon's epistemology lies in his common-law jurisprudence, a conclusion that has not been prominent in the conventional text-analysis literature. Fact-seeking is more isolated from Bacon's other intellectual pursuits. The utilitarian promise of science and the centralized organization of the scientific quest, embraced by Bacon's followers, were not emphasized by him. Bacon's use of different topics varies notably with intended audience and chosen medium. Our results have direct implications for the interpretation of the determinants of political and economic development in 17th-century England.

Keywords: Francis Bacon, culture, law, knowledge, scientific method, politics, religion

JEL Classifications: B31, Z10, C55, N73, K10, P10

† Department of Economics, The Williams School of Commerce, Economics, and Politics, Washington and Lee University, Lexington, VA 24450, USA and CESifo, Munich. Email: grajzlp@wlu.edu

‡ Department of Economics, University of Maryland, College Park, MD 20742, USA. Email: murrell@econ.umd.edu

* For helpful comments and suggestions we thank Michael Anderson, Nona Karalashvili, Richard Nielsen, Martin Schmidt, Dustin Tingley, participants at the workshop on Empirical Legal Studies at the University of Amsterdam, and three anonymous reviewers.

> 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

Francis Bacon plays a central role in debates on the sources of economic, institutional, and scientific development. In a recent influential 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 emphasized the possibilities of scientific advance and invention, encouraging their application to economic activities, thereby stimulating technological progress and economic growth. Mokyr identifies Bacon, the preeminent English philosopher, statesman, scientist, jurist, and author, as providing a unique stimulus to the rise of this 'Baconian' culture (Mokyr 2005, 2006, 2010).

The prominence given to Baconian culture leads directly to questions about the contribution of Bacon himself. What are the defining features of Bacon's ideas, in particular about scientific methodology, and from what spring did they flow? We address these questions using machine-learning methods applied to a large corpus of Bacon's works, for the first time producing pertinent quantitative evidence. These methods produce estimates of the salient themes emphasized by Bacon and clarify which of the elements of Baconian culture were not in fact emphasized by him.

Using the estimated themes, we then investigate the interconnectedness of the many disparate areas of Bacon's intellectual activity, focusing especially on how such connections can help explain the genesis of his scientific methodology. The interest of scholars in the sources of Bacon's ideas, and their connectedness, has inspired a voluminous literature in the history of science, political philosophy, and intellectual history more broadly. Many scholars portray Bacon as a one-of-a-kind genius alone as the 17th century began (see, e.g., Rees 2004, Urbach 1987: 24, Jardine 1974: 2). Others suggest that Bacon's ideas were heavily influenced by factors such as his religious beliefs, alchemy, magic, natural crafts, and classical philosophy, rhetoric, and science (see, e.g., Peltonen 2004, Rossi 1968: xii, 15, 39, 71; Zagorin 1998: 28; Henry 2002: 11).

Our quantitative examination of the interconnections between Bacon's varied intellectual pursuits provides strong evidence that important elements of Bacon's methodological ideas had their origins in his common-law jurisprudence. This conclusion on the genesis of Bacon's scientific method stands in contrast to the prevailing view in the existing scholarly literature on Bacon. That literature has devoted comparatively little attention to Bacon's legal background—even though he rose to the pinnacle of the legal profession and was a stout defender

of the common-law (Peltonen 1996b: 290).[1] (We justify this characterization of the literature with an extensive survey that is documented in the Appendix.) Only a very small subset of the multitude of Bacon scholars has attempted to explicitly investigate a connection between Bacon's scientific methodology and his jurisprudential background (Kocher 1957, Wheeler 1983, Simonds 1986, Cardwell 1990, Martin 1992: 164-171, Shapiro 2000: 107-112, Serjeantson 2014: 701-704).[2] While thought-provoking, the hypotheses of these scholars focus on a limited set of Bacon's ideas and works and are not accompanied by quantitative evidence.[3] Certainly, the arguments proposed by this small subset of scholars have not been widely accepted in the vast literature on Bacon and, in fact, remain contentious (see the Appendix). Even among those scholars who have investigated the possibility that Bacon's immersion in law provided some input into the development of his scientific method, there is broad disagreement on which aspects of that method may have been most shaped by his legal ideas. There is also a lack of agreement about which aspect of Bacon's jurisprudence was most influential, with some scholars emphasizing Bacon's knowledge of the civil law tradition despite his immersion in the common-law.[4] Our empirical examination sheds evidence of a new kind on these conflicting views on the connection between Bacon the lawyer and Bacon the methodologist of science.

---

[1] Supporting the even more general claim that Bacon's legal background has been understudied, Coquillette (2004: 313) argues that Bacon "has never been given the proper credit" for his "juristic writing, long overlooked". Shapiro (1980: 333) comments that "…the role of Bacon in law reform has been…obscured by the dearth of scholarly writings on the legal aspects of Bacon's thought". Earlier, Holdsworth (1927: 10) noted that "Of all Bacon's claims to greatness, his claims as a lawyer are the least of all known". De Montmorency (1905: 263) remarked that "…the world of thought is apt to forget that Francis Bacon, the Master of Laws, was a lawyer".

[2] As we clarify in the Appendix, the connection between Bacon's scientific method and his legal background has wider currency as a hypothesis that has not been investigated.

[3] The precise set of Bacon's works considered by these scholars is often not made explicit. Kocher (1957) refers to a number of Bacon's methodological works, but among these focuses almost exclusively on the *Maxims of the Law*. In his short contribution, Simonds (1986) refers to, but hardly scrutinizes, aspects of the *Maxims of the Law*, *Reading on the Statute of Uses*, and an unnamed legal dictionary. Wheeler (1983: 114) mentions, but never discusses in detail, "an informal survey" that the author undertook for purposes of his analysis. In drawing the analogy between Bacon's inquisitorial practice in court and his later examination of nature, Cardwell (1990) draws on parts of *Novum Organum* and a select set of further works. When using Bacon's own works to draw parallels between Bacon's law and science, Martin (1992: 164-170) draws primarily on Bacon's contributions about reform of the legal system and the science: "Bacon's reformed science of the law and his reformed natural philosophy have the same purpose, the same techniques, the same vocabulary and the same hierarchical organization" Martin (1992: 170). Shapiro (2000: 12) briefly refers to Bacon's legal work in the context of the need to establish facts in the legal system, but references no specific legal works by Bacon when discussing Bacon's contributions to the development of ideas on fact-verification in general (Shapiro 2000: 107-112). In conjecturing that "Bacon's general theory of the 'interpretation of nature' may perhaps also be regarded as having a significantly legal foundation", Serjeantson (2014: 701-702) draws a parallel between Bacon's study of nature and the role of investigation in the legal process in the context of "a little-studied treatise" included in a volume published decades after Bacon's death.

[4] Shapiro (2000: 107-112) argues that Bacon's legal training may have influenced his scientific method primarily via the emphasis on fact-verification. Kocher (1957) and Wheeler (1983) suggest that Bacon's jurisprudence was central to his inductive reasoning and, thus, his broader epistemological framework. Martin (1992: 2, 4, 164-170, 172, 175) mentions inductive epistemology and fact-seeking in the context of Bacon's legal theorizing but argues that Bacon's natural philosophy should be understood as reflecting Bacon's political plans for the reform of the law and the state rather than his common-law background: "…this study insists that Bacon's legal and political career was crucial in the creation of his natural philosophy and that his natural philosophy cannot be separated from his political ambitions" (Martin 1992: 2). Simonds (1986), Cardwell (1990), and Serjeantson (2014: 701-704) suggest that Bacon developed his ideas about the importance of fact-gathering and investigation based on his knowledge of continental civil and canon law. A further set of scholars who have conjectured, but not explicitly investigated, the influence of Bacon's legal background on his scientific methodology are similarly heterogeneous in their suppositions concerning which specific aspect of Bacon's natural philosophy may have been influenced by his immersion in law; see the Appendix.

Given law's relatively minor role in the intellectual-history literature on the origins of Bacon's ideas and given the lack of systematic, quantitative evidence on this subject, it is not surprising that the hypothesis of a large role for law in the development of Bacon's thought has not diffused into the economics literature.[5] Consequently, this hypothesis has not influenced debate on the sources of England's rise, a paradigmatic event that has reached iconic status in the analyses of economic and political development (see, e.g., North and Weingast 1989, Acemoglu and Robinson 2012, Mokyr 2016, Murrell 2017). Yet, evidence on a direct connection between Bacon's immersion in common-law and his scientific method carries the potential for a reinterpretation of that paradigmatic event. In particular, our evidence of a common-law origin of Bacon's scientific methodology suggests that the "culture of growth" and England's subsequent economic rise are, at least in part, rooted in the common-law tradition. We expand on this supposition in the concluding section of this paper, after we have completed our analytical inquiry.

Departing from existing analyses of Bacon, we examine the features and the cultural origins of Bacon's works using topic modeling, 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.[6] The Latent Dirichlet Allocation (LDA) model (Blei et al. 2003, Blei 2012) in particular has been applied by scholars to understand monetary policymaking (Fligstein et al. 2014, Hansen and McMahon 2016, Hansen et al. 2017), financial market performance (Larsen and Thorsrud 2015, Huang et al. 2016), and corporate behavior (Bandiera et al. 2016, Bellstam et al. 2016). We move beyond the application of LDA, and use Structural Topic Modeling (Roberts et al. 2014, 2016a), an innovation that integrates document-level data into the analysis. The inclusion of such metadata into the analysis allows us to provide state-of-the-art statistical estimates of the major features of Bacon's opus as well as assess their interrelatedness and common determinants.

Given their tremendous scope and breadth, Bacon's writings are particularly suitable for topic modeling.[7] In contrast to human analysts, topic models use a '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 especially useful in the study of an icon like Bacon, where his own numerous, disparate writings have been overlain by centuries of commentary, criticism, interpretation, and reinterpretation. Consequently, later scholars of Bacon are likely to be very sensitive to the presence of themes and aspects of Bacon's works that have been emphasized by earlier scholars, and less sensitive to the presence of other relevant features of Bacon's work. By providing a direct path to Bacon's own words and allowing the central themes in a corpus to be 'discovered' by an unsupervised machine-learning algorithm, the use of topic modeling reduces the

[5] For example, it does not appear in Mokyr's (2016: 72-74) discussion of the many sources of the ideas that inspired Bacon's work.

[6] See Gentzkow et al. (2017) for an overview of recent research using computational text analysis that is relevant for economists.

[7] Blaydes et al. (2018) emphasize that text-as-data approaches, such as topic modeling, are particularly valuable when examining a large number of texts simultaneously in order to investigate the broader themes in such texts.

dependence of the conclusions on the history of scholarly interpretation. Nevertheless, we emphasize very strongly that we are not making the argument that these techniques are superior to existing textual-analysis methods. Instead, we view machine-learning as one more, hitherto unexplored, route to obtaining insight into the nature and sources of cultural ideas.

To study Bacon's ideas 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 metadata covariates 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 with the audiences that he addressed and with the form his writing took. Thus, the inclusion of such metadata is particularly useful in the current context.

We then specified and estimated a Structural Topic Model (STM) to provide machine-learning-based, macroscopic, statistical evidence on the salient features of Bacon's work. STM identifies sixteen topics that are central in Bacon's writings, topics that would not necessarily be obvious to a reader of many disparate documents. These topics provide novel interpretations of the ideas that spanned Bacon's works. By closely examining the specific words and the particular documents that STM associates with each topic, we give each topic an evocative name suggestive of its contents. Thus our unsupervised machine-learning algorithm answers a question that is still much debated in the critical literature: What themes are emphasized in Bacon's work? (see, e.g., Peltonen 1996a, Vickers 1992). In particular, STM clearly pinpoints two themes that the interpretive literature invariably places at the center of Bacon's ideas: the emphasis on probing for facts and the epistemology of how to generalize from the facts. Our analysis also addresses the question of what themes are *not* emphasized in Bacon's work. Our evidence very clearly shows that ideas about the utilitarian promise of science and the central direction of scientific efforts, greatly emphasized by many of Bacon's followers, are not prominent in Bacon's own work. This result continues to hold in robustness tests using models estimated at much higher levels of topic granularity.

In order to develop insights into the origins of the principal features of Bacon's work, as identified by the STM, we then examine the positive correlations across documents in the use of disparate topics and the 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. Because Bacon was first immersed in law and turned his scholarly attention to natural philosophy only later, this result provides evidence in favor of the argument that the origins of Bacon's epistemological ideas lie in his common-law jurisprudence. In contrast, we find a weaker connection between Bacon's emphasis on probing for facts and his legal ideas. We interpret this as evidence that Bacon's emphasis on pursuing and ascertaining facts is comparatively closer to being a product of his own, individual-specific, emphases than is his common-law-inspired inductive epistemology. This finding also indicates that it was primarily Bacon's immersion in common law rather than his understanding of civil-law inquisitorial

practices that shaped his scientific methodology. Hence, in contrast to the few existing qualitative studies that have drawn an explicit connection between Bacon's jurisprudence and his methodology (Kocher 1957, Wheeler 1983, Martin 1992: 164-171, Shapiro 2000: 107-112; Simonds 1986, Serjeantson 2014: 701-704), our machine-learning analysis of a large corpus of Bacon's works allows us to pinpoint which specific aspect of Bacon's methodological paradigm is most closely connected to which facet of his jurisprudential thought.

We then make direct use of the defining element of STM, the integration of document-level information into estimation, to explore how the topics that Bacon emphasized varied with the intended audience and the form taken by his works. 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 form of communication. We provide quantitative evidence that supports this argument. Finally, we use the STM to illuminate additional features of Bacon's work. For example, our examination of negative topic correlations demonstrates that Bacon's opus cannot be understood as consisting of one grand, unitary project, a result providing new evidence on another long-standing controversy among Bacon scholars (Peltonen 1992: 279).

Apart from its specific implications for deliberations on early English development, our paper also has broader relevance for the comparative economics literature. We contribute to the study of culture (e.g., Guiso et al. 2006, Alesina and Giuliano 2015). Culture has been awarded a prominent role in the recent literature on economic and institutional development, with that literature often focusing on the deep historical roots of culture.[8] Our methodology shows how machine-learning can be used to isolate broad cultural ideas and to estimate the connections between them, providing insights into their historical origins. In particular, our evidence on the common-law origin of Bacon's scientific methodology suggests that the common-law, in addition to shaping a wide array of past and present-day institutional arrangements (see, e.g., Djankov et al. 2003, Beck and Levine 2003, La Porta et al. 2008, Guerriero 2016), may have been central to the emergence of the culture of science and invention that some claim made the modern world. Therefore, our analysis also provides insight into how a set of cultural ideas applicable in one domain may be influenced by cultural paradigms prevalent in other domains. Finally, understanding the importance of Bacon himself in laying the foundation for the eighteenth century 'Baconian' paradigm provides perspective on the comparative-economics question of how important the ideas and actions of influential individuals are in shaping fundamental societal and institutional change, relative to more deeply rooted forces (see e.g., Hughes 1986, Jones and Olken 2005, Besley et al. 2011, Brender and Drazen 2013).

The rest of the paper is organized as follows. We first briefly lay out the nature of structural topic modeling, emphasizing its novel elements. Then, we outline the steps we took in the

[8] On the importance of culture see, for example, Barro and McCleary (2003), Licht et al. (2007), Tabellini (2008), Kuran (2009), Algan and Cahuc (2010), Bowles and Gintis (2011), and Nunn (2012). For emphasis on deep historical roots, see Nunn (2008), Becker et al. (2016), and Guiso et al. (2016).

construction of a large machine-readable corpus of Bacon's works. 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 found within Bacon's corpus; that is, we present a statistical estimate of the defining features of Bacon's thought. We draw lessons using distinctive elements 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 features and the cultural 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 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 STM 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', that is, the principal features 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 most important words for each topic and the particular texts that feature a topic most prominently. However, the topics themselves are purely a product of model estimation. Therefore, the estimated topics identify the emphases within a corpus that might not be readily apparent to a reader of many disparate documents (Blei 2012). Importantly, our estimated topics are certainly not a product of the many interpretations of Bacon that have overlain his works in the four centuries since his death. They are also not produced in order to match words and documents to concrete issues specified in advance by the researcher as they would be in a supervised model.

