Journal of Comparative Economics 47 (2019) 111–135

Contents lists available at ScienceDirect

Journal of Comparative Economics

journal homepage: www.elsevier.com/locate/jce

Toward understanding 17th century English culture: A structural topic model of Francis Bacon's ideas

Peter Grajzl[a,b], Peter Murrell[c,*]

[a] Department of Economics, The Williams School of Commerce, Economics, and Politics, Washington and Lee University, Lexington, VA 24450, United States
[b] CESifo, Munich, Germany
[c] Department of Economics, University of Maryland, College Park, MD 20742, United States

ARTICLE INFO

Keywords:
Francis Bacon
Culture
England
Law
Knowledge
Scientific method

JEL classifications:
B31
Z10
C55
N73
K10
P10

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 of a cultural paradigm that spurred modern economic development. Bacon's works are the data in an estimation of a structural topic model, a recently developed 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. The utilitarian promise of science and the centralized organization of the scientific quest, embraced by Bacon's followers, were not emphasized by him. Using strategic communication, Bacon facilitated reception of his scientific methodology, targeted influential groups, and finessed powerful opponents. We provide the first quantitative evidence that the genesis of Bacon's epistemology lies in his experience in the common-law. Combining our findings with accepted arguments in the existing literature, we suggest that the effects of common-law culture can help explain the coincidence of early political and economic development in England.

> 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

Arguably the fundamental question of comparative economics is why some countries develop and why other countries fail to do so. In the economics literature on this question, the early development of England provides a paradigmatic example, perhaps the

This is a revised version of a paper previously entitled "A Structural Topic Model of the Features and the Cultural Origins of Bacon's Ideas". We are indebted to three anonymous reviewers and especially the editor, Timur Kuran, for insightful comments that have led to significant revisions in this paper. For further helpful remarks and suggestions we thank Michael Anderson, Nona Karalashvili, Richard Nielsen, Martin Schmidt, Dustin Tingley, and participants at the workshop on Empirical Legal Studies at the University of Amsterdam and at the 22nd annual conference of the Society for Institutional and Organizational Economics in Montreal.

* Corresponding author.
E-mail addresses: grajzlp@wlu.edu (P. Grajzl), murrell@econ.umd.edu (P. Murrell).

https://doi.org/10.1016/j.jce.2018.10.004
Received 23 October 2017; Received in revised form 20 September 2018; Accepted 4 October 2018
Available online 23 October 2018
0147-5967/ © 2018 Association for Comparative Economic Studies. Published by Elsevier Inc. All rights reserved.

single most important data point (North and Weingast, 1989; Acemoglu and Robinson, 2012; Mokyr, 2016). The most popular causal interpretation of England's unique early success emphasizes the effect of institutions after the Glorious Revolution of 1688. But this interpretation has two weaknesses. First, supporting evidence is scant, especially quantitative evidence (Murrell, 2017). Second, this interpretation offers no explanation of why England was unique at that time.

A different explanation of England's development might invoke an alternative approach to addressing the fundamental question of comparative economics—a distinctive culture. Mokyr (2016) offers such an alternative approach, pointing to the influence of a new "culture of growth". This culture stressed the possibilities of scientific advance and invention, encouraging their application to economic activities, thereby stimulating technological progress and economic growth. Yet despite its appeal, the emphasis on the culture of growth does not quite get to the heart of an explanation of England's rise. In particular, given that the culture of growth is really an 18th century Western European phenomenon, this cultural argument does not directly address the unique precocity of England. If the culture of growth is key, why was England comparatively early in development? One possible answer suggests itself immediately within the confines of Mokyr's argument. Mokyr (2016) identifies Francis Bacon (1561–1626), the preeminent English philosopher, statesman, scientist, jurist, and author, as providing a unique stimulus to the emergence of the culture of growth. Indeed, the central ingredient of the culture-of-growth was 'Baconianism', an amorphous set of philosophical and methodological principles advocated by those who adopted and built on Bacon's ideas.

Yet a focus on Bacon immediately leads one to ask: What exactly were his core ideas? And where did those ideas come from? For two important reasons, answering these questions is a key first step in any effort to use a culture-of-growth explanation for England's early economic rise. First, Baconianism and the culture of growth were fully solidified only in the 18th century (Mokyr, 2016, 2005). Thus, to the extent that elements of Baconianism were instrumental in England's rise, it is Bacon's ideas, dating from the late 16th and early 17th century, that would have been particularly important at the onset of England's development, not the 18th century interpretations of Bacon. Second, an understanding of the genesis of Bacon's ideas sheds light on whether those ideas were original with him, or whether he was simply an influential mouthpiece for a culture that already existed. Identifying a pre-existing culture as a crucial input to Bacon's thought, and subsequently into Baconianism, would lead to a very different interpretation of English history than concluding that Baconianism began to emerge *only* as a result of the contributions of the unique genius, Bacon, standing alone.

Methodologically, the focus on culture, and in particular the ideas of one person, raises the thorny question of how to bring to bear the favored tools of economics, quantitative empirics, to understand more about the characteristics and development of specific cultures within historical studies. In this paper, we show how new, machine-learning methods can provide the required methodology and thereby aid in understanding both the features and the origins of Bacon's ideas, a key component of 17th-century English culture. This distinguishes our contribution methodologically from that of the prevailing intellectual history on that period. But as we note in many places in the text below, we also differ from that literature in substantive conclusions about the nature and genesis of Bacon's thought.

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 (Blei et al., 2003; Blei, 2012; Roberts et al., 2014, 2016a). Topic models produce estimates of a small set of themes (the topics) that are present in an author's work. Applied to Bacon's corpus, the estimated topics identify his core ideas and, by omission, allow identification of those elements of Baconianism that were not in fact emphasized by him. This immediately provides an answer to the first of the questions posed above, on the difference between Bacon's own ideas and Baconianism. Our estimated topic model clearly pinpoints two themes that the intellectual history literature invariably places at the center of Bacon's ideas: the emphasis on probing for facts and the epistemology of generalizing from facts. Importantly, we show that ideas about the utilitarian promise of science and the central direction of science—often attributed to Bacon by his followers—are not prominent in Bacon's own work.

The topic model naturally produces estimates of the connections between the use of different topics, information that can help explain the genesis of Bacon's ideas. The estimates provide strong evidence that Bacon's methodological ideas had their origins in his common-law jurisprudence. Our paper provides the only quantitative evidence to date that bears on this issue, providing insight into the broader culture that existed at the beginning of England's early economic development. The finding about the legal origins of Bacon's methodological ideas is also directly relevant to contentious debates in the existing intellectual history literature on Bacon. As we demonstrate in an extensive literature survey in Appendix A, 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 legal background (Kocher, 1957; Wheeler, 1983; Simonds, 1986; Cardwell, 1990; Martin, 1992: 164–171; Shapiro, 2000: 107–112; Serjeantson, 2014: 701–704). Moreover, the conclusions of these scholars have not been widely accepted in the literature. Additionally, even among this small set of scholars, there is disagreement about which aspect of Bacon's legal background was most influential. In contrast, our methodology produces quantitative evidence in favor of the hypothesis that Bacon's ideas about scientific methodology emerged from his immersion in English common law. As we discuss in the conclusion, this finding has important implications for an understanding of 17th century English culture and for the interpretation of England's comparatively early economic rise.

To set the stage, Section 2 provides a primer on Bacon in the context of the time in which he lived. We offer an overview of key aspects of Bacon's personal and professional life, as well as introduce the powerful actors, social groups, and intellectual paradigms that shaped his actions.

Our formal analysis begins with the specification and estimation of a Structural Topic Model (STM; Roberts et al., 2014, 2016a), providing machine-learning-based, macroscopic, statistical evidence on the salient features of Bacon's work. STM is a recent innovation in topic modeling that integrates document-level data into the analysis. Given STM's general unfamiliarity to economists,

Section 3 details the generative model that underpins STM. With their tremendous scope and breadth, Bacon's writings are particularly suitable for topic modeling.[1] 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.) By providing a direct path to an author's own words and allowing the central themes in a corpus to be 'discovered' by an unsupervised machine-learning algorithm, the conclusions from topic-model estimates are divorced from the history of scholarly interpretation. Nevertheless, we emphasize that we do not argue that these techniques are superior to existing textual-analysis methods. Instead, we view machine-learning as one more, hitherto unexplored, method of obtaining insight into the nature and sources of cultural ideas.