The basic structure of the STM is as follows (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 $k \in \{1,\ldots,K\}$.[9] That process uses a K-dimensional vector of the parameters of a distribution that generates one of the topics $k \in \{1,\ldots,K\}$ for each position in the document. This is the topic-prevalence vector, which 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 $v \in \{1,\ldots,V\}$. This choice is determined by a topic-specific V-dimensional

[9] 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.

vector specifying the probabilities that each element of the vocabulary will be chosen to fill an empty position given that topic.[10]

Early versions of topic models viewed the topic-prevalence vector as drawn from the same distribution for all documents. Thus, all documents were interchangeable regardless of the characteristics of the document (e.g. textual form, intended audience). The key innovation of STM vis-à-vis earlier topic models is to make the topic-prevalence vector a function of the characteristics of the document. STM thus allows the researcher to incorporate data that characterizes documents into the model of the data-generating process for the corpus. Incorporation of this extra information improves the identification of topics and allows the researcher to estimate the relationship between document characteristics and topic prevalence (Roberts et al. 2014). For ease of reference, this document-level information is called the metadata, separating it conceptually from the core data input, the text in the corpus.

The key features of the data-generating process are the following (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 the metadata. Let the metadata be given by a matrix $\mathbf{X}$, each row (denoted $\mathbf{x}_d$) listing the values of all metadata covariates for document d. Then, the topic-prevalence vector for document d ($\boldsymbol{\theta}_d$) is drawn from a logistic-normal distribution with parameters that are a function of the covariate values:

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

$\boldsymbol{\Gamma}$ is a matrix of coefficients relating covariate values to topic-prevalence. With $\boldsymbol{\Sigma}$ a general variance-covariance matrix, there is the possibility of correlations across documents in the topic-prevalence vector.

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

$$z_{dn} \sim \text{Multinomial}(\boldsymbol{\theta}_d), \tag{2}$$

where the k^{th} element of z_{dn} is unity and all other elements are zero when 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:

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

[10] 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.

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:

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

The data to be used for estimation are the metadata matrix, $\mathbf{X}$, and all words in all documents, that is, w_{dn}, $n = 1,\ldots,N_d$ and $d = 1,\ldots,D$. Given these data, one estimates $\boldsymbol{\Gamma}$, $\boldsymbol{\theta}_d$, m_v, κ_{kv}, and β_{dkv} by maximizing the posterior likelihood that the observed data were 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 (Roberts et al. 2016a, 2016b). To address problems due to non-convexity, we rely on the spectral initialization approach advocated by Roberts et al. (2016b, 2016c).

3. Data

3.1. The Corpus of Bacon's Works

The construction of the corpus began with a search for all works of Bacon that had been digitized accurately enough 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 corpus and omitted from all subsequent processing steps.

Because Bacon, like all his educated contemporaries, was fluent in Latin, a very large proportion of the words that did not produce a match were Latin words. 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, replacing accented letters with their nearest ASCII equivalent.

Those 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, with stemming being a standard process by which the variant forms of a word are translated into standard forms. In Latin this process gives the inflection-less form of a word.

Once the Latin words in the corpus were in their standard forms, they were translated into English. The Latin-English dictionary relied initially on 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 assign values of the metadata covariates to each document. We coded two metavariables that capture two different characteristics of Bacon's varied opus, both highlighted as important by Bacon scholars. Peltonen (1992, 1996a: 10) argues that Bacon purposefully targeted different audiences with different ideas and strategically chose to disseminate different ideas in different forms. Therefore, our two metavariables are, first, the

intended audience—historians, methodologists, lawyers, politicians, scientists, or philosophers—and, second, the form of the finished work—case reports, apothegms, letters, essays, book-length tomes, or speeches. The coding of these metavariables was based on our own judgment using the large literature characterizing the context of Bacon's writings. 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.[11] Table 2 provides the document frequencies within each cell 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 prevalence of each topic (element of θ_d) varies with document type.

4. Estimating and Interpreting the Features of Bacon's Work

4.1. The Topics

A key initial decision to be made concerned the choice of the (fixed) number of topics to be estimated. There exists no clear-cut or unified approach to this decision, with the literature advocating the use of both computational statistical measures and human judgment (Roberts et al. 2014: 1068-1070; 2016b). We therefore proceeded by estimating a series of STM models, by varying the number of topics. We considered all models containing from 5 to 25 topics.[12] We first examined the variations across the models in the held-out likelihood and the size of residuals to assess how goodness-of-fit varied with the number of topics (Wallach et al. 2009, Taddy 2012, Roberts et al. 2016b). We then examined the set of estimated models that fit the data particularly well and compared them using their scores on average semantic coherence and exclusivity. (Coherence measures the internal consistency of the topics and exclusivity captures the extent to which topics in the model can be differentiated from each other). Our initial goal was to identify the set of models that were not strictly dominated by other models in terms of semantic coherence and exclusivity. Then, using our own judgment, we examined the cohesiveness and exclusivity of the topics in the set of non-dominated models, that is, those models located on the 'semantic coherence-exclusivity frontier' (Roberts et al. 2014). This process resulted in choosing 16 for the number of topics. We verified that all of our substantive findings were robust to small variations in the number of estimated topics.

[11] Specifically, to code the intended audience, we examined each work included in our corpus, especially the prefaces, which often make clear the intended audience. Moreover, a large proportion of the documents are letters and speeches, where there is no doubt about who is the intended recipient or the actual listeners. In some cases, where the intended audience is not clear from the nature of the document itself, we relied on the existing secondary literature on Bacon, usually that accompanying publications of the documents themselves.

[12] In a robustness check, reported in Section 4.2, we considered additional models featuring an even greater number of topics.

Table 3 presents detailed results for the topics estimated by STM. The words listed in Table 3—those that are most highly associated with 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'). Note that it is the statistical procedure that identifies the topics, together with the words and documents most highly associated with them. This is the sense in which one can truly say that STM estimates those ideas that are most prominent in Bacon's writings without using any pre-existing ideas on Bacon.

When observing some of the mundane words in Table 3, it might be worth recalling that the STM estimates are driven by correlations across documents in the occurrence of words even though, in isolation, some words provide little direct information about the fundamental ideas underlying a topic. (For example, 'thing' is quite highly associated with 9 of the 16 topics).[13] Indeed that is the aspect of STM that cannot be easily matched by human readers. An author's employment of one topic might be characterized by the subtle use of particular combinations of mundane words in ways that a reader would find very hard to detect.

The names assigned to each topic in Table 3 are our interpretations of the set of ideas 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 ideas in the documents for which a topic is most important (reflecting estimates of the θ_d in (1) above). For each estimated topic, we examined closely the twenty documents that featured a given topic most prominently. The assignment of names to the estimated topics is a crucial step in the analysis because our ultimate conclusions about the features and the cultural origins of Bacon's thought are based on which particular topics are reflected in Bacon's corpus, which topics are correlated across documents, which topics entail overlapping vocabulary, and how Bacon's usage of different topics varies with both intended audience and document form.

Readers will note below that we are able to easily identify the underlying ideas for all 16 topics and 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 dubious ones. We present all estimated topics and provide evocative names that clearly capture the content for each of the topics.

We present two lists of the 30 most important words for each topic, with the two lists reflecting different criteria of what is most important. The highest probability ('Highest prob') words are those that are most common 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 for a given topic reflect a weighted combination of two criteria:

[13] The word 'thing' (or 'particular') was often used as indicating the more modern 'fact' during Bacon's time. It was only during the 17th century that the modern usage of 'fact' became much more common (Shapiro 2000). In Bacon's works, the most common use of 'fact' was in a legal proceeding indicating the deed of which someone was accused.

association with that topic with a high probability and a greater likelihood of being associated with that topic than with other topics.[14]

Table 3 begins with two topics that were part of the core set of ideas that Bacon's later followers came to refer to as the "Baconian program" (Mokyr 2005, 2010, 2016). 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 documents are sections of the *Novum Organum* that urge that "Our only hope, then, is in genuine induction…But a really useful induction for the discovery and demonstration of the arts and sciences, should separate nature by proper rejections and exclusions, and then conclude for the affirmative, after collecting a sufficient number of negatives….[Some] may raise this question rather than objection, whether we talk of perfecting natural philosophy alone according to our method, or the other sciences also, such as logic, ethics, politics. We certainly intend to comprehend them all. And as common logic, which regulates matters by syllogisms, is applied not only to natural, but also to every other science, so our inductive method likewise comprehends them all" (Bacon 2014, Sections XIV CV CXXVII).

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 answers with evidence. 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."[15] 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, the three that are highly associated with this topic do not contain the famous descriptions of the infrastructure of science on Bensalem. This topic is

[14] 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).

[15] To enhance readability, we have edited the direct historical quotes that we provide by updating to modern spelling and punctuation.

clearly not about organization of science, the subject often associated with *New Atlantis* (Sargent 1996).

There are two 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.[16] 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): "an umbrella ideology for people to collect data" Mokyr (2016: 92).[17] Notably however, STM has found this topic in a different set of documents and expressed in a different way than is usual in the conventional textual-analysis 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 work for the Crown especially as Lord Chancellor, the chief judge of the Chancery court, which had a large role in land cases arising out the law of trusts 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 land transfer. The document most associated with this topic is Bacon's disquisition on the Statute of Uses, a 1535 law on the inheritance of 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 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. 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 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 persuaded the judges to extend the protections of the common law (within England) to all

[16] 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.

[17] Note, however, that this topic is slightly different from the aspect of Bacon's ideas emphasized in Shapiro (2000). She focuses on fact verification, while this topic centers on fact seeking as the process of asking the questions aimed at eliciting possible facts. The aspects of legal fact verification emphasized by Shapiro are not prominent in the documents highly associated with the topic.

those born in Scotland after 1603. The four documents with which this topic is most highly associated all concern this issue.

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, especially canon law. 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. In another document strongly associated with this topic, Bacon suggests updating the laws relevant to the Church of England, a long-neglected subject.

The Legal Theory topic is the one in which Bacon discusses ideas that are broader than those on one single legal issue. Many of the documents associated with this topic are from his *Maxims*, in which he 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 closely connected to either law or scientific method. Religion & Diplomacy centers on foreign affairs, at a time when these were always intertwined with religion. 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.

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 label Political Strategy. It is about specific aspects of civic engagement, which Bacon explored from a philosophical standpoint (Peltonen 1996a: 7), commenting on 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 (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."

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 challenges by pretenders to the throne—hence 'Perkin'. 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 very carefully.

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 mythological creatures. Notably, law is absent within this topic, although Bacon wrote much about government and law in Greek and Roman times. 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 often was the case, 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.

4.2. What Is and What Is Not Emphasized by Bacon?

The analysis above has given us a machine-learning, macroscopic, statistical answer to the fundamental question: What themes are emphasized in Bacon's work? (Peltonen 1996a, Vickers 1992). Above all, the STM estimates illustrate the tremendous breadth of Bacon's opus, a subject much emphasized by modern scholars (Peltonen 1992). Two core features of the set of ideas usually associated with Bacon are revealed quite clearly in STM's identification of the two topics relating to scientific method. 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 work 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).

The analysis also allows us to address the question: What themes are *not* emphasized in Bacon's work? For example, some view Bacon as inspiring the later (18th-century) emphasis on the production of useful knowledge (e.g., Rossi 1968, Gaukroger 2001: 14, Losee 2001: 61, Mokyr 2005, 2010). Others have viewed Bacon as emphasizing the necessity of large-scale, centralized organizational arrangements for scientific investigation (see, e.g., Jonas 1984, Sargent 1996, Gaukroger 2001, Harkness 2007). Interestingly, our 16-topic STM estimates do not reveal any emphasis in Bacon on either the utilitarian value of useful knowledge or on the centralized organization of the scientific quest. To explore this issue further, as a robustness check, we also estimated a series of additional STM models featuring an increasing number of topics: 32, 48, 64, and, finally, 100. Increasing the number of topics decreases standard measures of both model-fit and the average coherence of estimated topics (see above), rendering such models clearly inferior to our chosen 16-topic STM. However, the estimation of models featuring a number of topics larger than 16 allows us to explore whether an emphasis on utilitarianism and centralized organization of science can perhaps be discovered at a greater level of topic granularity than suggested by standard criteria for the choice of the number of topics. For each estimated model we followed the same process for examining the topical content as described earlier in this section. In particular, after closely inspecting the word lists for all estimated topics for a given model, we read the twenty documents that featured a given topic most prominently for all those estimated topics for which the highest probability and FREX words were even remotely suggestive of ideas that could be potentially linked to utilitarianism or organization of science.

For the 32-topic, 48-topic, and 64-topic STMs, respectively, we found no evidence whatsoever of an emphasis by Bacon on utilitarian ideas or centralized organization of science. We found virtually no indication of such ideas even in the 100-topic STM, the highest level of granularity that we explored.[18] For example, the 100-topic STM contains a topic where Bacon discusses (what we might term today) social-welfare criteria. Yet there is no evidence of an emphasis on the utilitarian promise of science in that topic. In fact, in a document from *Novum Organum* that features this topic particularly prominently, Bacon articulates a distinctly non-utilitarian view: "Truth, therefore, and utility, are here perfectly identical, and the effects are of more value as pledges of truth than from the benefit they confer on men" (Bacon 2014, Section CXXIV). Similarly, in another topic where Bacon surveys the tasks required in generating and evaluating evidence, there is no discernible emphasis on the centralized organization of the scientific quest. Indeed, in a document featuring that particular topic Bacon comments: "…in the habits and regulations of schools, universities, and the like assemblies, destined for the abode of learned men and the improvement of learning, everything is found to be opposed to the progress

[18] Detailed supporting evidence is available from the authors upon request.

of the sciences" (Bacon 2014, Section XC). Finally, neither of these two topics from the 100-topic STM were in the top half of the ranking of topics based on their relative importance in the corpus.

Given that neither the utilitarian promise of science nor the centralized organization of the scientific quest are readily identifiable within the set of STM-estimated topics, even at high levels of topic granularity, it follows that an emphasis on these themes must have been a product of later interpretations of the followers of Bacon, rather than a product of Bacon's own emphases. As Mokyr remarks (2016: 64): "The exact content of the writings of cultural entrepreneurs sometimes mattered less than the message that future generations chose to distill from it." Or as Snider (1991: 120) argues, 'Baconianism' should be understood as "a dialectical process of production and reception contracted between an author and his readers". The lack of emphasis in Bacon on utilitarianism that is revealed by our estimates is fully consistent with Rossi's (1996: 35-36) and Hill's (1997: 84) argument that Bacon's own interest in science was driven primarily by his quest for truth per se, and was thus distinctly non-utilitarian. Similarly, the absence of a distinct emphasis in Bacon on the centralized organization of the scientific quest, implied by our estimates, is fully consistent with Zagorin's (1998: 170) characterization of the *New Atlantis*: "What *New Atlantis* contains…is a partial relation of an imaginary society with a concentration on the role of science. Its political side is omitted, and the most we can find are a few clues to the latter's character and features."

Importantly, Bacon's contributions as a legal practitioner and a legal scholar have been relatively neglected in the literature, despite clear evidence in the historical record that these pursuits were important elements of his intellectual efforts.[19] The set of topics identified by our STM unmistakably reveal that Bacon was indeed, above all, a lawyer and in particular a jurisprudential scholar. Figure 1 illustrates the relative importance of the various topics in the corpus and demonstrates that legal topics are among those most prevalent in the corpus. Of course, the sheer proportion of the corpus attributed to legal topics is in part a reflection of the underlying composition of our corpus. However, recall that the STM estimation algorithm identifies topics based on the co-occurrence of words in disparate documents and conditional on metadata. This means that identification of topics by the STM should not be affected solely by the balance of different types of documents in the corpus, assuming that a full range of the author's works appears in the empirical corpus. Thus, the fact that law is featured prominently in no less than five substantively different STM-estimated topics (King, Court, & Law; Land Law; Religious Law; Legal Theory; Law & Nation) is evidence that legal reasoning is indeed a very important feature of the set of ideas underpinning Bacon's work. Law, in addition, contributes significantly to the documents associated with Probing for Facts, and is present in Religion & Diplomacy. Furthermore, as discussed above, four out of the five STM-estimated legal topics (Legal Theory; Religious Law; Law & Nation; Land Law) are featured prominently in documents where Bacon

[19] See footnote 1.

engages in jurisprudence and scholarship as opposed to writing about comparatively more mundane legal matters in the context of his professional practice of law.