Section 4 documents the process of building the comprehensive, machine-readable corpus of Bacon's major writings, comprising 282 documents and nearly 900,000 words. Additionally, we associated with each document the values of variables characterizing the form of Bacon's work (e.g. an essay, a letter etc.) and its targeted audience (e.g. philosophers, politicians etc.).

As we discuss in Section 5, 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. This is the way in which our unsupervised machine-learning algorithm addresses the question of which ideas Bacon emphasized (and implicitly which parts of Baconianism he did not).

In Section 6, we develop insights into the origins of the principal features of Bacon's work by examining the correlations across documents in the use of disparate topics and the commonalities between topics in the use of distinctive vocabulary. Our analysis demonstrates that Bacon's epistemological reasoning is strongly connected with 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 the first systematic quantitative evidence that the origins of Bacon's epistemological ideas lie in his common-law jurisprudence.

Consistent with the comparative social science literature that emphasizes the importance of strategic communication by influential actors (see, e.g., Kuran, 1995), Bacon has often been thought of as communicating his ideas strategically, taking into account multiple powerful constituencies and societal taboos. In Section 7, we make direct use of the defining element of STM, the integration into estimation of document-level information, to explore the way in which Bacon's emphases varied with intended audience and medium of communication. We find, for example, that Bacon emphasized his methodological ideas when addressing groups that he viewed as particularly effective messengers of his thought, choosing to disseminate those ideas in the form that made them most widely available, print. However, he de-emphasized his methodological ideas when addressing monarchs and other politicians. When communicating with politicians, Bacon instead emphasized matters of law that pertained to high politics. Moreover, he communicated those ideas primarily in private letters, carefully shielding them from public discourse. Perhaps, we have found one reason why Bacon's methodological ideas have usually been interpreted as applying more immediately to science than social science or governance, even though he emphasized the universal applicability of those ideas.

Our methodology and our findings complement and supplement the existing literature in a number of additional ways. Apart from its specific implications for the interpretation of the rise of England, to which we return in the conclusion, our paper has broad relevance for the literature on comparative economic development. We contribute to the study of culture (e.g., Guisoet al., 2006; Alesina and Giuliano, 2015), and particularly to the strand of research on culture that has focused on identifying culture's deep historical roots.[2] Our paper shows how machine-learning can be used to uncover broad cultural ideas and to estimate the connections between them, providing insights into their historical origins. In particular, we provide a new view within economics on why the common-law was important, focusing on its cultural influence rather than on the shaping of institutional arrangements (see, e.g., Djankov et al., 2003; Beck et al., 2003; La Porta et al., 2008; Guerriero, 2016). Additionally, by isolating the source of Bacon's ideas, we provide perspective on the question of how important the ideas and actions of influential individuals are in shaping fundamental social 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).

We also contribute by applying machine-learning in economics and in comparative economics in particular. Our paper is unique in economics in applying such methods to historical-cultural questions. It lies in a class of work that is beginning to appear in other social sciences—quantitative analyses of culture produced by examining the recorded ideas of important social actors (see, e.g., Blaydes et al., 2018; Lucas et al., 2015: Section 4; Roberts et al., 2014). Within economics, topic modeling has been employed only very recently.[3] The Latent Dirichlet Allocation (LDA; Blei et al., 2003; Blei, 2012) approach has been applied in understanding monetary policymaking (Fligstein et al., 2014; Hansen and McMahon, 2016; Hansen et al., 2018), financial market performance (Larsen and Thorsrud, 2015; Huang et al., 2016), and corporate behavior (Bandiera et al., 2017; Bellstam et al., 2016). We move beyond LDA, and use Structural Topic Modeling (Roberts et al., 2014, 2016a), integrating document-level data into our analysis.

Lastly, this paper is relevant to the voluminous intellectual history literature on Bacon and Baconianism. The interest of scholars in the features of Bacon's ideas, and their connectedness, has inspired a voluminous literature in the history of science, political philosophy, and intellectual history more broadly. Within this literature there are many contentious debates on the nature of Bacon's

[1] 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.

[2] On the importance of culture see, for example, Barro and McCleary (2003), Licht et al. (2007), Tabellini (2008), 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).

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

ideas and the source of those ideas. In the details of the paper, we comment on how our results are relevant to those debates, which hypotheses our paper rejects, and which it accepts.

2. Bacon in his times

Francis Bacon (1561–1626) lived during an age when English culture flowered, personally observing the first performances of Shakespeare's plays and the publication of that great literary and religious work, the King James Bible. Parliament was beginning to erode the power of the monarch, and the legal system was edging towards independence. Across Europe, the ideas that later constituted the scientific revolution were emerging (Wootton, 2015). Contributing to all of these areas—culture, politics, law, and science—Bacon took "all knowledge to be [his] province", coming as close as anybody ever has to fulfilling that promise.

Bacon was born into the elite. Under Queen Elizabeth (1558–1603), Nicholas Bacon, his father, became Lord Chancellor, the head of the legal system. He introduced the precocious young Francis to the Queen. But Nicholas died when Francis was eighteen and left him without any source of income. Family, then, seems to have conditioned Bacon's professional ambitions for the rest of his life: a desperate search for status and wealth, always seeing service to the monarch as the obvious route.

By the time of his father's death in 1579, Bacon had graduated from Cambridge, begun his legal studies, and traveled in Europe. At Cambridge and in Europe, he received grounding in the renaissance humanist tradition, with its emphasis on the practical use of knowledge especially within politics, and in the near-dominant philosophical approach to science, Aristotelianism, with its emphasis on an unchanging universe to be understood using deductions from first principles. But, being the youngest of six children in an era of primogeniture, Bacon needed a profession. He took up law, completing an apprenticeship at England's 'third university', London's Inns of Court, where lawyers studied, practiced, and lived. There he learned the rigors of legal rhetoric and absorbed the enormous stock of case material. But he would have also imbibed an evolutionary common-law outlook that diverged greatly from Aristotelianism. This approach emphasized that law could be found through observation of local customs and that jurisprudence could rely on experts sifting through past cases, distilling common elements (Grajzl and Murrell, 2016).

Bacon succeeded brilliantly as a legal scholar in Elizabethan England, but foundered in seeking higher office. Elizabeth distrusted him and appointed him only to lower-level positions. Perhaps because of his disappointments in rising to higher office, as the 16th century closed Bacon turned to the avocation that secured his place in history, the study of natural philosophy (i.e., science) and particularly to deliberation on how scientific knowledge came to be established.[4]

In seeing the need for a new epistemology of science, Bacon was far ahead of his time since science itself was only beginning to stir. When he began, no major work of experimental science had been published for centuries, the notion of discovery was just entering the language, and Galileo had not yet revolutionized the understanding of planetary motion (Wootton, 2015). Nevertheless, earlier experiences in the law had led Bacon to see the importance of finding fundamental principles (his *Maxims*) within the inchoate body of cases and customs that constituted the common law (Kocher, 1957). Moreover, in the common culture of London, experimentation in action, if not words, was common-place (Harkness, 2007).

Grappling with the intellectual challenge of developing a new epistemological paradigm, Bacon had to confront the prevailing Aristotelian philosophy, which emphasized that all science ought to be deduced from generally accepted first principles. Bacon saw that this approach was not productive in a world that was generating new discoveries (Wootton, 2015). But Bacon's world was also one where the very idea of a fact, in its modern sense, was only beginning to appear. Consequently, the pursuit of new facts was a goal that was far from common-place even amongst those practitioners now viewed as the most forward-looking thinkers of that era. Even less common were ideas about how to sift facts systematically when searching for scientific laws and generalizations of higher order. Working in a prevailing culture that thought of the scientific method as contemplative understanding of an unchanging world, Bacon's genius was to see the need for a radical departure from this prevailing paradigm, a new epistemology that emphasized the discovery of facts and the process of generating reliable knowledge from them. This was the task he set himself: "Bacon was the first person to try to systematize the idea of a knowledge that would make constant progress" (Wootton, 2015: 1691).

Even though he viewed his epistemology as universally applicable (Bacon, 1859), Bacon was fully aware of the heterogeneity of audiences that he would address with his proposed scientific method. In slaying the great dragon of his day, Aristotelian scientific method, Bacon felt resistance from the great academic centers, Cambridge and Oxford, redoubts of Aristotle and civil law (Sgarbi, 2012: 32). But he would also obtain support from those, perhaps a majority of the educated classes in Bacon's London, who had been steeped in the philosophy and methodology taught at the Inns. Attacking the old intellectual paradigm was surely a pursuit from which Bacon derived expressive utility.

Bacon understood the importance of carefully choosing methods of communication (Peltonen, 1992; Zagorin, 1998: 180). The printing press had not yet become standard, but it had decreased the cost of public dissemination. However, to use print was to be a professional author rather than an aristocrat, amongst whom broad circulation was still viewed as "degrading" (Peltonen, 1996a: 10). The debate on printing was especially vigorous within the legal profession. Those pitching arguments in favor of broader dissemination stood against those favoring a closed process emphasizing the comparative advantage of lawyers in interpreting existing materials (Ross, 1998). Bacon was generally on the side of the publicists, despite his aristocratic pretensions.

The importance of different audiences was also apparent in Bacon's work as a jurist and political appointee. When Elizabeth was succeeded by a foreign king, James I, in 1603, Bacon's legal and political career prospered: after securing appointment as Solicitor General and then Attorney General, he eventually rose to Lord Chancellor in 1618. But this placed Bacon in a position where his

[4] Vickers (1978: 24) dates the first development of Bacon's key ideas on scientific methodology to the period between 1603 and 1608.

public statements on law were in the service of a King who had no education in the common law and little appreciation for English legal customs and institutions (Coquillette, 1992: 104–105). As the King's lawyer, Bacon was duty-bound to present the King's position; failure would have resulted in loss of income, or even a head. At the same time, Bacon's rise through officialdom placed him in direct competition with Edward Coke, the most distinguished member of the increasingly powerful legal profession and Bacon's long-time personal rival: to call the story of their relationship a soap opera would give it too much dignity. Bacon would surely have placed winning this competition above any considerations of integrity (Baker, 2004: 257). Especially on law, therefore, analysis of Bacon's works must take into account the divergent interests of different impacted parties as well as a possible disconnect between Bacon's private convictions and public utterances.

Finally, as Galileo's circumstances exemplify, science in the 16th and 17th century had to navigate an environment dominated by religion. Certainly, religious issues were central within English political struggles at that time and discourse on religion was intertwined with discourse on politics, law, and science (Hill, 1997). For Bacon, the influence of his devoutly Calvinist mother seems to have conditioned his fervor in the pursuit of knowledge more than determined his doctrinal stance. If anything, he followed the pragmatic Anglicanism of his father (Zagorin, 1998:6), an easy choice given the predispositions of James I (Coquillette, 1992: 6). The main area in which Bacon's work verged on religious controversy was in his emphasis on a strict separation between religion and science, the former to focus on God's will and the latter on God's works (Gaukroger, 2001: 80; Zagorin, 1998: 48). While advancing such views did invite controversy with some theologians, such arguments were common enough as the 17th century began (Kocher, 1953).

3. Structural topic modeling

With rising computer power and the increasing availability of machine-readable texts, new approaches to the quantitative analysis of culture have become increasingly feasible by examining the recorded ideas of important social actors (see, e.g., Blaydes et al., 2018; Lucas et al., 2015: Section 4; Roberts et al., 2014). Focusing on Bacon, this paper provides an example of such analysis for comparative economics, using a topic model (Blei et al., 2003). As the name suggests, the key output of a topic model is a set of 'topics', that is, the principal features of a corpus as identified by an unsupervised machine-learning algorithm. The estimated topics identify the emphases within a corpus that might not be readily apparent to a reader of many disparate documents (Blei, 2012). Topic modeling is therefore ideally suited to identify ideas that are prominent in a particular corpus and reflect broader social culture.

The particular approach that we use is Structural Topic Modeling (STM), introduced by Roberts et al. (2014, 2016a). The key innovation of STM vis-à-vis earlier topic models is to use information on the documents themselves in the estimation process, rather than just the words in the documents. Thus, for example, an author's use of a cultural topic can be related to the type of audience that the author is addressing, leading to insights on which groups share a particular culture. In this spirit, recent work by comparative politics scholars has used STM, or related methods, to understand the timing and patterns of historic divergence between Christian and Islamic political thought (Blaydes et al., 2018), worldwide perceptions of the rise of China (Roberts et al., 2016a), communication strategies of Muslim clerics (Lucas et al., 2015: Section 4.1), and international views of the U.S. (Lucas et al., 2015: Section 4.2).

Falling under the umbrella of 'big data' analysis, topic models are machine-learning tools that use generative probabilistic models developed for quantitative investigation 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). These groups of words are the topics, formally conceptualized as probability distributions over vocabulary. 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 featuring a topic most prominently. However, the topics are purely a product of model estimation. They are not produced 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 $k \in \{1,\ldots,K\}$.[5] 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 vector specifying the probabilities that each element of the vocabulary will be chosen to fill an empty position given that topic.[6]

Early versions of topic models viewed the topic-prevalence vector as drawn from the same distribution for all documents. In STM the topic-prevalence vector is a function of variables characterizing documents. Incorporation of the data on these variables improves the identification of topics and allows the researcher to estimate the relationship between document characteristics and topic

[5] 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.

[6] 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.

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 (θ_d) is drawn from a logistic-normal distribution with parameters that are a function of the covariate values:

$$\theta_d \sim \text{LogisticNormal}(\mathbf{\Gamma x}_d', \mathbf{\Sigma}). \tag{1}$$

$\mathbf{\Gamma}$ is a matrix of coefficients relating covariate values to mean topic-prevalence. With $\mathbf{\Sigma}$ 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:

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

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:

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

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

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

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 $\mathbf{\Gamma}$, θ_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).

4. Data

4.1. The corpus of Bacon's works

The construction of the corpus began with a search for all Bacon's works that had been digitized accurately enough that most words could be machine read. Sources 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

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 (various essays, letters, speeches, case reports)	172
Total	282

Notes: Some names have been edited for rendering in modern English.

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 MorphAdorner (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 MorphAdorner. 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.

4.2. The metadata

The last step in organizing the data was to assign values of the metadata covariates to each document in order to use a more complete model that facilitates directly examining the effect of metadata covariates on topical prevalence. For reasons discussed in Section 7, we coded two metavariables that capture two different characteristics of Bacon's varied opus. The first is Bacon's intended audience—historians, methodologists, lawyers, politicians, scientists, or philosophers. The second is the form of his 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 intended audience was coded after careful scrutiny of each document, simultaneously taking into account existing scholarship on Bacon's opus.[7] The form of each text document was readily ascertained. Table 2 provides the document frequencies within each cell of the cross-classification of the two covariates.

In the estimations we describe below, we model topical prevalence (or importance) as a function of a linear combination of the dummy variables reflecting our two metadata covariates (see Eq. (1) above). Thus, 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.

5. Estimating and interpreting the features of Bacon's work

We now proceed to ask which topics are contained in Bacon's work. This is an absolutely fundamental research agenda, given Bacon's centrality in English cultural life. We are seeking to characterize the most prominent elements of the thought of a "cultural entrepreneur" whose ideas are believed to have spurred modern economic growth (Mokyr, 2016). Importantly, we are doing this in a novel way, using unsupervised machine-learning methods. Thus, our estimates are neither a product of our prior beliefs nor of the many interpretations of Bacon that have overlain his works in the four centuries since his death. Stated starkly, the topic model is

[7] Specifically, to code the intended audience, we examined each work, 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.

Table 2
Frequency of documents by audience and form of communication.

Form of communication 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

discovering the central elements of Bacon's ideas, which in turn influenced the development of the culture of growth. Our analysis thereby contributes to the emerging body of social science research that draws on text-as-data and utilizes novel computational methods to study social thought, perceptions, and attitudes as key ingredients of culture (see, e.g., Roberts et al., 2016a: Sec. 3.1; Lucas et al., 2015: Section 4; Blaydes et al., 2018; Gutman et al., 2018: 283–284).

5.1. Choosing the number of topics to be estimated

A key initial decision to be made in estimating a topic model is 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 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.[8] We first examined 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.

5.2. Representation of topics in the STM

Table 3 presents detailed results for the topics estimated by STM. Recall that a topic is formally a distribution over vocabulary. 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 are often 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. The names assigned to each topic in Table 3, however, are not produced by STM. They are our interpretations of the set of ideas that underlie each topic, interpretations justified in subsection 4.3.

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).[9] 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.

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: association with that topic with a high probability and a greater likelihood of being associated with that topic than with other topics.[10]

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

[9] 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.

[10] 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).

Table 3
Topics and top words for the estimated 16-topic STM.

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

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

(*continued on next page*)

Table 3 (*continued*)

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

Note that the 'words' listed above 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.