5. Evidence on the Origins of Bacon's Ideas

The importance of legal ideas among the estimated topics raises the question of the potential link between Bacon's jurisprudential thought and his scientific method. We now examine this question explicitly. Given that Bacon was first and foremost educated in law (Holdsworth 1924: 139) and "played all the major legal roles" (Coquillette 1992: 2) during his professional engagement, STM-based statistical evidence of a connection between Bacon's legal topics and his methodological topics would be evidence in favor of the hypothesis that Bacon's methodological ideas are rooted in his legal background. In contrast, a connection between Bacon's non-legal topics and his methodological topics would be evidence consistent with the argument that the former were influenced by the latter or, alternatively, that influence was perhaps bidirectional. Similarly, a lack of a connection between Bacon's methodological topics and other topics would be evidence in support of the view that Bacon's scientific methodology was largely *sui generis* with him, in the sense of not being influenced by other themes in his work.

We examine two different aspects of the connection between Bacon's legal and non-legal topics and his methodological ideas: one based on document-level correlations among topics and one based on the overlap of vocabulary use across topic pairs.

5.1. Topic Correlations

An important advantage of STM over LDA is that the generative model underlying STM explicitly incorporates the possibility that topic usage is correlated across documents.[20] We use this aspect of STM to examine the interconnectedness at the document level of the various sets of ideas in Bacon's opus, developing insights into their origin. Figure 2 provides a visualization of the links between topics, where those links capture how the importance of two topics positively covaries across documents (reflecting the estimate of $\boldsymbol{\theta}_d'\boldsymbol{\theta}_d$, which in turn reflects the metadata in $\boldsymbol{\Gamma}$ and the estimates of $\mathbf{x}_d$ and $\boldsymbol{\Sigma}$).[21]

A key empirical fact about Bacon's opus, evident from Figure 2, is that Epistemology is directly correlated only with Physics, Energy and, importantly, with Religious Law, where Bacon focuses on the legal aspects of religious controversies and organizations (as opposed to religion per se; see Section 4.1). The link between Epistemology and Religious Law is slightly stronger than that between Epistemology and Physics, Energy. (The estimated correlation coefficient equals 0.0417 for Epistemology and Religious Law and 0.0189 for Epistemology and Physics, Energy.) Via Religious Law, Epistemology is further indirectly linked with another legal topic, Law & Nation. These document-level topic correlations show that Bacon combined epistemological

[20] Topic correlation is allowed also in the correlated topic model of Blei and Lafferty (2007).

[21] See Roberts et al. (2016b: 23) for 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 Figure 2 captures the 6% of correlations that are the strongest positive ones.

reasoning with discussion of both natural philosophy and the law. Given Bacon's early training in law and a later interest in scholarship about natural philosophy[22], this is one piece of evidence consistent with the hypothesis that a core aspect of Bacon's methodological ideas emerged from his immersion in the epistemology of the common law.[23] 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.

Bacon was certainly also familiar with continental legal thought and spent part of his professional career in non-common-law courts (see, e.g., Holdsworth 1938). Thus, had Bacon viewed the rules of civil-law inquisitorial procedure as providing the paramount methodological paradigm applicable to scientific methodology (see, e.g, Simonds 1986: 499, Cardwell 1990: 270, 274, 276; Serjeantson 2014: 701-702), we would have expected to detect a link between Bacon's legal topics and Probing for Facts, that is, the part of Bacon's methodological thought in which he stresses the need to collect and investigate facts. Yet we see no such connection in Figure 2; indeed, Probing for Facts is not correlated with any other STM-discovered topic. We interpret this as evidence that law's effect on Bacon's scientific methodology did not arise from his civil jurisprudence. Rather since Epistemology concerns making generalizations from facts (a characteristic of common-law reasoning) and because it is Epistemology that is correlated with a legal topic, our results indicate strongly that the roots of Bacon's methodological ideas sprang primarily from his immersion in English common law (see Wheeler 1983, Kocher 1957).

---

[22] For the topics Religious Law and Epistemology we examined the available historical records to date each of the top twenty documents that feature a given topic most prominently. Consistent with the claim that Bacon's immersion in law preceded his work in natural philosophy and scientific methodology more generally, we found that the documents featuring the Religious Law topic prominently are on average more than six years older than the documents featuring the Epistemology topic prominently.

[23] We use the term 'epistemology' in the general sense of the process by which beliefs about the world come to be established and viewed as justified beliefs. Adding connotations to this term that implied 21st-century precision would be inappropriate for the inchoate ideas on epistemology that existed in Bacon's time.

Referring to the 'epistemology of the common law' raises the question what comprised that epistemology. Since there exists in the literature no well-accepted answer to this question, since constructing an answer would require in itself a separate paper, and since any highly specific answer would invite unproductive controversy, we only provide a list of its core features, which is sufficient for the macroscopic view that this paper provides. Two elements of the method of the common-law are most relevant for our paper, the epistemology of finding/creating the law (or laws) and the process of deciding on facts in specific cases.

Finding the law depended on an implicit assumption that the underlying implicit law or even, perhaps, the optimal laws, were to be discovered in an ongoing decentralized process that depended on the deciding of real disputes, not through abstract speculations. This process involved observations of custom and an overview of past analogous cases, with analysis by trained professionals. Custom was important because it was thought to produce rules that were most useful. Rulings in past cases provided historical wisdom (see Grajzl and Murell 2016). Expert professionals were essential because they had been through a long process of training their mind in the processes of legal reasoning and absorbing a large stock of knowledge on the procedures and precedents of law. That knowledge was made available by the legal profession in the form of records of past arguments and judgments, the collection of which became more systematic over time. Conclusions about the underlying law were drawn from the common elements of both custom and many cases, using precedent, analogy, and rudimentary induction. This process was ongoing, with middle-level legal rules discovered as time passed, often stated in the form of aphorisms or maxims.

The process of deciding on facts will be familiar to most readers because the common-law model was the conduct of jury trials (see, e.g., Damaška 1986). The impartial juror gradually became central in the process. The courts relied on witness testimony and developed criteria for judging the credibility and competence of witnesses. Multiple witnesses increased the reliability of fact-finding. The expertise of witnesses was considered. The publicity of proceedings enhanced the credibility of the process. The model was one where the criterion of success was the implementation of a process that was best suited to pursue the likely answer to a highly specific question, rather than one where success was judged on whether truth had been discovered. Process rather than product was the mark of a satisfactory decision-making based on facts.

Finally, aside from the already noted link between Bacon's epistemological ideas and one aspect of his science (Physics, Energy), no other non-legal topic—including Classical Thought—is connected to any of the two methodological topics (see Figure 2). At the very least, our empirical evidence is therefore consistent with the interpretation that Bacon's exposure to non-legal ideas—including classical philosophy (see, e.g., Peltonen 2004, Zagorin 1998: 68)—was comparatively less critical for the development of his methodological thought than was his immersion in law.

5.2. Word Usage Across Topic Pairs

To further explore the connection between Bacon's jurisprudential thought and his scientific methodology, we now examine whether selected pairs of topics use similar vocabularies.[24] The degree of overlap in vocabulary use is indicative of the degree to which the contrasted topics share a common semantic foundation. Note that the analysis of word usage across topic pairs exploits different patterns in the data than do the topic correlations. An overlap in vocabulary is possible even when the topics being compared are featured prominently in quite disparate documents. In contrast, using correlations to establish a link between two topics required that the topics be featured prominently in overlapping sets of documents. In this sense, the examination of overlapping word usage adds to and complements the previous analysis of connections between topics that was based on positive correlations. The analysis of overlapping word usage may be able to detect connections between topics that could not be as easily detected via the analysis of document-level topic correlations (and vice versa).

Given the positive document-level correlation between Bacon's use of legal topics and aspects of his methodology, we produce a series of examinations of the overlap in the usage of words between topic pairs that each compare one legal topic with one methodological topic, particularly focusing on words relating to either inductive reasoning or fact-finding. Because of Bacon's education and his early professional immersion in law, evidence of an overlap in such word usage between the two types of topics would be another piece of evidence in favor of the hypothesis that Bacon's legal reasoning exerted an influence on his non-legal methodological ideas. Furthermore, following a difference-in-differences type of reasoning, we examine whether the extent of overlap in inductive or fact-finding vocabulary between a legal topic and a methodological topic is weaker than the extent of overlap in the usage of such vocabulary between a science topic and a methodological topic. If this is not the case, then this is evidence pointing to the legal origins of Bacon's methodological ideas, rather than origins outside the legal domain.

To implement the test, in Figures 3-6, we present a series of plots of the top 80 FREX words within chosen topic pairs.[25] In order to keep the number of figures to a minimum, for each of law

[24] Interestingly, Bacon himself endorsed the investigation of specific words that recur in the depiction of different domains (in his case, natural processes) as means to understanding the world (see Jardine 1974: 106-107).

[25] 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

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 such probabilities that occurs in the set of 80 words (Roberts et al. 2016b, fn. 20).[26] 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 are therefore very important in the analysis. They provide the vocabulary that the two topics have in common.

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.[27] These are words that 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. They are also at least as important (that is, used as much) in the combination of Epistemology with Religious Law as in the combination of Epistemology with Physics, Energy. Thus, at a minimum, the connection of Epistemology to Physics, Energy is 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 using Figure 4, which examines the words that Epistemology shares with Legal Theory and Botany, respectively.

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*

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.

[26] The vertical configuration of the words is random.

[27] 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'.

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 perhaps the most celebrated element of Bacon's thought has its genesis in the law. As the Appendix shows, this conclusion is not prominent, and sometimes explicitly rejected, in 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 analysis suggest that, at least for the field of law, Kocher's (1957) largely ignored argument should have been given more heed.

Figures 5 and 6 carry out the same exercise for another much-celebrated Baconian contribution, Probing for Facts. Commonalities between the pairs of topics in word usage are few. 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. Hence, the word-usage diagrams imply that Probing for Facts is, in contrast to Epistemology, much more *sui generis*, in the sense of being unrelated to other topics. This conclusion is consistent with the evidence based on the examination of positive topic correlations discussed in Section 5.1. STM-based evidence thus once more does not confirm the hypothesis that Bacon's jurisprudential background influenced his scientific method via an emphasis on fact-seeking and fact-verification (Shapiro 2000: 109, Martin 1992: 165, Wheeler 1983: 113; Simonds 1986: 499, Serjeantson 2014: 701-702). Our STM results instead suggest that Bacon's immersion in law, and in particular common law, was comparatively much more influential in providing a foundation for his inductive reasoning and the broader epistemological framework.

Yet even though Epistemology and Probing for Facts might differ with respect to their origins, there is evidence of commonality between the two topics. In Figure 7, 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 the words 'cause', 'understand', 'reason', 'know', 'therefore', 'find', 'yet', 'ask', 'mean', 'think', 'matter', 'observe', 'like', and 'method' are all shared to some degree, even if by no means evenly between the two topics. This finding is noteworthy because 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, upon recourse

to the text and using statistical methods, 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.

6. Further Evidence on the Features of Bacon's Opus

Our STM-based evidence sheds light on additional features of Bacon's work debated in the scholarship on Bacon. In what follows, we focus on the questions about the unity of Bacon's opus as a whole (Section 6.1) and the impact of Bacon's intended audience and chosen medium (Section 6.2).

6.1. Assessing the Unity of Bacon's Opus as a Whole

According to Peltonen (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" Peltonen (1992: 279). In our discussion of Figure 2, we have already commented on the presence or absence of positive connections between specific topics emphasized by Bacon. To assess the unity of Bacon's works, we now examine the topics that are inversely correlated across documents. Although negative correlations are not typically highlighted in the topic modeling literature, they can provide valuable insights about the features of Bacon's opus. For example, the existence of topics that are negatively correlated with multiple other topics would be one indication of the non-unitary character of Bacon's work.

Figure 8 shows the connections between topics arising from the set of correlations that have the largest negative values.[28] The central feature of this diagram is the inverse correlation of King, Court & Law with five other topics. In particular, the negative correlation between King, Court & Law and Epistemology demonstrates that when Bacon was discussing matters of law relating to or otherwise close to the monarchy he was less likely to use his epistemology than elsewhere. As indicated in Section 4.1, 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 other legal domains (Land Law, Legal Theory).[29]

In addition to having identified a topic that is negatively correlated with multiple other topics, there are also topics—such as Dynastic Politics, Classical Thought, and Probing for Facts—that are neither positively nor negatively correlated with any other topic (see Figures 2 and 8) and are

[28] 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 8 captures the 7% of correlations that are the strongest negative ones.

[29] Figure 8 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 8 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 and does not hold in general.

in this sense isolated from the rest of the salient themes in the corpus. Thus, we hardly find evidence of a single consistent grand project. We therefore endorse the conclusion of Peltonen (1992: 281) that "Bacon's different writings did not form a consistent and coherent system".

6.2. The Role of Intended Audience and Writing Form

Does Bacon's emphasis on specific topics vary systematically with observable document-level characteristics and if so how? We now make direct use of the defining element of STM—the inclusion of metadata covariates into topic model estimation—to obtain further insights into the features of Bacon's ideas by analyzing 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 devoted to a topic (reflecting estimates of the θ_d in (1) above) and the values of a particular covariate for all documents, thereby obtaining estimates of the pertinent column of Γ in (1) above (Roberts et al. 2016b: 17-18). We then plot either mean estimated topic proportions for different document-level covariate values (Figures 9-11, 13, and 18) or mean differences in estimated topic proportions for two different values of a document-level covariate (Figures 12, 14-17, 19).

We first examine the role of the intended audience. Peltonen (1992) and Zagorin (1998: 180), for example, emphasize that an understanding of Bacon's opus requires 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 two central topics that were much emphasized by Bacon's followers, Probing for Facts and Epistemology, varies with the intended audience. Figures 9 and 10 summarize the results.

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 his Epistemology most of all when addressing methodologists, thus directing his work toward those social actors who would find it most useful. Interestingly, none of the core elements of Bacon's ideas on scientific methodology seems to be have been targeted at either historians or politicians.

Were the core methodological elements of Bacon's thought relatively favored in some writing forms more than in 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 11 shows that when articulating his convictions about Probing for Facts, Bacon preferred apothegms to other writing forms. Consistent with Kocher (1957), Figure 12 shows that Bacon relied more on apothegms when he expressed some of his legal ideas (Legal Theory; Land

Law). Bacon, however, does not use apothegms to convey his Epistemology.[30] Indeed, as Figure 13 illustrates, Bacon strove to develop his epistemological ideas primarily in book-length volumes and, to some extent, essays. This finding resonates with Peltonen's (1996a: 10) interpretation that in order to reach new audiences Bacon deliberately chose to disseminate his groundbreaking views about the genesis of knowledge in a broadly-available printed form, despite the fact that such publication had at the time been viewed as "degrading" for a member of the aristocracy.

Our STM can address additional questions about the content of Bacon's opus that have been much debated by modern-day scholars (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". We find that Bacon indeed used essays (Figure 14) primarily to communicate his theories of Political Strategy, a particular facet of civil knowledge in which he was passionately interested (see Section 4.1). However, Bacon also relied on his essays to express his views about foreign policy and religion (Religion & Diplomacy).

Second, consistent with our observations on Figures 9 and 10, when Bacon addressed historians (Figure 15), he did not emphasize his methodology. 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 16), his methodological ideas took a back seat, a finding suggesting that his supposed unstinting advocacy of his scientific methodology did have limits. When he addressed politicians, Bacon emphasized religion, in connection with both law and diplomacy.[31] Interestingly, Bacon did not choose politicians as his audience when communicating his insights about political tactics and the games politicians play (Political Strategy), but rather intended those insights as contributions to philosophical debates (Figures 17, 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. He did, however, communicate extensively to politicians about official legal matters of the crown (King, Law & Court), but he did so in private letters, as opposed to published works (Figure 19). These findings are consistent with the prevailing view of Bacon as a shrewd political actor, adept at strategically manouvering across the political landscape and relentless at pursuing his personal political goals (Peltonen 1996b, Drinker Bowen 1993, Powell 1996).

7. Conclusion

Using the new methodological tool of structural topic modeling (STM), we have drawn on machine learning to examine the features and the origins of Francis Bacon's ideas. Our analysis identifies the salient themes in Bacon's work, examines their connectedness, and highlights the key

[30] 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.

[31] In our coding of the metadata, monarchs are politicians.

role of his choice of intended audience and chosen medium. In particular, our quantitative examination of a large corpus of Bacon's works allows us to revisit the enduring scholarly debate about the genesis of Bacon's ideas on scientific methodology. This is of direct interest to economists because of Bacon's pivotal role as a "cultural entrepreneur" in the emergence of the "culture of growth" that was a central factor in England's economic rise (Mokyr 2016).