5.3. Characterizing the topics

The usual procedure in characterizing the topics by 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. In contrast, we present all estimated topics and provide evocative names that clearly capture the content for each of the topics.

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."[11] 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 for the five sections of *New Atlantis* that are included as separate items in our corpus, the three that are highly associated with this topic do not contain the descriptions of the infrastructure of science on Bensalem. This topic is 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.[12] Thus, STM identifies a key aspect of Bacon's scientific method that has been highlighted by multiple scholars (see e.g.,

[11] To enhance readability, we have edited the direct historical quotes that we provide by updating to modern spelling and punctuation.
[12] Our use of the word 'facts' might be somewhat anachronistic because the modern use of that word only gradually developed during the 17th

Peltonen, 1996a: 17; Mahlerbe, 1996; Mokyr, 2005: 289, 304): "an umbrella ideology for people to collect data" Mokyr (2016: 92).[13] 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 and engagement with high politics: 'solicitor', 'attorney', 'chancellor' are all associated with this topic. 'Essex' was a patron of Bacon whom Bacon later helped legally prosecute for treason. 'Coke' was Bacon's professional and personal archrival Edward Coke, for many the greatest lawyer of all times. Coke viciously opposed the idea of royal prerogative that Bacon, perhaps due to his professional position, was in contrast at least publicly more sympathetic to (Coquillette, 1992).[14] 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 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

(*footnote continued*)

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.

[13] 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.

[14] Hence Bacon's famous dictum from his essay *Of Judicature* that judges "be lions, but yet lions under the throne".

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.

5.4. 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). More broadly, the answer provides a window into an understanding of the main ideas that underpinned early 17th-century English culture, at least as absorbed or generated by Bacon.

One crucial insight provided by our results is best framed by contrasting our findings with some standard lessons from the existing intellectual history literature. Within the literature on Bacon's influence on later ideas, one finds emphasis on four distinctive contributions that Bacon has been said to make to Baconianism and, thus, the culture of growth. First, there is the set of ideas on the inductive logic of interpreting the world given the ascertained facts (see, e.g., Peltonen, 1996a; Rossi, 1996). Second, there is the emphasis on an intensive effort to find out about, and catalogue, the world (see, e.g., Malherbe, 1996: 79, 83). Third, 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). Fourth, 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).

We find only two of these four sets of ideas in STM's estimated topics. Epistemology matches the first and Probing for Facts the second. 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). However, there is no emphasis in Bacon on either the utilitarian value of useful knowledge or on the centralized organization of the scientific quest. Appendix B demonstrates that these results are fully robust to estimating STMs at much higher levels of granularity. Given that monarch would not have been generally opposed to ideas on the utilitarian benefits from science and the centralized organization of scientific quest, Bacon would have had no incentive to purposefully suppress them or hide them "between the lines" to avoid persecution (see, e.g., Strauss 1952: 24–26; Kuran 1995: 338). It follows that an emphasis on utilitarian value of science and the centralized organization of the scientific quest must have been a product of later interpretations of the followers of Bacon, rather than a product of Bacon's own emphases.[15]

We have therefore discovered a fundamental way in which the ideas emphasized by Bacon are very different from those usually associated with Baconianism. This has important implications because Bacon's ideas were so prominent in his own society that they provide a window into the emphases in 17th century English culture. In contrast, Baconianism was both later—the 18th century—and more widespread—across Western Europe. Our results suggest that it is the first two elements of Baconianism listed above that made England's culture distinctive at the time that the seeds of its development were being planted, in the 17th century. The second two elements listed above were not part of that culture. Inductive empiricism was crucial, but perhaps not the broader elements of the culture of growth.

Turning to a broader overview of the topics, the STM estimates illustrate the tremendous breadth of Bacon's opus, a subject much emphasized by modern scholars (Peltonen, 1992). Importantly, the set of topics unmistakably reveal that Bacon was, above all, a

[15] 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*.

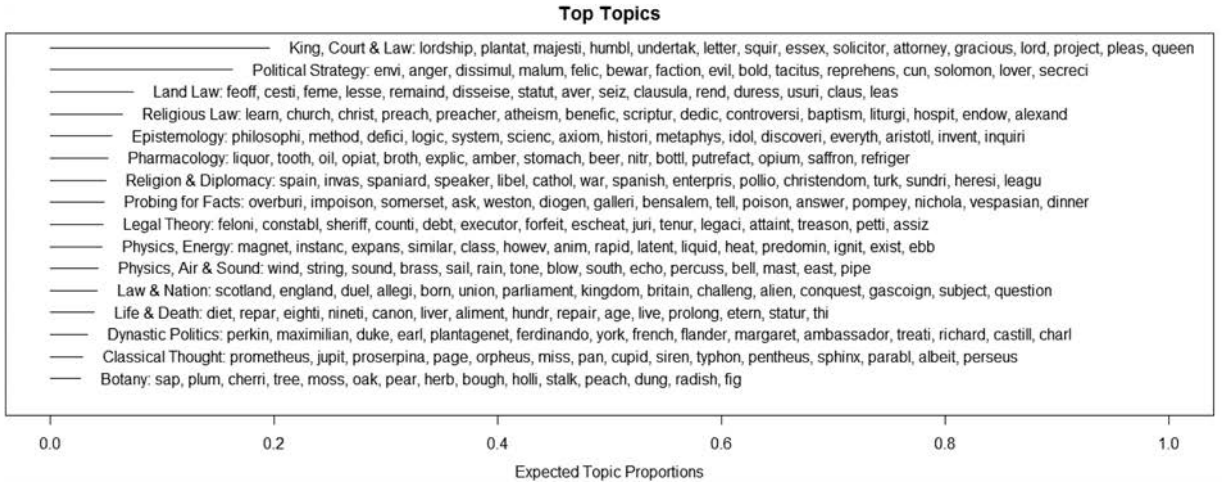

Fig. 1. Expected values for topic proportions in the Bacon corpus
Notes: For each topic, the figure displays the assigned names and the top fifteen FREX words. The size of the bars to the left of each topic is proportional to the probability that a random word drawn from the whole corpus has been generated by that particular topic.

lawyer and in particular a jurisprudential scholar. Fig. 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 engages in jurisprudence and scholarship as opposed to writing about comparatively more mundane legal matters in the context of his professional practice of law.

The finding that legal topics are many and prominent in Bacon's opus is a significant one for scholars of Bacon because 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.[16] At the same time, for scholars of comparative development, it is a noteworthy result that the father of the modern scientific method emphasized law and jurisprudence since it raises the question of the potential connection between the two, suggesting the possibility of an influence of legal reasoning on the development of scientific methodology.

6. 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 legal background and his inductive empiricism. Given Bacon's central role in the development of the culture of growth (Mokyr, 2016), quantitative evidence on the existence (or lack thereof) of a link between Bacon's experience in the law and his scientific methodology would cast light on the origins of the cultural ideas associated with England's rise. Moreover, given the distinctiveness of England's legal system and the widespread knowledge of law in England during Bacon's times, one would have an explanation of why Baconianism came earlier to England and diffused more quickly there than elsewhere.

We use STM to inquire into the connection between Bacon's ideas that emanated from his experience in the law and his scientific method based on the following reasoning. Given that Bacon was first and foremost educated in law and played all of the major legal roles during his professional engagement, statistical evidence of a connection between Bacon's legal topics and his methodological topics would support the hypothesis that Bacon's methodological ideas were rooted in his legal background. In contrast, a connection between Bacon's non-legal topics and his methodological topics would be consistent with the argument that the former were influenced by the latter. Similarly, a lack of a connection between Bacon's methodological topics and other topics would be consistent

[16] 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".

with 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.

6.1. 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.[17] 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. Fig. 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 θ_d' θ_d, which in turn reflects the metadata in $\mathbf{\Gamma}$ and the estimates of $\mathbf{x}_d$ and $\mathbf{\Sigma}$).[18]

A key empirical fact about Bacon's opus, evident from Fig. 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 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,[19] 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.[20] 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 Fig. 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

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

[18] 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.

[19] 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.

[20] 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 Murrell, 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.