Nearly a century ago the venerable Holdsworth (1927: 10) commented: "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." STM allows us to bring quantitative evidence to bear on the validity of Holdworth's still largely ignored (see Appendix) conjecture on the genesis of Bacon's scientific methodology.

Interestingly, Bacon himself viewed his insights on scientific methodology "the offspring of time rather than wit", a statement that itself has clear origins in 17th-century common-law ideas about the development of law (Grajzl and Murrell 2016).[32] Resonating with Holdsworth and consistent with Bacon's view, our STM estimates point to the legal, and in particular common-law, roots of Bacon's epistemological reasoning. However, our evidence also shows that not all of Bacon's ideas about scientific methodology can be decisively attributed to his immersion in law. Specifically, in the empirical patterns discovered by STM, Bacon's emphasis on fact-seeking is much less strongly related to his legal themes than is his emphasis on epistemological reasoning. This suggests that at least some aspects of Bacon's scientific methodology were to a notable extent *sui generis* with him.

STM also allows us to highlight which themes are not emphasized in a corpus, thereby offering the possibility of checking whether ideas attributed to a particular individual by a later interpretive literature can be readily identified in a macroscopic statistical analysis of the person's works. The promise of science as a means to improve the human condition and the need for centralized organization of scientific pursuits, for example, have often been identified as key components of the "Baconian program", the broad cultural paradigm that existed in the centuries after Bacon (Mokyr 2005, 2010, 2016). Our evidence, however, reveals no emphasis in Bacon on the utilitarian argument for the importance of science or on the centralized organization of the scientific quest. In this sense, our findings echo Mokyr's (2016: 78) view that "The adoption of [Bacon's] ideas by the eighteenth century Enlightenment intellectuals was highly selective and was made to suit [his followers'] agenda". Our analysis therefore suggests the necessity for further research on how Bacon's ideas were buffeted by the forces of cultural evolution, eventually turning into the Baconian program that was so influential.

[32] 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.

We close with some implications of our results for the interpretation of 17th-century English economic and institutional development, a central case study in the economics literature (North and Weingast 1989, Acemoglu and Robinson 2012, Mokyr 2016, Murrell 2017). We have provided evidence that there was a large role for common-law in the development of Bacon's thought. Since an education in law was prevalent among the upper reaches of English society at that time, this could explain why the diffusion of Bacon's ideas about scientific methodology occurred earlier in England than in the rest of Europe. Given Mokyr's (2005, 2006, 2010) compelling case for the influence of Bacon's methodological ideas on European economic development, this early diffusion, in turn, could have been one factor causing the earlier economic development of England. Additionally, scholarship in political science and law has argued that English constitutionalism was fostered by England's common-law culture (see, e.g., Cromartie 2006). Perhaps, then, the oft-emphasized association between institutional changes and subsequent economic expansion in 17th- and early 18th-century England reflects the presence of an omitted variable, the culture of the common-law, rather than a causal connection. Of course, an examination of these conjectures extends far beyond the substance and scope of this paper, but our results certainly provide impetus for future research on these vital topics.

References

Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York, NY: 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.

Bacon, Francis. 2014. *Novum Organum Or True Suggestions for the Interpretation of Nature*. Edited by Joseph Devey. Project Gutenberg EBook #45988. https://www.gutenberg.org/files/45988/45988-h/45988-h.htm. Accessed February 2018.

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.

Beck, Thorsten, Demirguc-Kunt, Asli, and Ross Levine. 2003. "Law and Finance: Why Does Legal Origin Matter?" *Journal of Comparative Economics*, 31:4, 653-675.

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.

Blaydes, Lisa, Justin Grimmer, and Alison McQueen. 2018. "Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds." *Journal of Politics*, forthcoming.

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.

Cardwell, Kenneth W. 1990. "Francis Bacon, Inquisitor." In: Sessions, William A. (ed.), *Francis Bacon's Legacy of Texts*. New York, NY: AMS Press, 269-289.

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, Danier R. 2004. *The Anglo-American Legal Heritage, Introductory Materials.* Second Edition. Durham, NC: Carolina Academic Press.

Coquillette, Daniel R. 1992. *Francis Bacon*. Stanford, CA: Stanford University Press.

Cromartie, Alan. 2006. *The Constitutionalist Revolution: An Essay on the History of England, 1450-1642*. Cambridge, UK: Cambridge University Press.

Damaška, Mirjan R. 1986. *The Faces of Justice and State Authority. A Comparative Approach to the Legal Process*. New Haven, CT: Yale University Press.

de Montmorency, J.E.G. 1905. "Francis Bacon." *Journal of the Society of Comparative Legislation*, 6:2, 263-283.

Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. "The New Comparative Economics." *Journal of Comparative Economics*, 31:4, 595-619.

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.

Gaukroger, Stephen. 2001. *Francis Bacon and the Transformation of Early-Modern Philosophy*. Cambridge, UK: Cambridge University Press.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy. 2017. "Text as Data." NBER Working Paper No. 23276.

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.

Guerriero, Carmine. 2016. "Endogenous Legal Traditions and Economic Outcomes." *Journal of Comparative Economics*, 44:2, 416-433.

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. 2017. "Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach." *Quarterly Journal of Economics*, forthcoming.

Harkness, Deborah E. 2007. *The Jewel House: Elizabethan London and the Scientific Revolution*. New Haven, CT: Yale University Press.

Hathi Trust. n.d. https://www.hathitrust.org. Accessed January-May 2016.

Henry, John. 2002. *Knowledge Is Power: Francis Bacon and the Method of Science*. London, UK: Icon Books.

Hill, Christopher. 1997. *Intellectual Origins of the English Revolution Revisited*. New York, NY: Oxford University Press.

Holdsworth, William Searle. 1938. *Some Makers of English Law*. London, UK: Cambridge University Press.

Holdsworth, William Searle. 1922. *A History of English Law, Volume V*. Boston, MA: Little, Brown, and Company.

Holdsworth, William Searle 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.

Hughes, Jonathan. 1986. *The Vital Few: The Entrepreneur and American Economic Progress.* New York, NY: Oxford University Press.

Internet Archive. n.d. https://archive.org. Accessed January-May 2016.

Jardine, Lisa. 1974. *Francis Bacon, Discovery and the Art of Discourse*. London, UK: Cambridge University Press.

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.

Kocher, Paul H. 1957. "Francis Bacon on the Science of Jurisprudence." *Journal of the History of Ideas*, 18:1, 3-26.

Kuran, Timur. 2009. "Preface: The Economic Impact of Culture, Religion and the Law." *Journal of Economic Behavior and Organization*, 71, 589-592.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature*, 46:2, 285-332.

Larsen, Vegard H. and Leif Anders Thorsrud. 2015. "The Value of News." Unpublished manuscript.

Licht, Amir N., Chanan Goldschmidt, and Shalom H. Schwartz. 2007. "Culture Rules: The Foundations of the Rule of Law and Other Norms of Governance." *Journal of Comparative Economics*, 35:4, 659-688.

Losee, John. 2001. *A Historical Introduction to the Philosophy of Science*. Cambridge, UK: Oxford University Press.

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.

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: Dolfsma, 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. 2017. "Design and Evolution in Institutional Development: The Insignificance of the English Bill of Rights." *Journal of Comparative Economics*, 45:1, 36-55.

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. 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. Accesse d 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.

Rees, Graham. 2004. "On Francis Bacon's Originality." *Intellectual News*, 14:1, 69-74.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2016c. "Navigating the Local Modes of Big Data: The Case of Topic Models." In: Alvarez, R. Michael (ed.), *Computational Social Science, Discovery and Prediction*. New York, NY: Cambridge University Press, 51-97.

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. Airoldi. 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. 1968. *Francis Bacon: From Magic to Science*. Chicago, IL: The University of Chicago Press.

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.

Serjeantson, Richard. 2014. "Francis Bacon and the 'Interpretation of Nature' in the Late Renaissance. *Isis*, 105:4, 681-705.

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.

Simonds, Roger T. 1986. "Bacon's Legal Learning: Its Influence on His Philosophical Ideas." In: McFarlane, D. (ed.), *Acta Conventus Neo-Latini Sanctandreani*, Proceedings of the Fifth International Congress of Neo- Latin Studies.. Binghamton, NY: Center for Medieval and Early Renaissance Studies.

Snider, Alvin. 1991. "Bacon, Legitimation, and the 'Origin' of Restoration Science." *The Eighteenth Century*, 32:2, 119-138.

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.

Urbach, Peter. 1987. *Francis Bacon's Philosophy of Science: An Account and a Reappraisal*. Chicago, IL: Open Court.

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. 1983. "The Invention of Modern Empiricism: Juridical Foundations of Francis Bacon's Philosophy of Science." *Law Library Journal*, 76, 78-120.

Whitaker's Words. 2006. http://archives.nd.edu/whitaker/wordsoth.htm and http://archives.nd.edu/whitaker/dictpage.htm. Accessed May 15, 2016.

Zagorin, Perez. 1998. *Francis Bacon*. Princeton, NJ: Princeton University Press.

Table 1: Works Included in the Corpus

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

Notes: Some names edited for rendering in modern English.

Table 2: Frequency of Documents by Writing Form and Audience

	Writing Form						
Audience	Apothegm	Case report	Essay	Letter	Speech	Tome	Total
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, bruge, 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, rosemari, mushroom, wheat, vine, escul, colewort, compost, mistleto, turnip, fruit

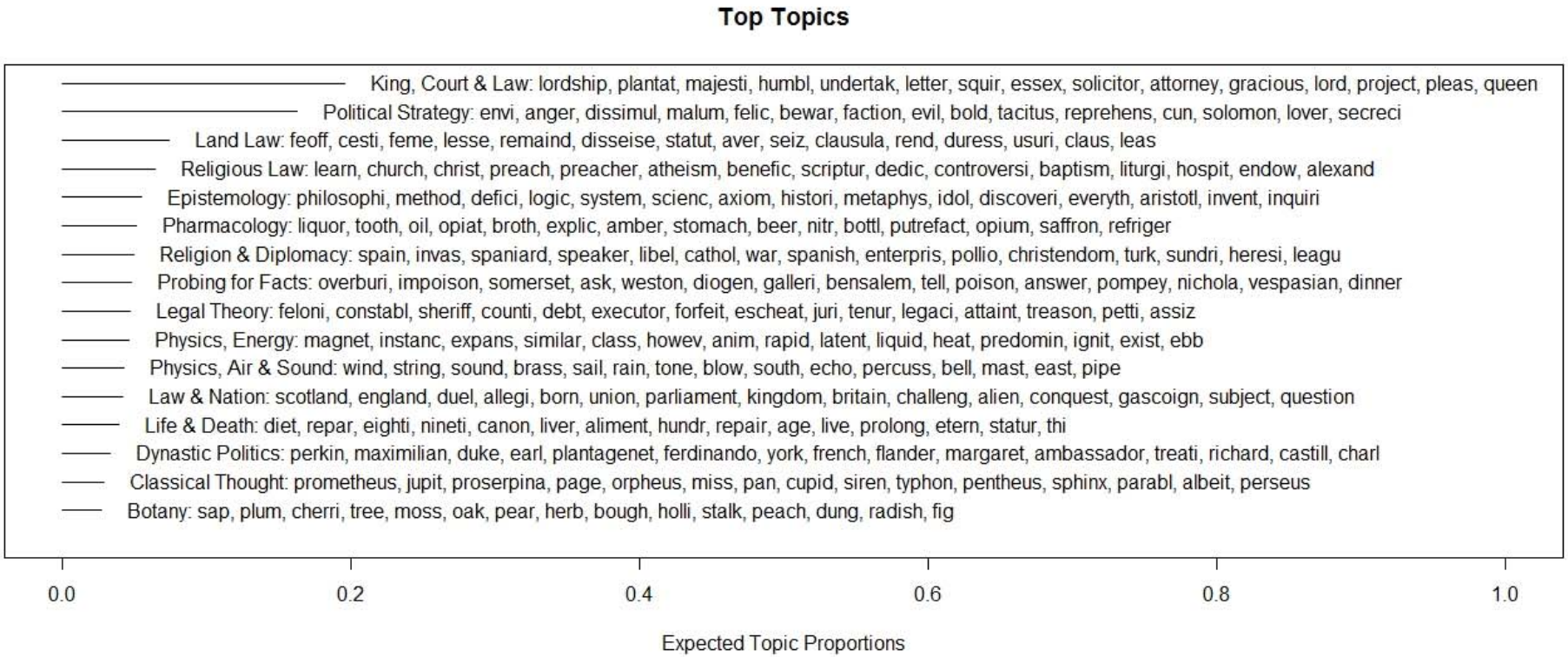
15. Life & Death

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

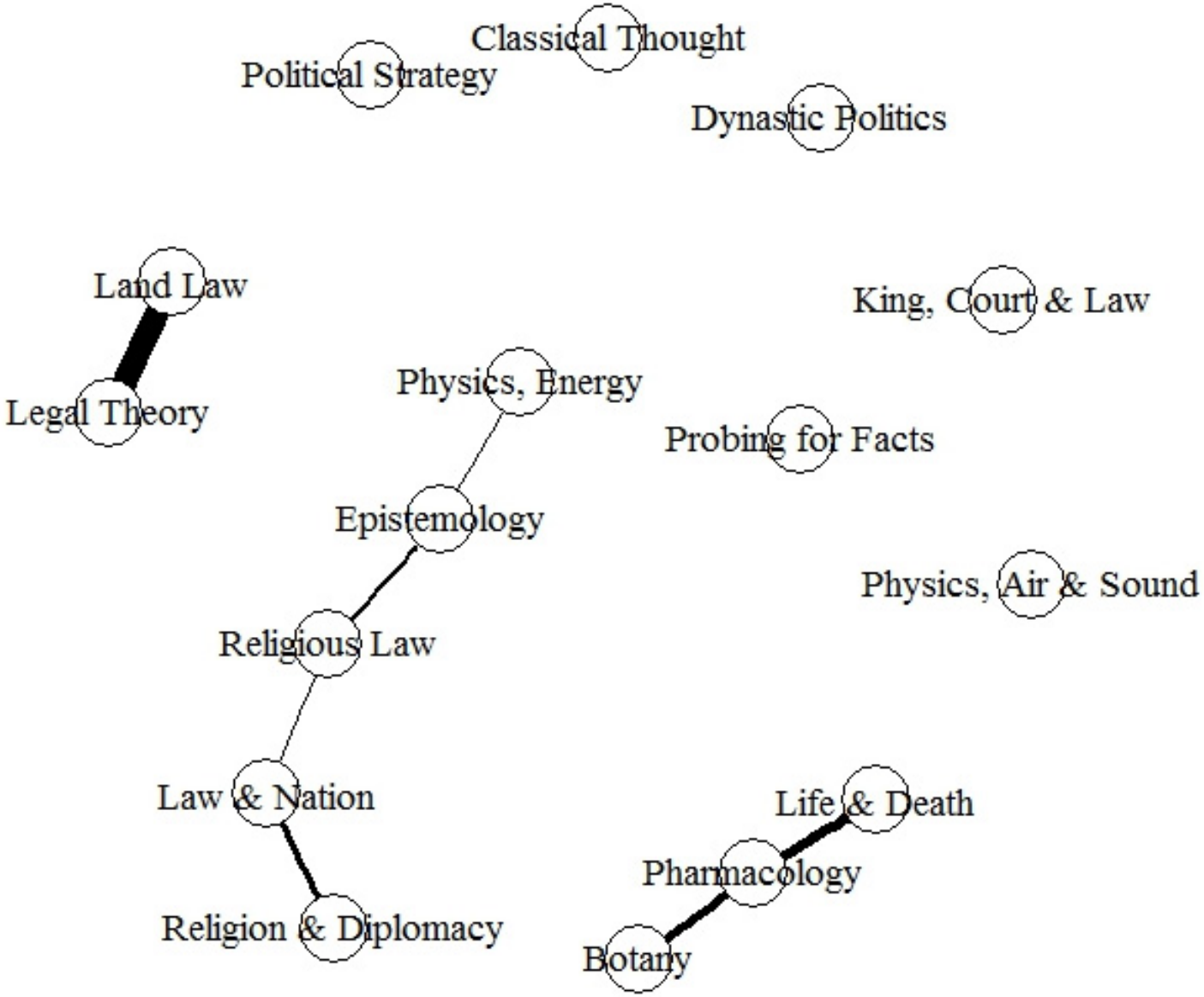
Highest Prob: bodi, spirit, will, water, part, may, make, upon, see, caus, also, heat, air, thing, great, therefor, like, take, much, use, man, natur, time, motion, cold, put, littl, good, one, come
FREX: liquor, tooth, oil, opiat, broth, explic, amber, stomach, beer, nitr, bottl, putrefact, opium, saffron, 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



For each topic, the figure displays 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.