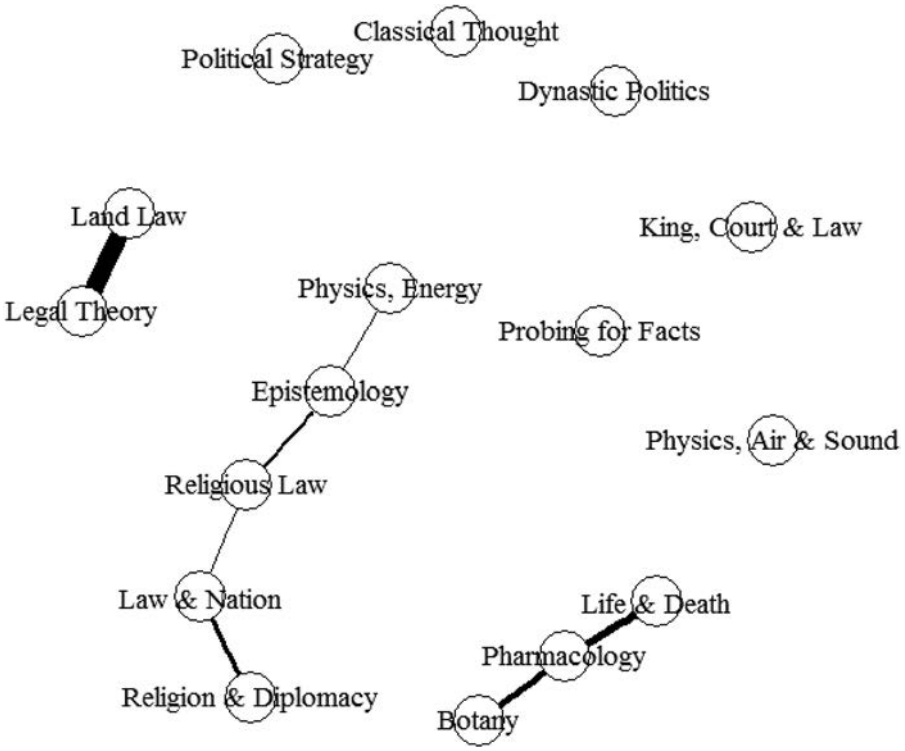

Fig. 2. Positive topic correlations
Note: The relative thickness of links reflects the magnitude of (positive) correlations between respective topic pairs.

results indicate strongly that the roots of Bacon's methodological ideas sprang primarily from his immersion in English common law (see Wheeler, 1983; Martin, 1992).

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 Fig. 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.

6.2. *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.[21] 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 requires 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 above results on correlations, we examine overlaps in the usage of words within topic pairs that include one legal topic and one methodological topic. We particularly focus 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 mode 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.

[21] 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).

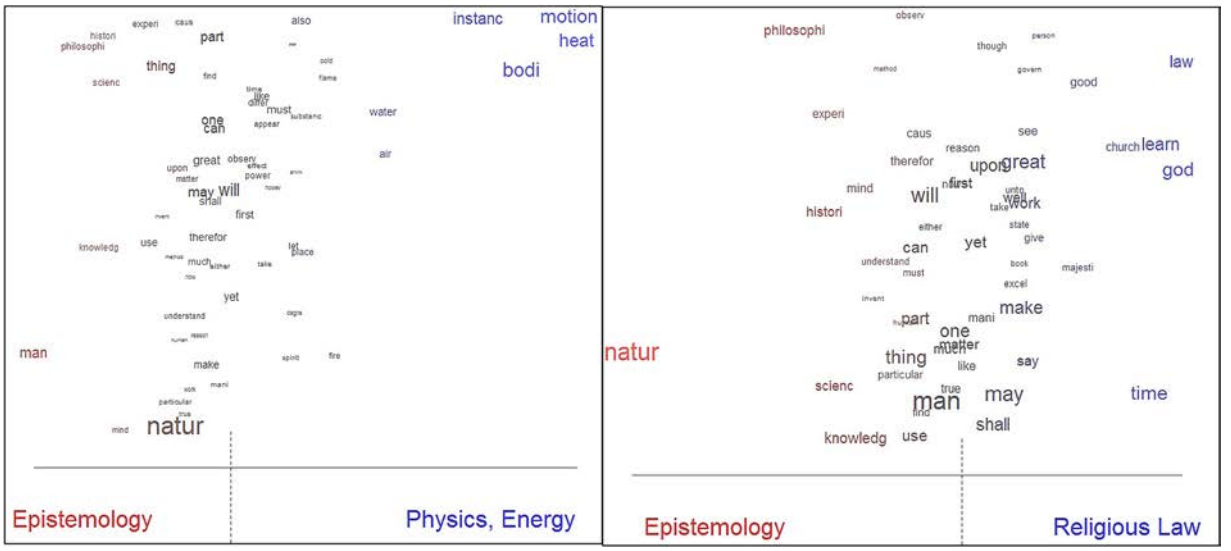

Fig. 3. Vocabulary use across topic pairs: Epistemology vs. Physics, Energy and Religious Law.

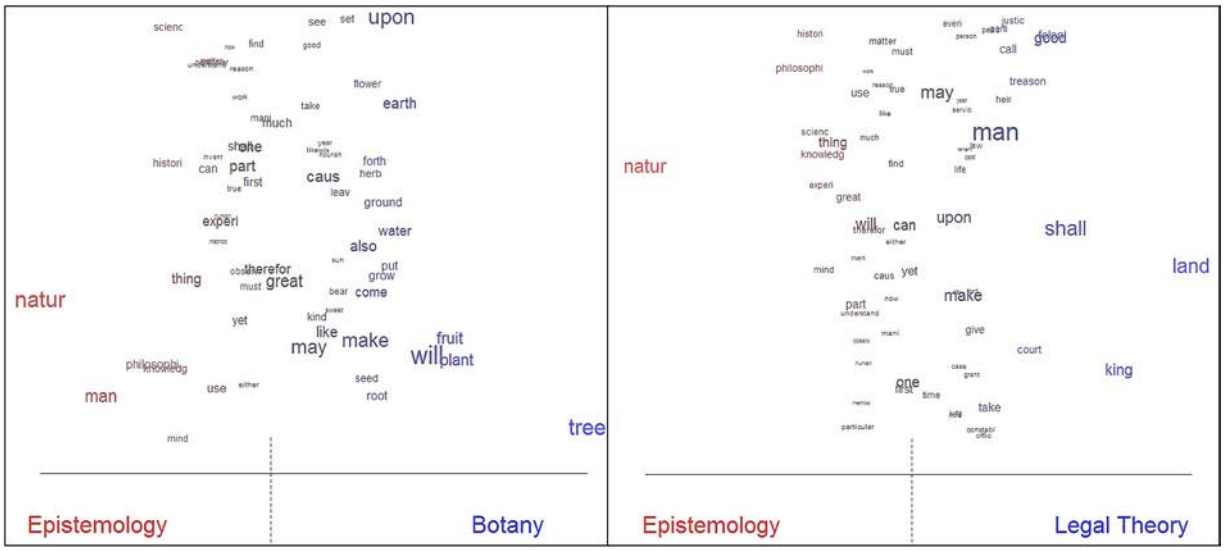

Fig. 4. Vocabulary use across topic pairs: Epistemology vs. Botany and Legal Theory.

To implement the test, in Figs. 3–6, we present a series of plots of the top 80 FREX words within chosen topic pairs.[22] In order to keep the number of figures to a minimum, for each of law and science, we choose one topic that is directly connected to Epistemology in the topic correlation diagram (Fig. 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).[23] 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.

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

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

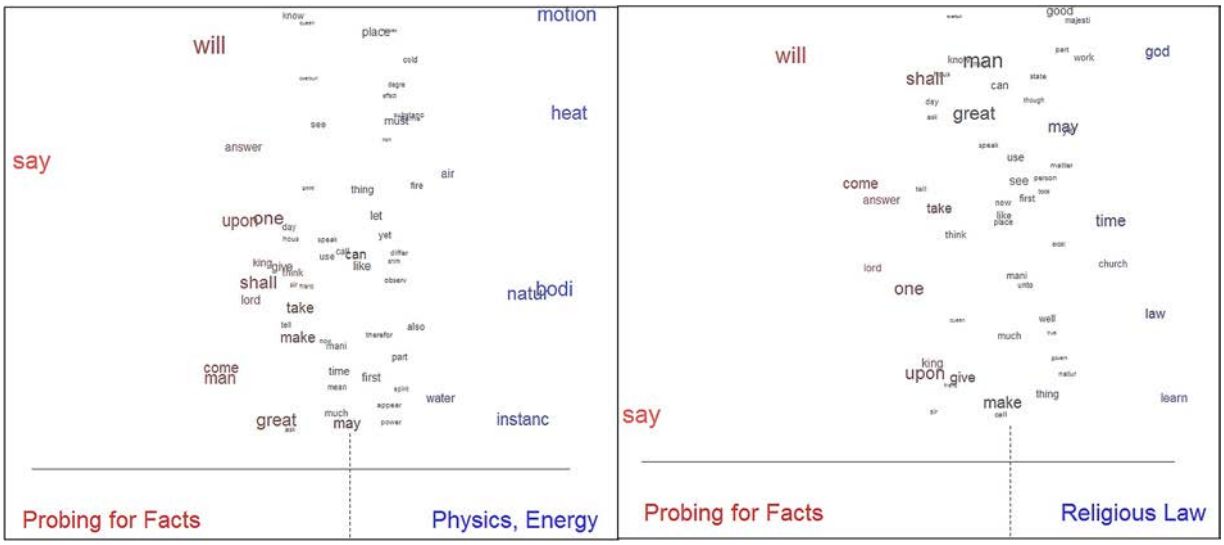

Fig. 5. Vocabulary use across topic pairs: Probing for Facts vs. Physics, Energy and Religious Law.