Figure 2: Positive Topic Correlations



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

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

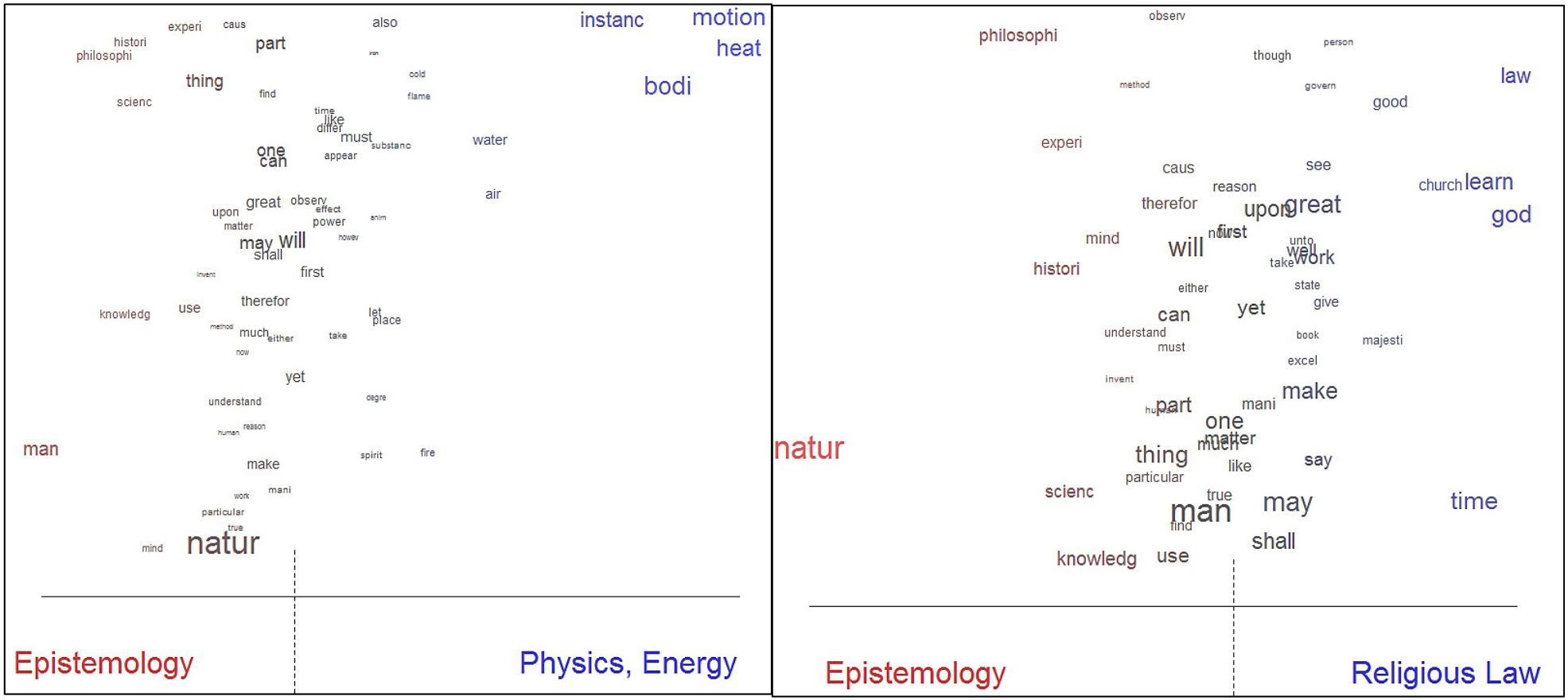


Figure 4: Vocabulary Use Across Topic Pairs: Epistemology vs. Botany and Legal Theory

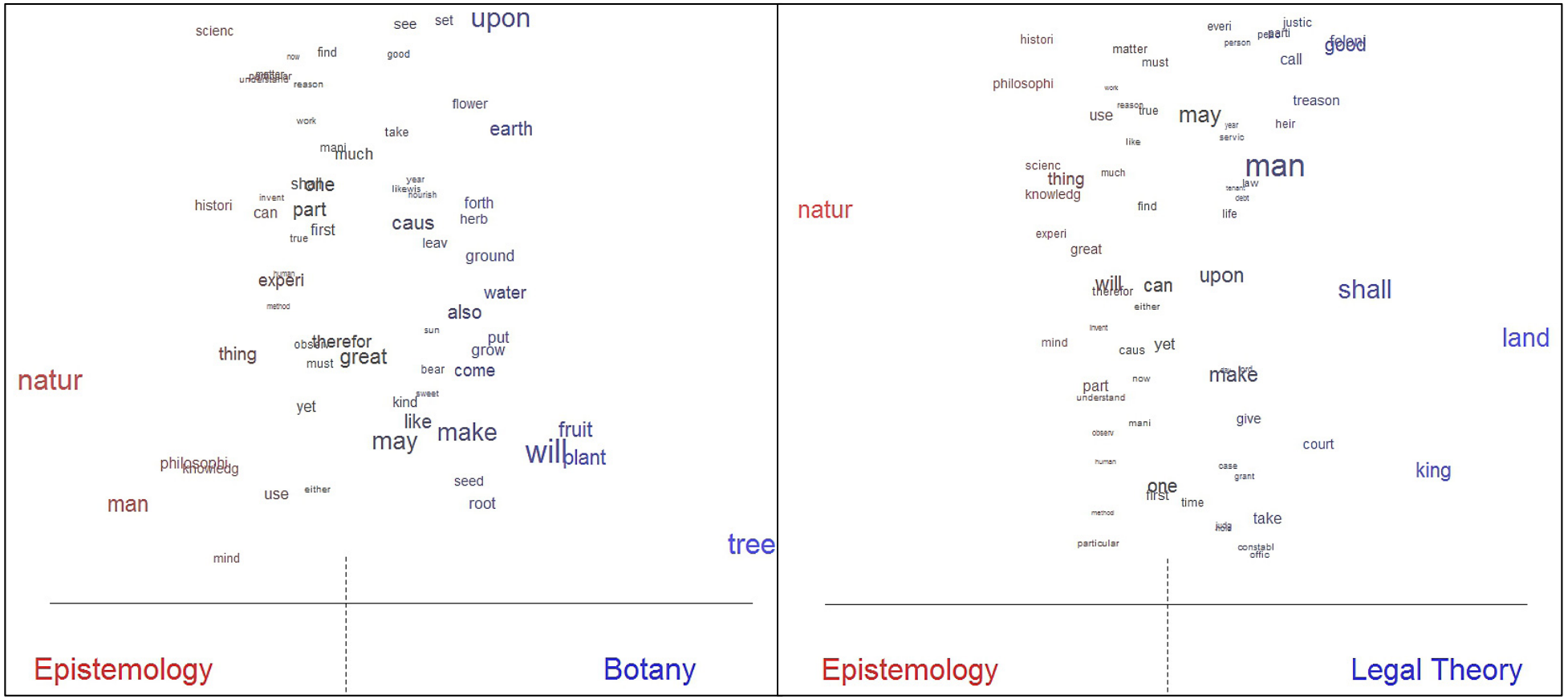


Figure 5: Vocabulary Use Across Topic Pairs: Probing for Facts vs. Physics, Energy and Religious Law

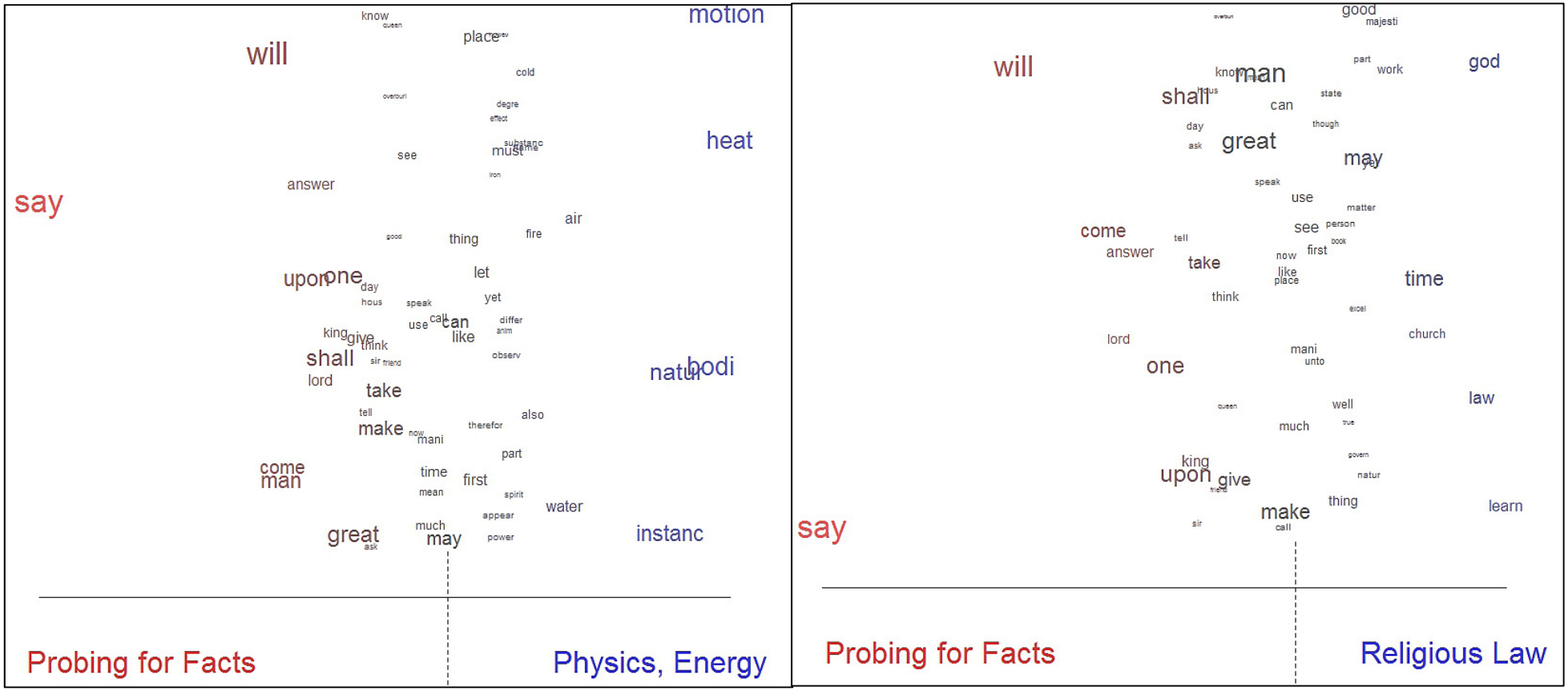


Figure 6: Vocabulary Use Across Topic Pairs: Probing for Facts vs. Botany and Legal Theory

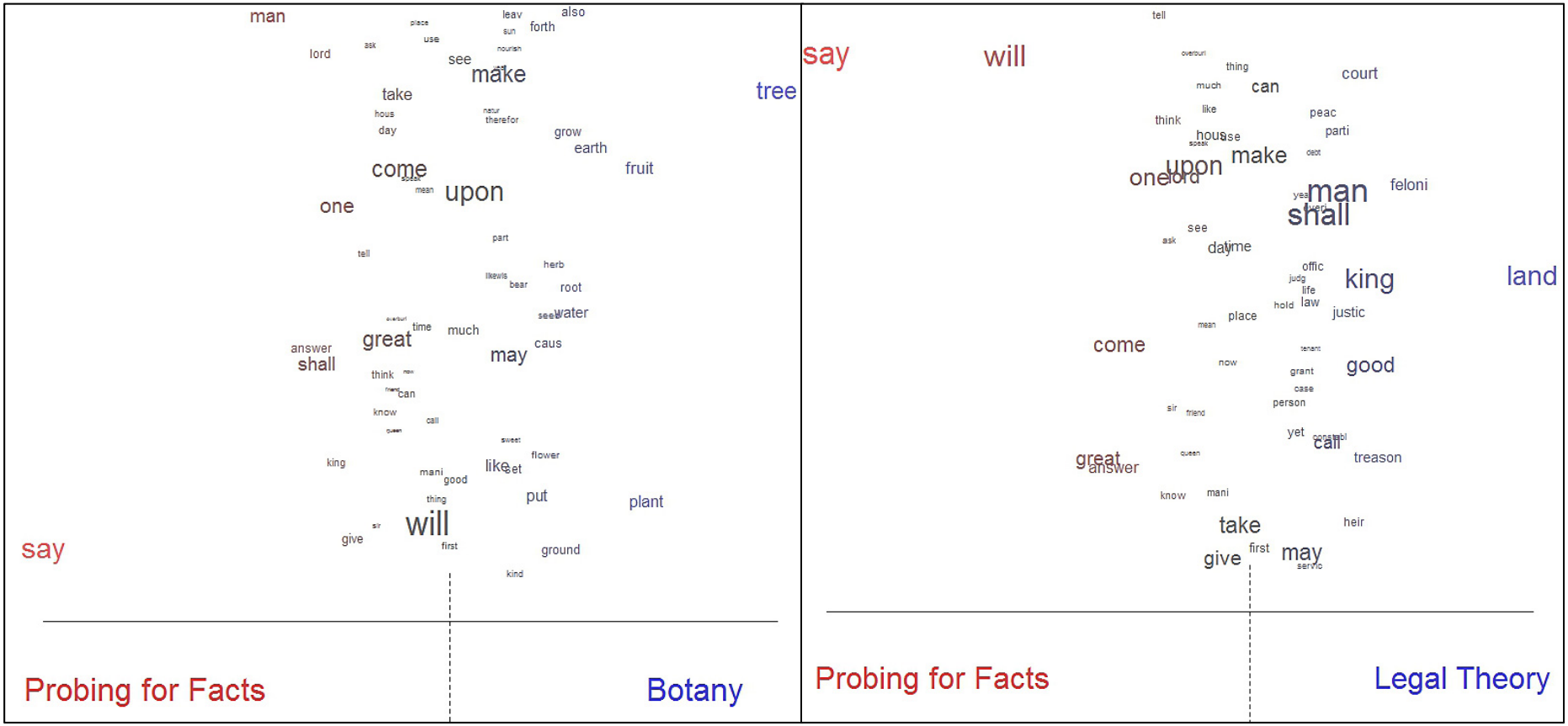


Figure 7: Vocabulary Use Across Topic Pairs: Epistemology vs. Probing for Facts

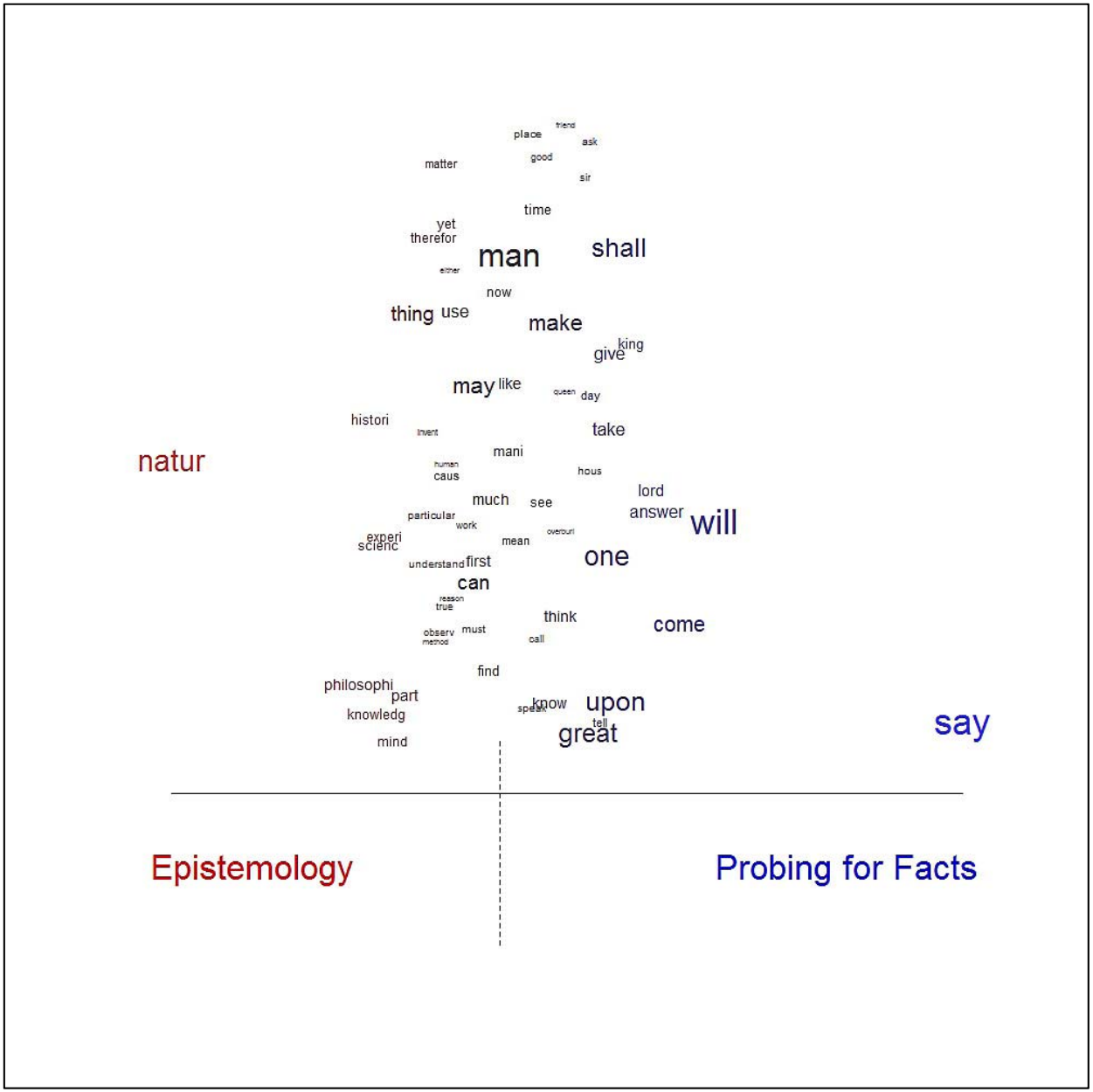
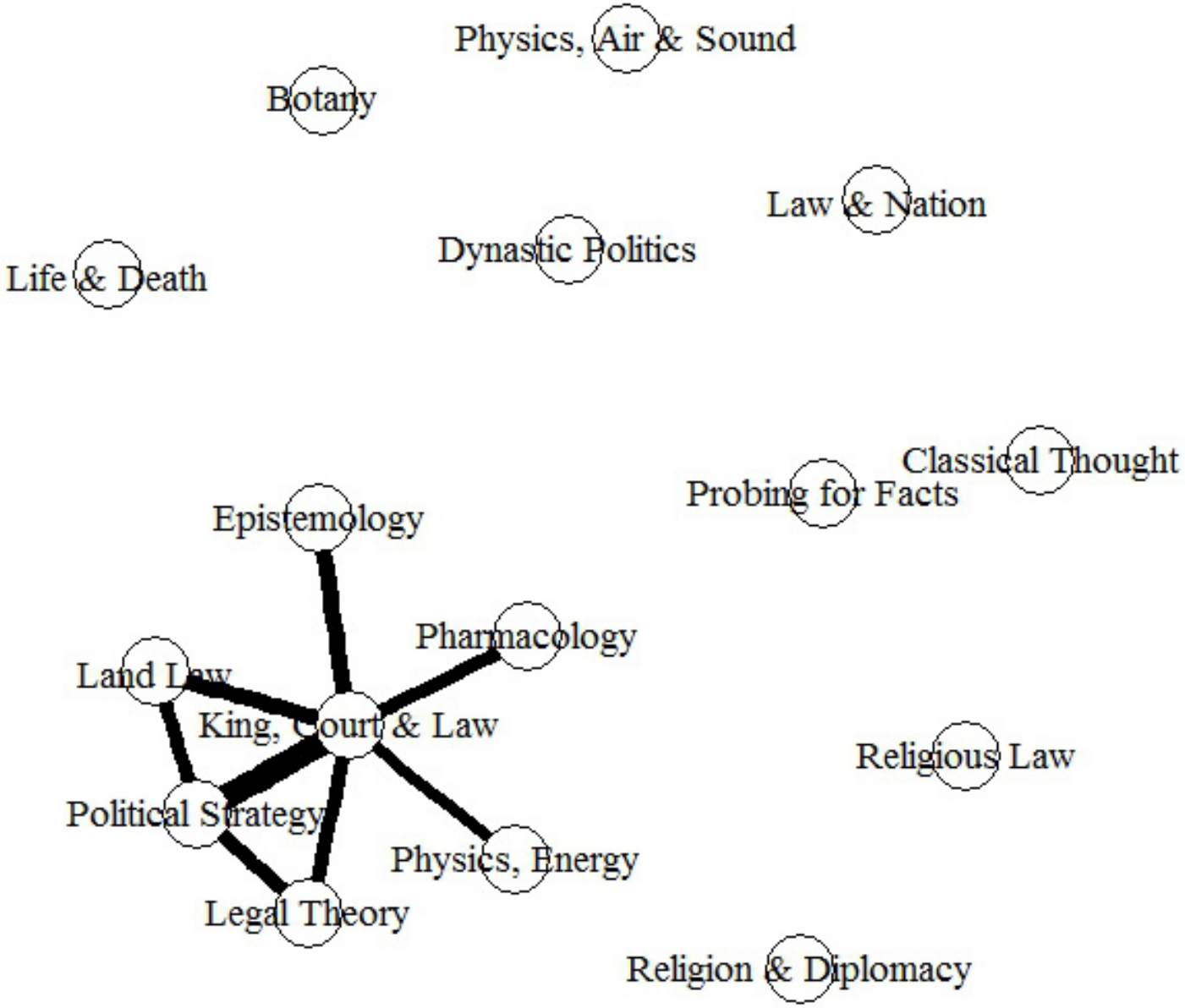
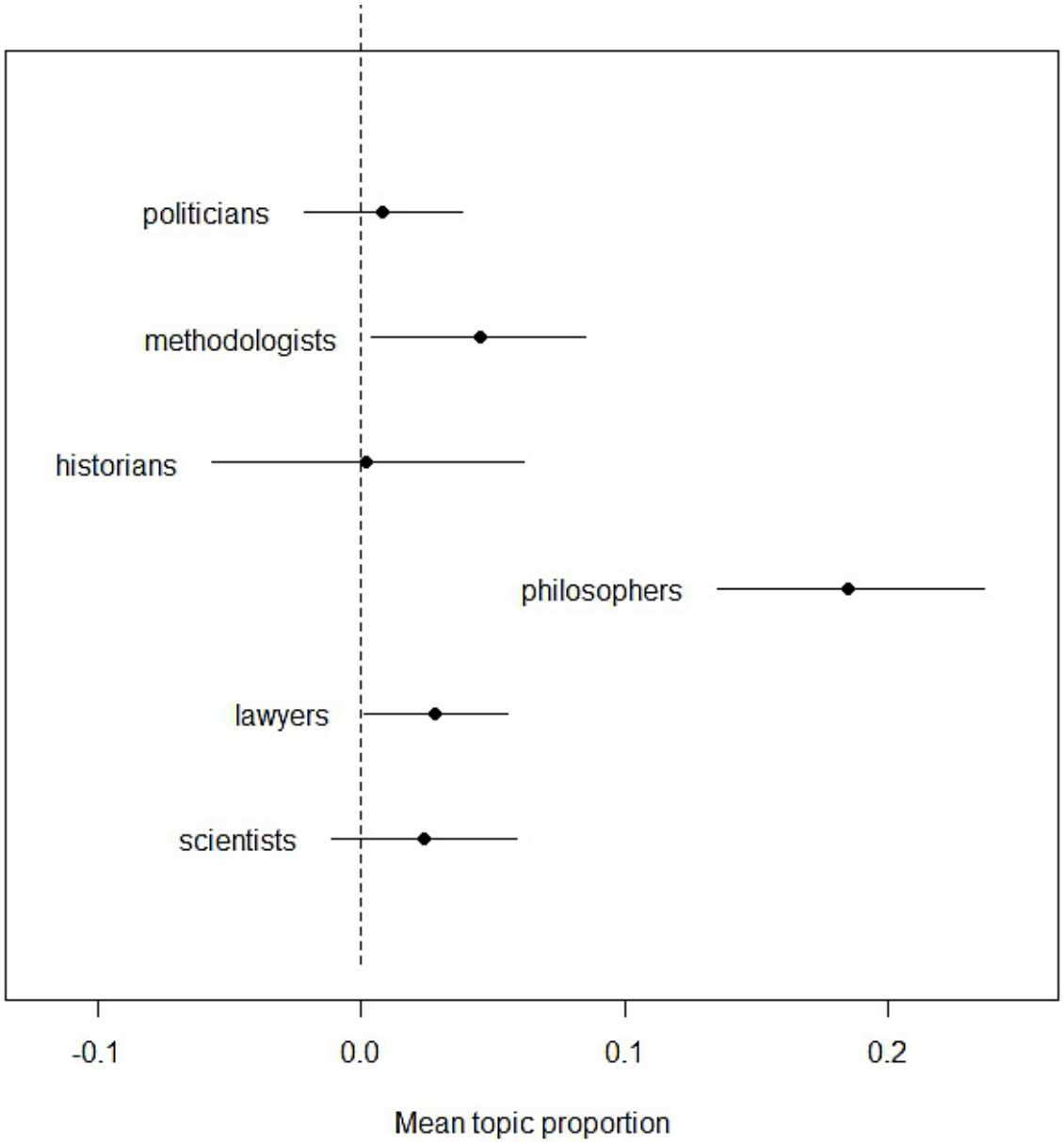


Figure 8: Negative Topic Correlations



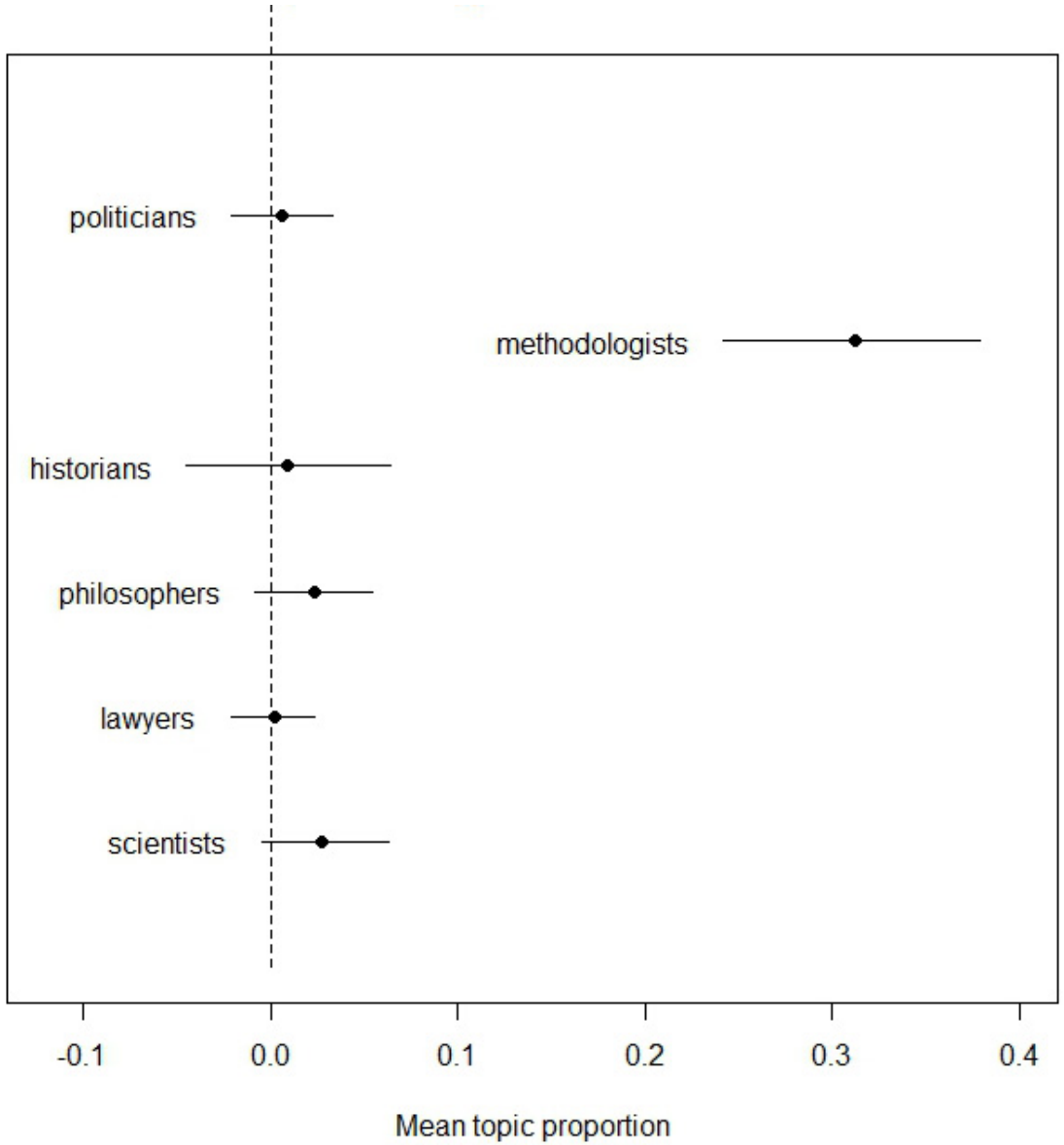
Note: The relative thickness of links reflects the magnitude of (negative) correlations between respective topic pairs.

Figure 9: The Effect of Audience on Prevalence of Probing for Facts



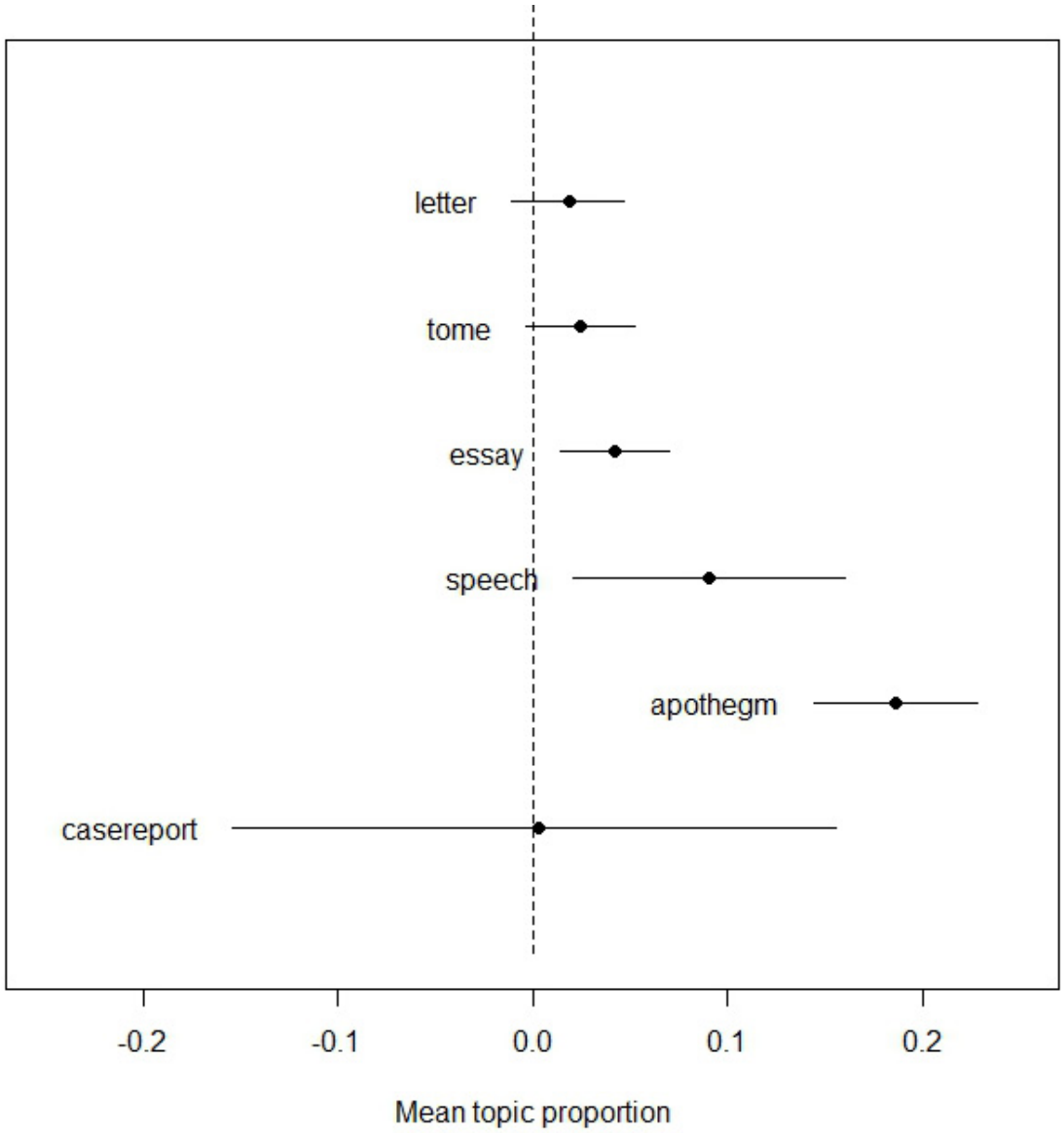
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Probing for Facts for each type of audience.