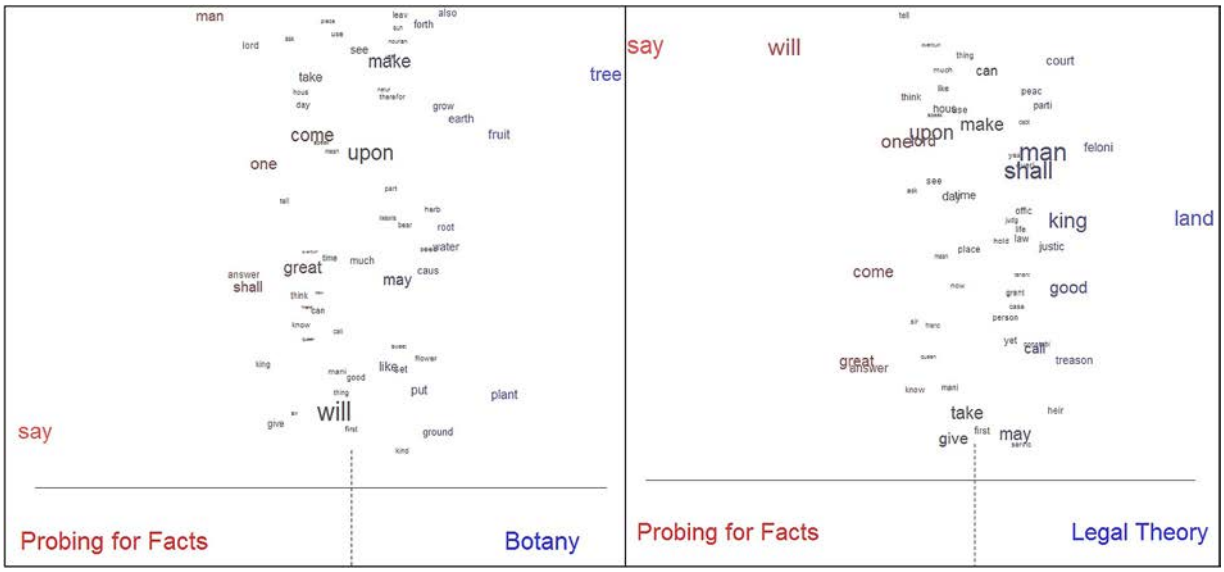

Fig. 6. Vocabulary use across topic pairs: Probing for Facts vs. Botany and Legal Theory.

Analyzing Fig. 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.[24] These are words that are central to inductive reasoning. Both Physics, Energy and Religious Law share these concepts equally with Epistemology. They are also at least as important 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 Fig. 4, which examines the words that Epistemology shares with Legal Theory and Botany, respectively.

[24] 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'.

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

Our STM-based evidence thus indicates that the most celebrated element of Bacon's thought—inductive reasoning—had its genesis in the law. As Appendix A shows, this conclusion is not prominent, and sometimes explicitly rejected, in the vast interpretative literature on Bacon. This is a finding of central interest to scholars of comparative development. It suggests that crucial elements of the very culture that was associated with England's economic rise (Mokyr, 2016) arose neither as a result of historical accident nor was exclusively a product of a one-of-a-kind genius, but rather grew from already established inductive principles embedded in the English common-law tradition.

Figs. 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. Again, STM-based evidence does not confirm the hypothesis that Bacon's background in the law 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).

7. Evidence on strategic communication of Bacon's ideas

Strategic communication of ideas, with its resultant effects on social equilibrium, has been a topic of direct interest to political economists at least since Kuran (1987, 1989, 1995), with recent research examining the importance of constraints on the communication strategies of powerful social actors (see, e.g., Coşgel et al., 2012; Rubin, 2014; Yanagizawa-Drott, 2014; Adena et al., 2015; Little, 2017). Bacon has often been thought of as communicating his ideas strategically, taking into account political constraints and the receptiveness of different constituencies (Peltonen, 1992; Zagorin, 1998: 180). We now explore how Bacon's emphasis on specific topics varied with the audience he was addressing and the form he used to communicate his ideas.

By producing evidence on Bacon's strategies for dissemination, we enhance the understanding of how Bacon's own strategic choices might have resulted in his ideas diffusing more strongly in some domains (e.g., science) than in others (e.g., history or politics). Because strategic communication of ideas by leading figures can alter public discourse, which in turn impacts the stock of private knowledge, it has the potential to shape the evolution of social order (Kuran, 1995: Sec. IV). With Bacon a cultural leader and with the rising interest in economics in the role of leaders in institutional change (Jones and Olken, 2005; Besley et al., 2011; Brender and Drazen, 2013), a closer understanding of Bacon's communication strategy provides insight into the evolution of English culture, directly resonating with concerns in the current literature on comparative economic development. Even more broadly, we show the power of STM in utilizing the original writings of individual social actors to discover which audiences were targeted with specific ideas and by which communication media. This is an approach that would be immediately applicable in many areas of comparative cultural history. For example, STM offers a means of quantitatively assessing the extent to which agents communicate their thoughts strategically, a phenomenon of central interest to scholars of comparative politics but difficult to measure empirically (Kuran, 1995: Ch.15; Jiang and Yang, 2016).

To analyze the impact of intended audience and form of communication, we proceed as follows. Using the 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 metadata covariate for all documents, thereby obtaining estimates of the pertinent column of $\boldsymbol{\Gamma}$ in (1) above (Roberts et al., 2016b: 17–18). We then plot either estimated topic proportions for different covariate values (Figs. 7, 8, 10, 12, 13, 14, 16) or differences in the estimated topic proportions for two different values of a covariate (Figs. 9, 11, 15).

We first examine the role of the intended audience. If Bacon was strategic about which audiences to address with specific ideas and, at the same time, mindful of the constraints imposed by powerful actors and social taboos, we would expect him to disproportionately emphasize specific topics to specific audiences. Our STM evidence shows that Bacon varied his use of the methodological topics that were much emphasized by his followers, Probing for Facts and Epistemology. Bacon's emphasis on questioning the world to establish facts (Probing for Facts) was addressed primarily to philosophers (Fig. 7), consistent with his eagerness to challenge the prevailing Aristotelian paradigm and despite the resistance he would have faced from important elements of this audience when emphasizing discovery of the new (see Section 2). Similarly, Bacon most emphasized his Epistemology when addressing methodologists (Fig. 8), strategically directing his work toward those social actors who would find it most useful and most likely to disseminate it more quickly.

Bacon did not emphasize the core elements of his ideas on scientific methodology when addressing other audiences (see Figs. 7 and 8). In particular, Bacon's methodological ideas took a back seat in his communications with politicians, even though it is well-

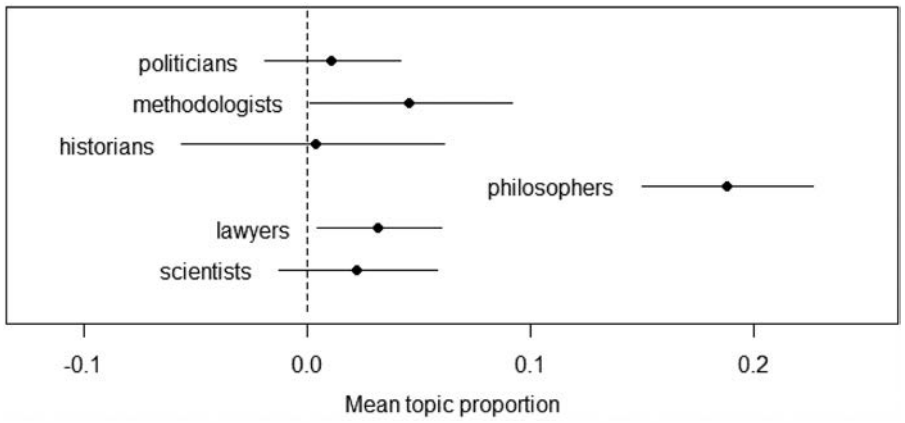

Fig. 7. The effect of audience on prevalence of Probing for Facts
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of probing for facts for each type of audience.