Figure 10: The Effect of Audience on Prevalence of Epistemology



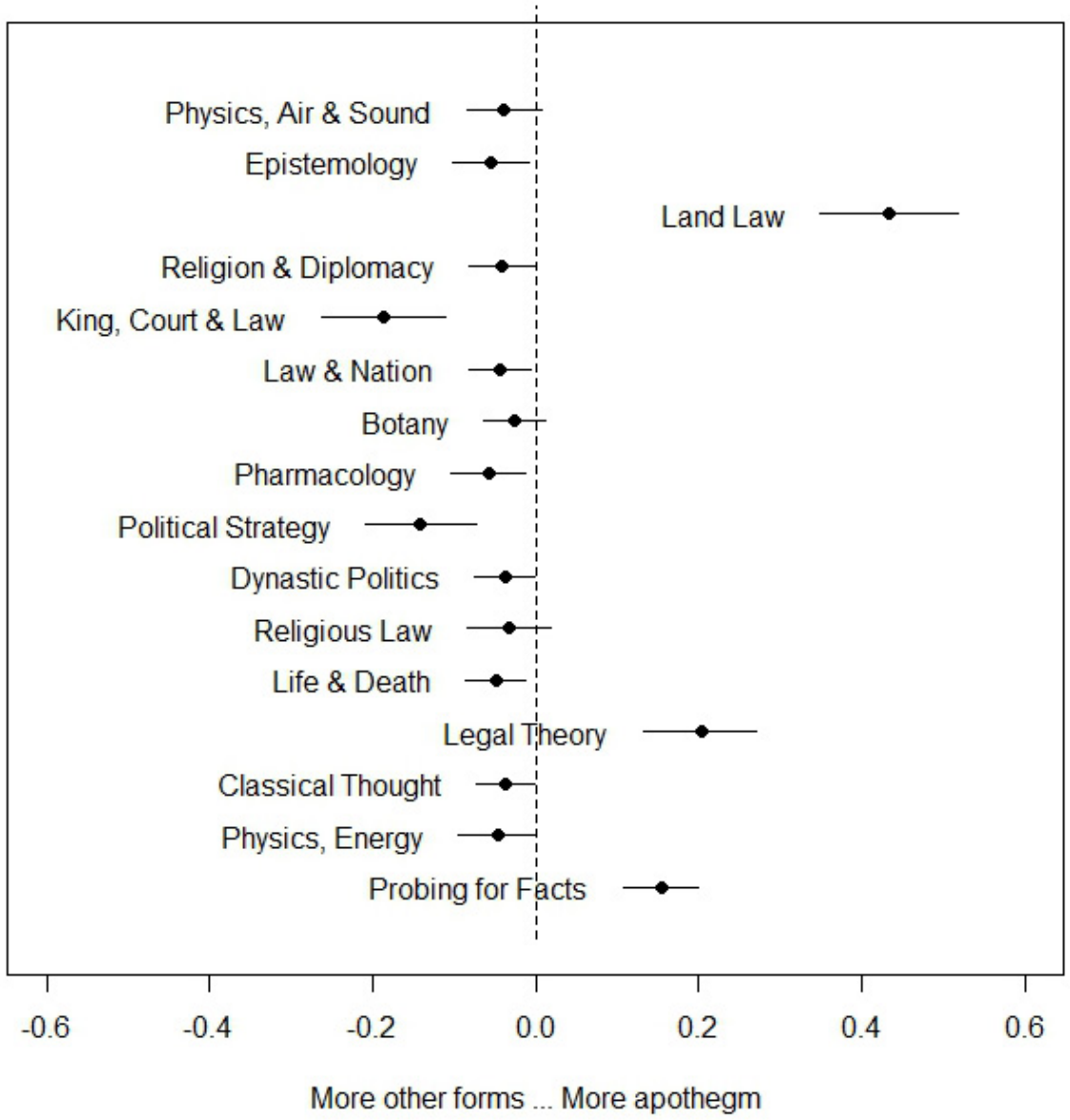
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Epistemology for each type of audience.

Figure 11: The Effect of Writing Form on Prevalence of Probing for Facts



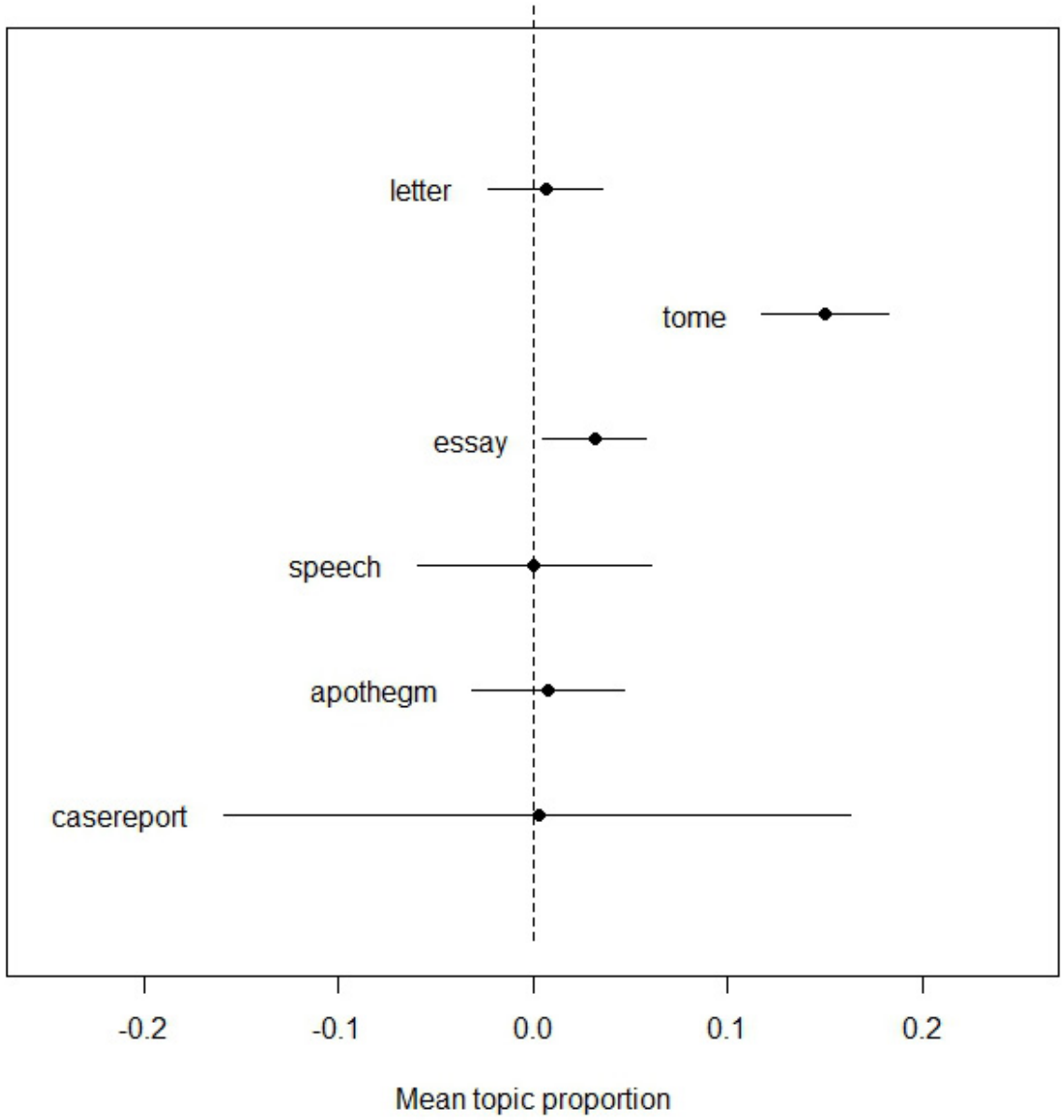
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Probing for Facts for each type of writing form.

Figure 12: The Effect of Writing Form, Apothegm



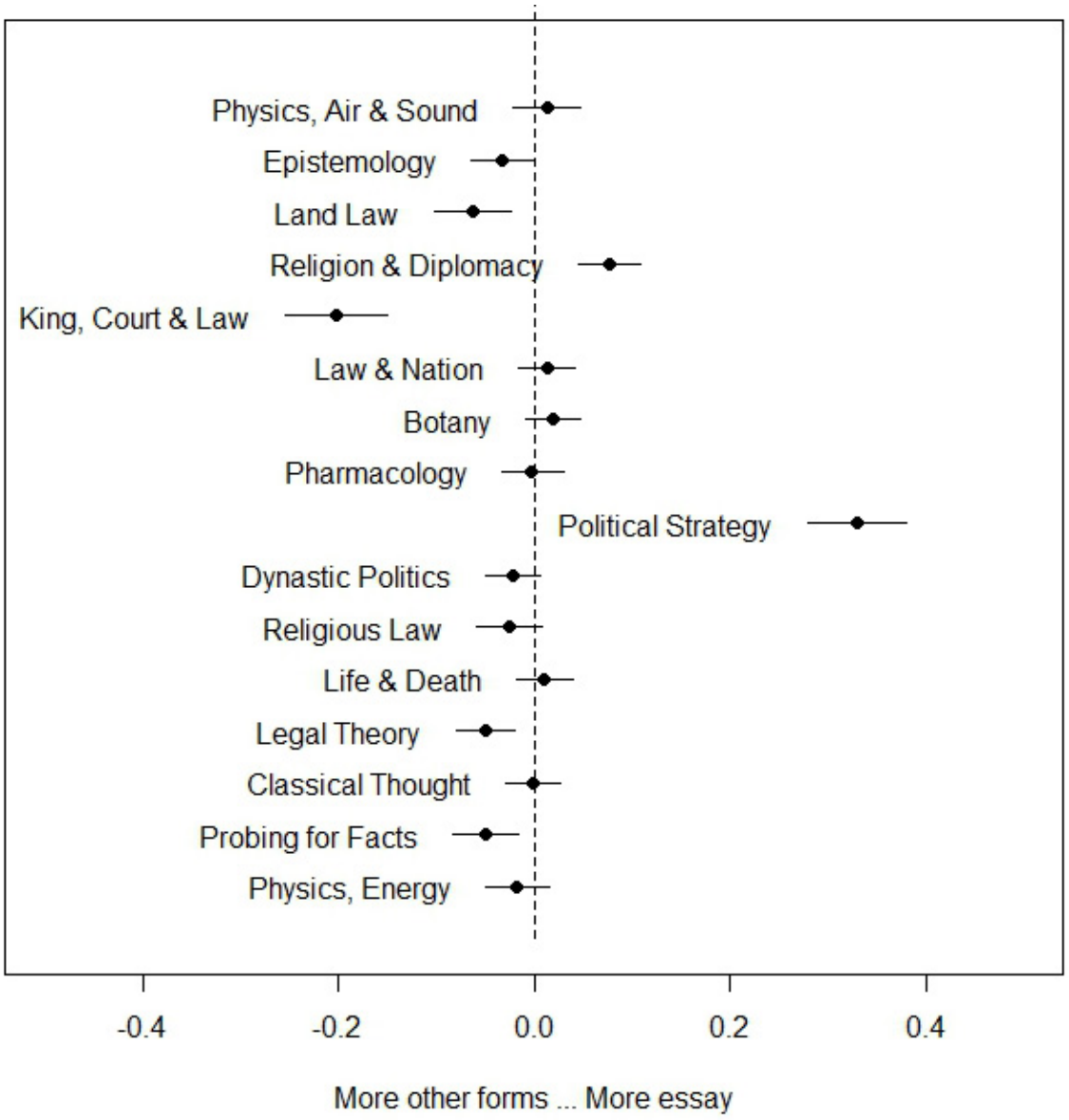
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for apothegms compared to all other document types.

Figure 13: The Effect of Writing Form on Prevalence of Epistemology



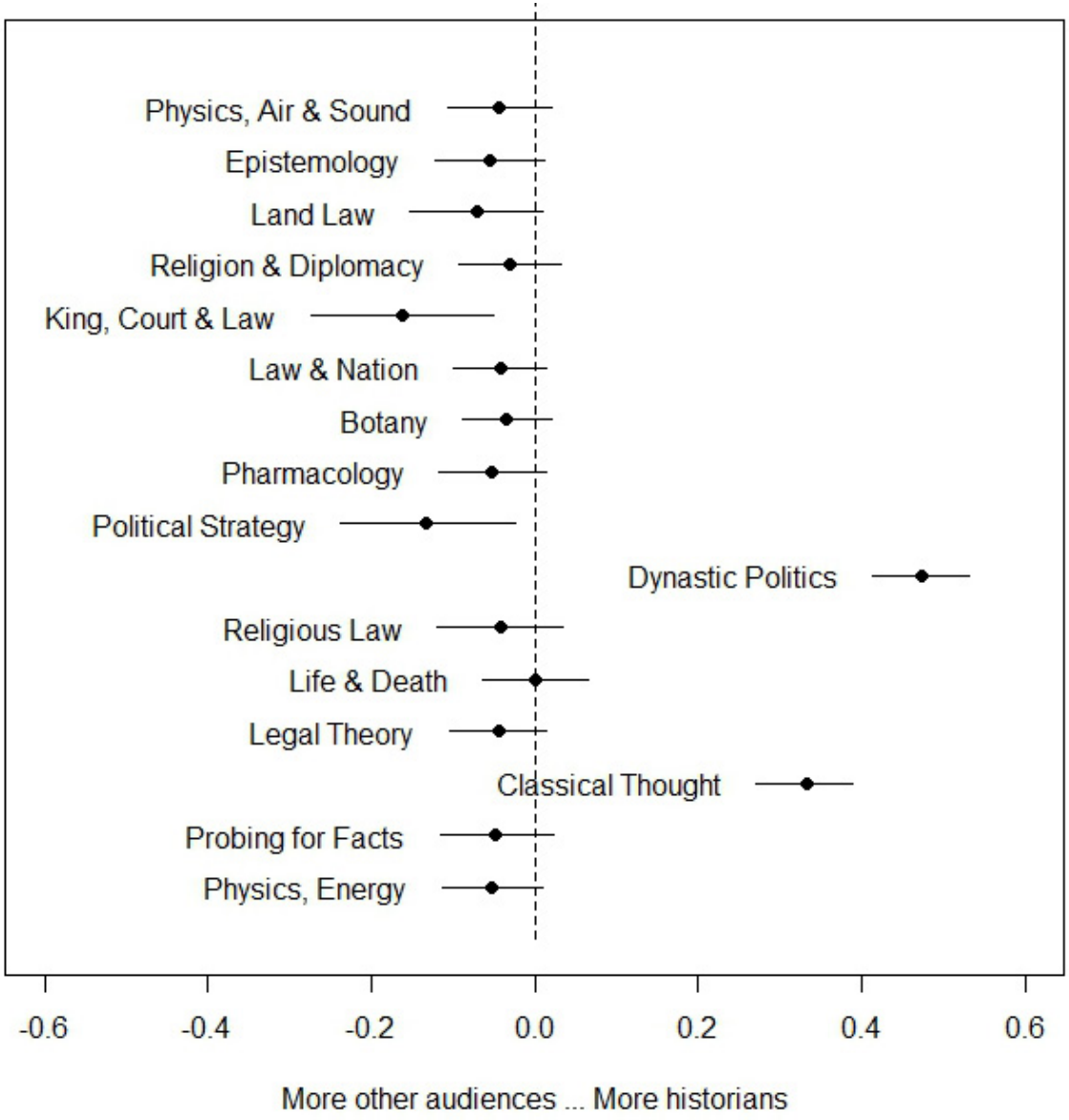
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Epistemology for each type of writing form.