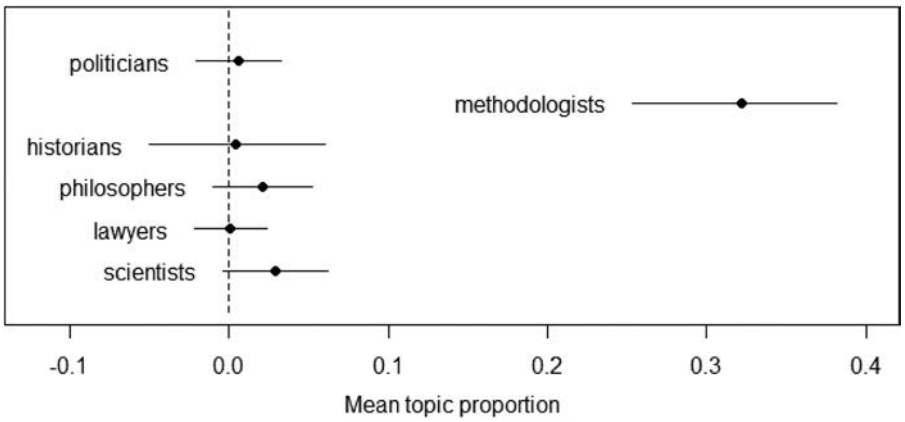

Fig. 8. The effect of audience on prevalence of Epistemology
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of epistemology for each type of audience.

known that Bacon had been eager to secure royal support for his work on natural philosophy (Berrios, 2012: 114).[25] Our results thus clearly capture the self-censorship that likely reflected Bacon's eventual "disenchant[ment] with the possibility of advancing his philosophical plans through a political medium…[after which he] divorced his philosophical ambitions from politics, became an obsequious politician and, employing an enigmatic style, confined his philosophical message to an exclusive audience" (Peltonen, 1996a: 9).

Hence, when addressing politicians (Fig. 9), Bacon emphasized King, Court & Law, a topic that blended Bacon's immersion in power politics with his expertise in law (see Section 5.3). Politicians, not fellow lawyers, were the primary intended audience for that topic (Fig. 10). Indeed, when he addressed lawyers, Bacon placed more emphasis on other legal topics (Land Law and Legal Theory in Fig. 11). Moreover, Bacon did not publish ideas on King, Court & Law in formats that would render them available to the broader public. Carefully distinguishing between the private and the public domain (see Kuran, 1995), he communicated them primarily in private letters (Fig. 12). Bacon also communicated extensively with politicians on religious matters in connection with both law and diplomacy (Fig. 9), a potentially contentious subject matter, but also one on which Bacon and James I shared broadly congruent views (see Section 2)

Bacon developed important ideas about political tactics (Political Strategy), including on dissimulation and the importance of strategic behavior (see Section 5.3). But he did not choose politicians as his primary audience when communicating these ideas. Instead, he contributed these insights to philosophical debates (Fig. 13). Perhaps Bacon, himself deeply immersed in political games, did not want to suggest strategies that could have been used against him. All of the above findings are consistent with the prevailing view of Bacon as a shrewd political actor, adept at strategically maneuvering across the political landscape, mindful of powerful constituencies, and relentless at pursuing his own political goals (Peltonen, 1996b; Drinker Bowen, 1993; Powell, 1996).

Our evidence shows Bacon preferring some communication forms more than others when disseminating core elements of his methodological thought. When emphasizing Probing for Facts, Bacon preferred apothegms (Fig. 14).[26] With apothegms constituting a

[25] In our coding of the metadata, monarchs are politicians.
[26] We distinguish here between Bacon's apothegms, short pithy disconnected statements, and his use of aphorisms in the longer works *Novum*

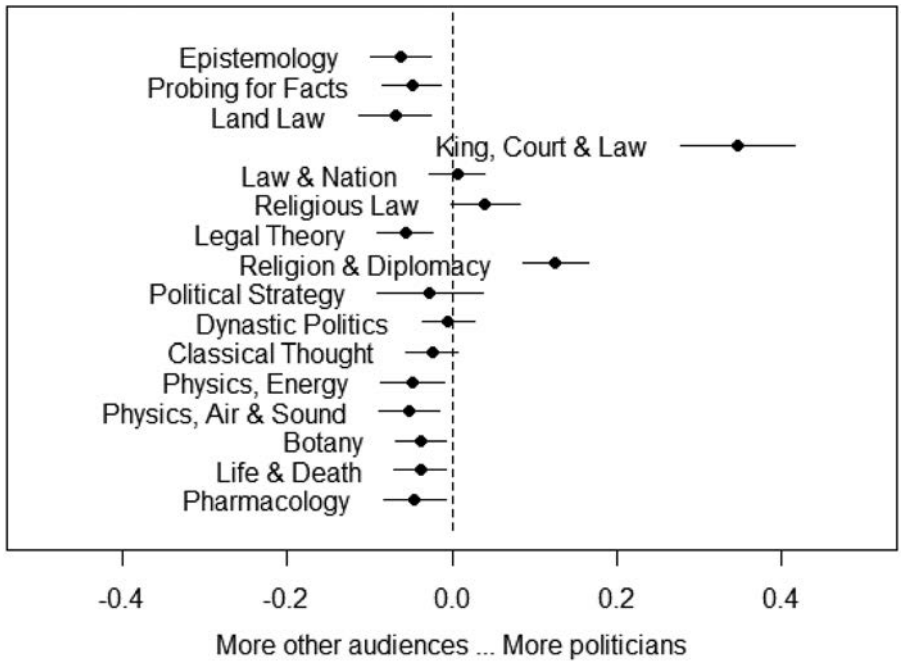

Fig. 9. The effect of audience, politicians
Note: The figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which politicians are the intended audience compared to works addressed to all other audiences.

widely adopted and popular approach utilized by authors of 'commonplace books' during the early modern period (Havens, 2002), this is evidence of Bacon's intention to disseminate his ideas about fact-seeking broadly and in line with his vision that the scientific enterprise should engage a broad swath of society. Since Bacon relied more on apothegms also when he expressed some of his legal ideas (Fig. 15), this is a further piece of evidence that he incorporated legal reasoning into his approach to scientific methodology.

In contrast, Bacon did not use apothegms to convey his Epistemology, but rather presented his epistemological ideas primarily in book-length volumes and, to an extent, essays (Fig. 16). 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 unworthy of an aristocrat. This is perhaps another way in which Bacon's immersion within debates in the legal profession, where he was on the side of the publicists (Ross, 1998), affected his work on scientific methodology.

In sum, evidence on the effects of both intended audience and form of communication demonstrates that Bacon was keenly aware of the political, professional, and social constraints that he faced and accordingly communicated his ideas strategically, with corresponding repercussions for their social reception.

8. Conclusion

The role of political and intellectual leaders in institutional change and economic development is of fundamental interest to comparative social scientists. We have focused on the features and the origins of the ideas of a preeminent social figure during a pivotal period in the history of modern development. Uniquely to-date in economics, we have illustrated how machine-learning methods for text analysis can provide a direct route to the ideas emphasized by important social actors, distinguishing between those ideas and the later interpretations of followers. Our methodology has broad applicability for the analysis of historical-cultural phenomena, and in particular for uncovering cultural ideas and the connections between them. We show how machine-learning can be used to draw conclusions about the wider cultural origin of specific ideas and about the extent and the nature of strategic communication. This provides insights into the determinants of the speed of diffusion of those ideas. These are all key steps in any analysis striving to characterize the evolution of ideas and, more broadly, culture.