Figure 14: The Effect of Writing Form, Essay



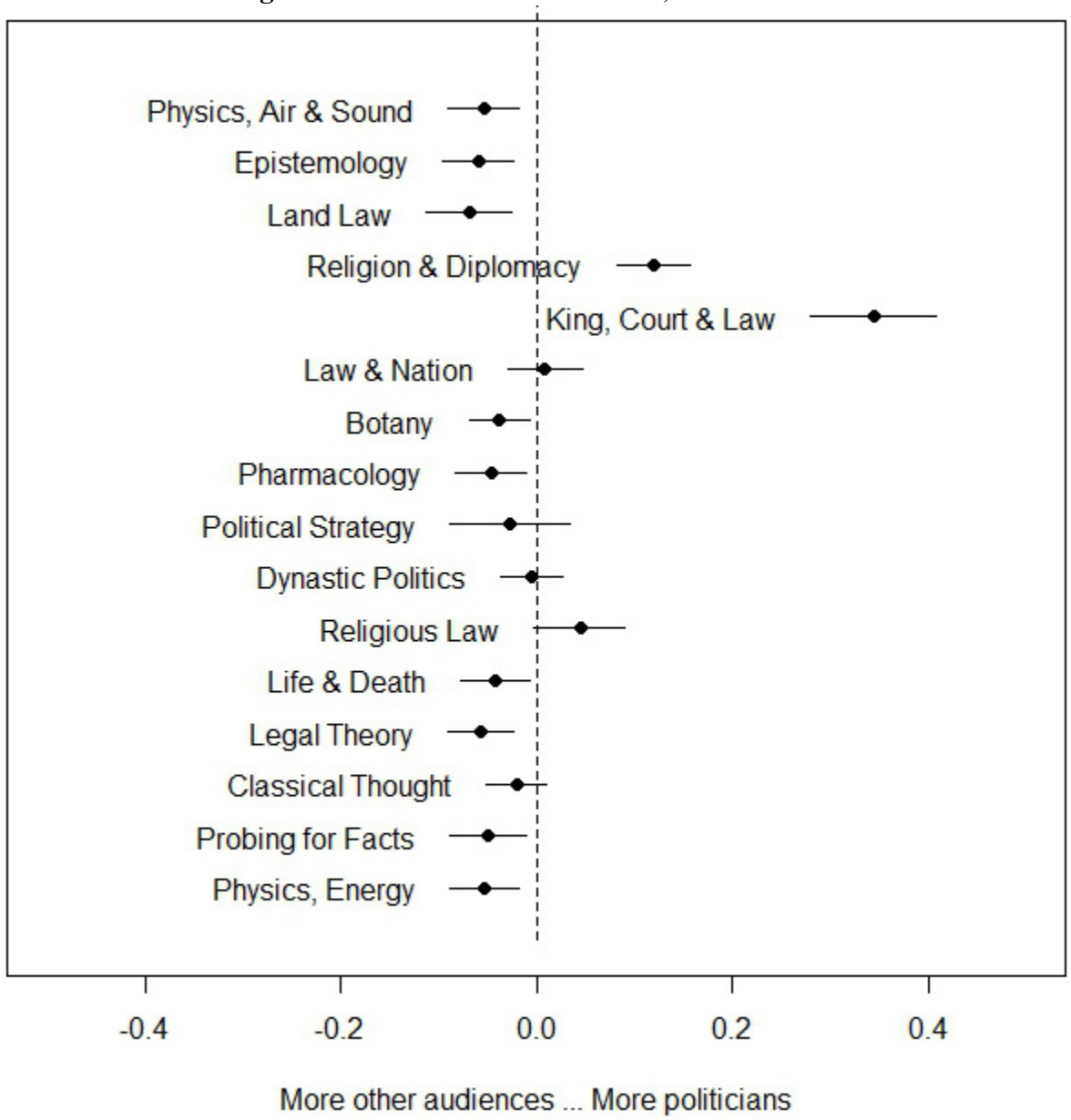
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for essays compared to all other document types.

Figure 15: The Effect of Audience, Historians



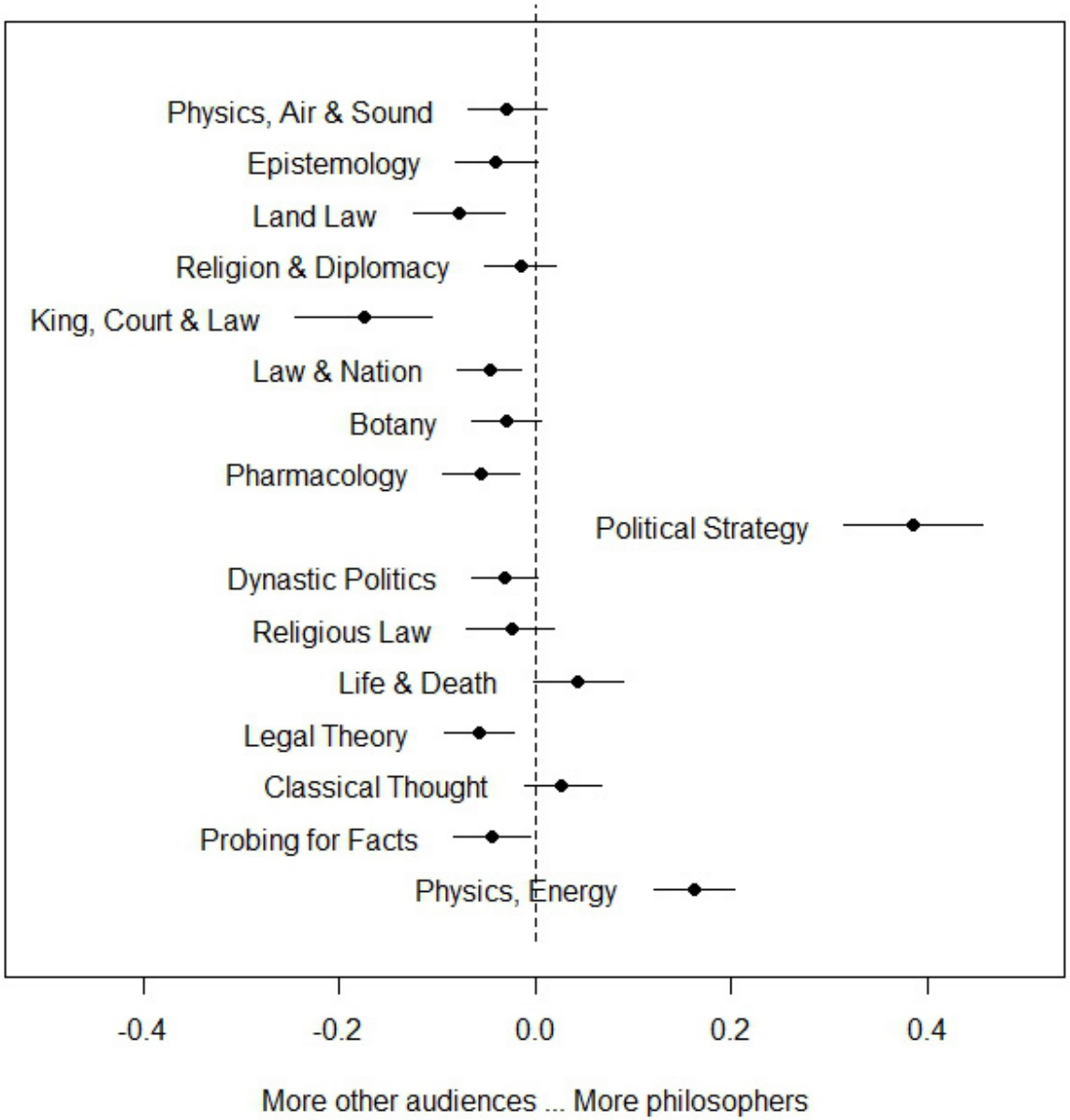
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which historians are the intended audience compared to works addressed to all other audiences.

Figure 16: The Effect of Audience, Politicians



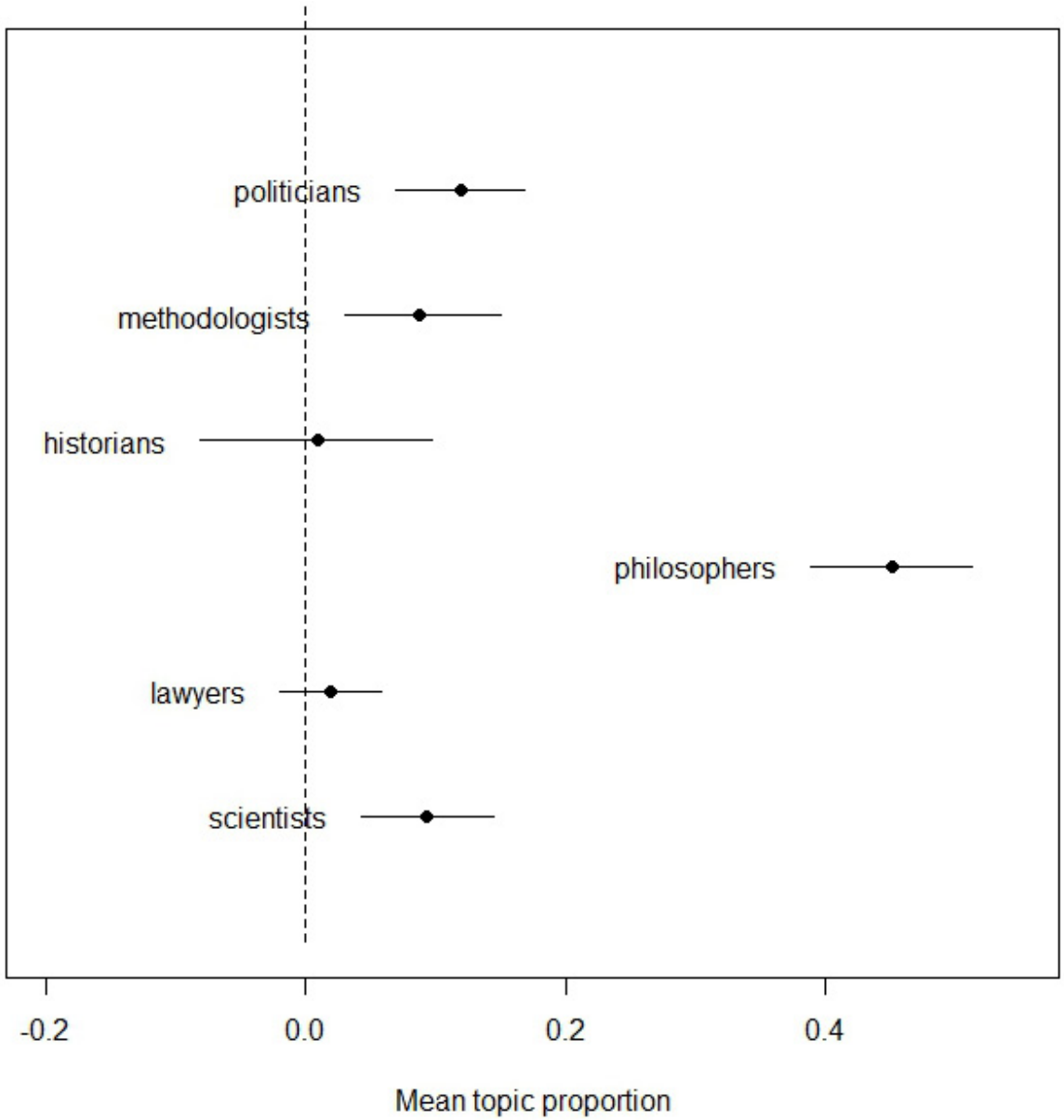
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which politicians are the intended audience compared to works addressed to all other audiences.

Figure 17: The Effect of Audience, Philosophers



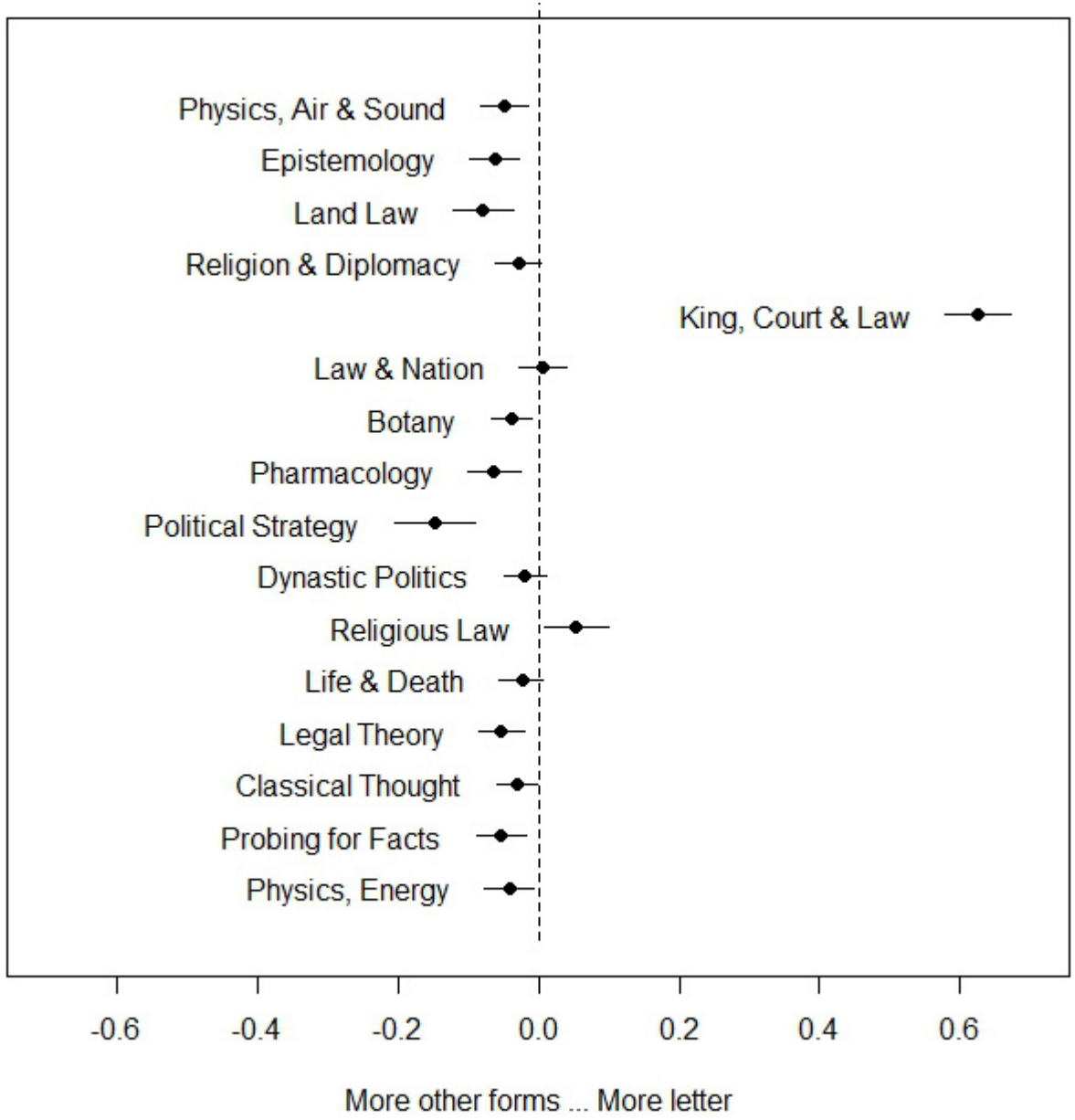
Note: This figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which philosophers are the intended audience compared to works addressed to all other audiences.

Figure 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.

Figure 19: 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.