We conclude by going beyond our detailed results to make a conjecture about a much broader implication of our findings. We place those findings within the broader sweep of the economics literature that focuses on what English experience can tell us about the complicated puzzle of how modern societies emerge. A key observation in that literature is that when England became the first

(*footnote continued*)
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.*

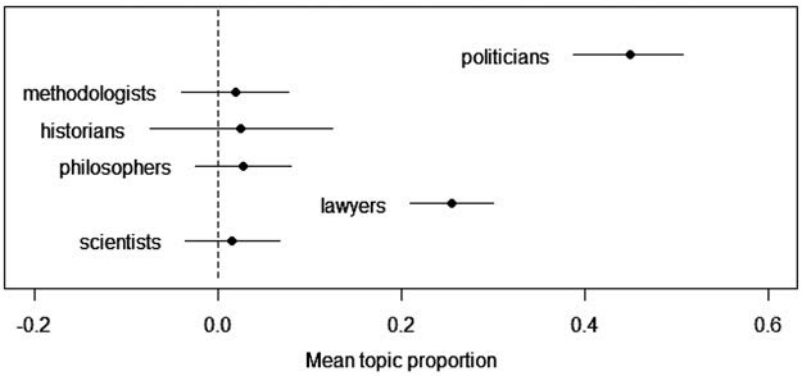

Fig. 10. The effect of audience on prevalence of King, Court & Law
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of king, court & law for each type of audience.

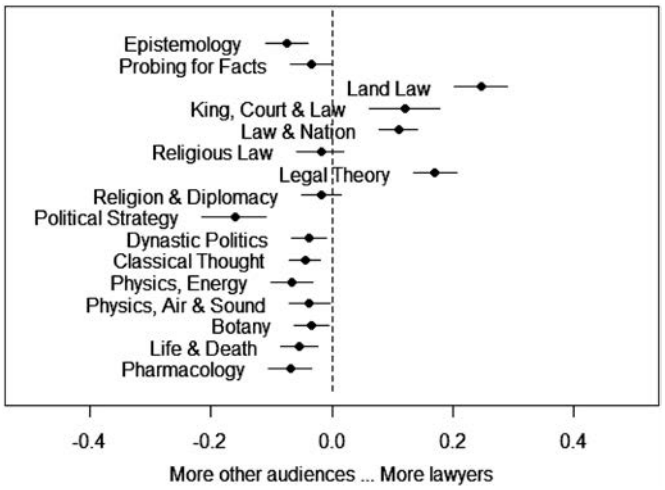

Fig. 11. The effect of audience, lawyers
Note: The figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for works for which lawyers are the intended audience compared to works addressed to all other audiences.

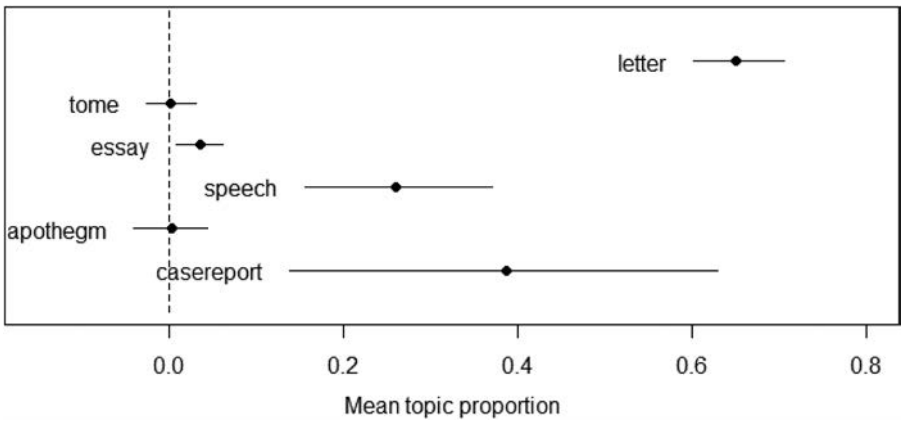

Fig. 12. The effect of form of communication on prevalence of King, Court & Law
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of king, court & law for each type of form of communication.

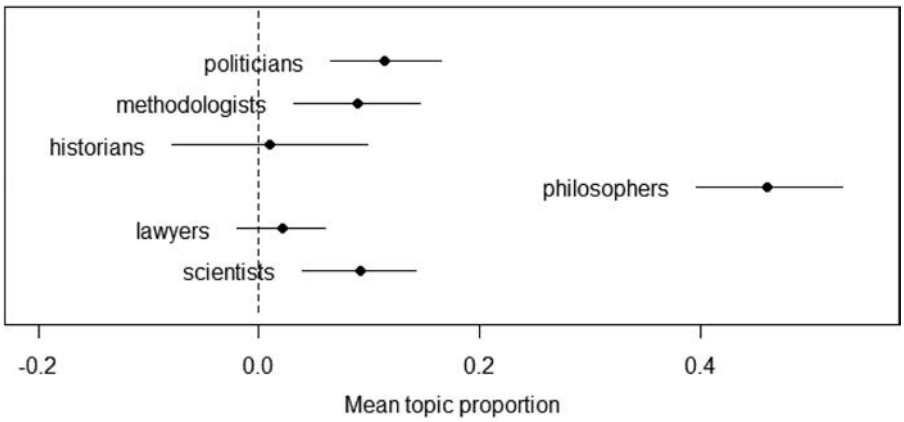

Fig. 13. The effect of audience on prevalence of Political Strategy
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of political strategy for each type of audience.

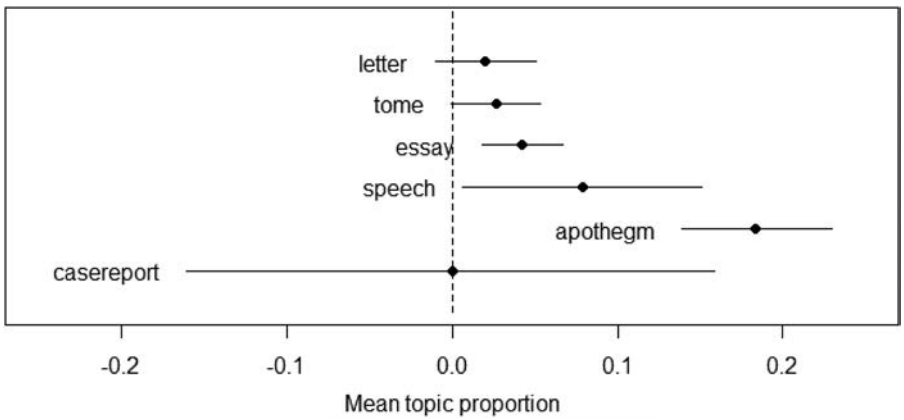

Fig. 14. The effect of form of communication on prevalence of Probing for Facts
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Probing for Facts for each type of form of communication.

country to evidence modern economic development, its political institutions were far superior to those of any other country (North and Weingast, 1989; Acemoglu and Robinson, 2012). Thus, it is tempting to conclude that it is the formal institutions that led to the economic development. Our findings suggest resisting the temptation: the coincidence of better political institutions and economic development could be the product of deeper phenomena rooted in English common-law culture.

The common law rose to a degree of sophistication and effectiveness over many centuries, reaching an unparalleled degree of centrality in English life in the 17th century (Holdsworth, 1927). Thus, a long line of research in history, political science, and law has argued that the influence of common-law thinking was a key ingredient in the development of political institutions in England during the 17th century (Nenner, 1977; Cromartie, 2006). This influence arose, at least in part, because the politico-legal ideas emanating from the experience with the common-law had diffused into the broader English culture, an education in law being prevalent among the upper reaches of English society at that time (Stone, 1964; Hill, 1997). In Nenner's (1977: x) words: "By the seventeenth century, England had developed a political culture completely comfortable with sophisticated legal concepts…[A] legal disposition of mind… was being increasingly brought to bear upon political and constitutional problems. It conditioned men's thought and language and ultimately their actions." The existing literature, therefore, gives us a direct link from the common-law culture of the early 17th century to the political institutions of the early 18th century.

The piece of evidence that our paper adds is in showing how key elements of Bacon's ideas were a reflection of aspects of the common-law culture in which he was immersed. As Bacon began his work on science, the ground was being prepared for a change in ideas about science by the everyday culture of London, dabbling in exercises to understand nature (Harkness, 2007), and by Europe's broader scientific elite, grappling with new ideas about discovery (Wootton, 2015). Bacon provided influential leadership, helping to shape the nature of public discourse about the methods and tools of the scientific quest. Perhaps one reason for his success was that he saw how to skillfully combine both new and old methods of communication and maneuver between different audiences, transmitting his ideas about scientific methodology strategically. Given the resonance of Bacon's ideas with the prevailing common-law culture of the elite, given Bacon's strategic dissemination of his thoughts, and since public communication of ideas by an influential leader can

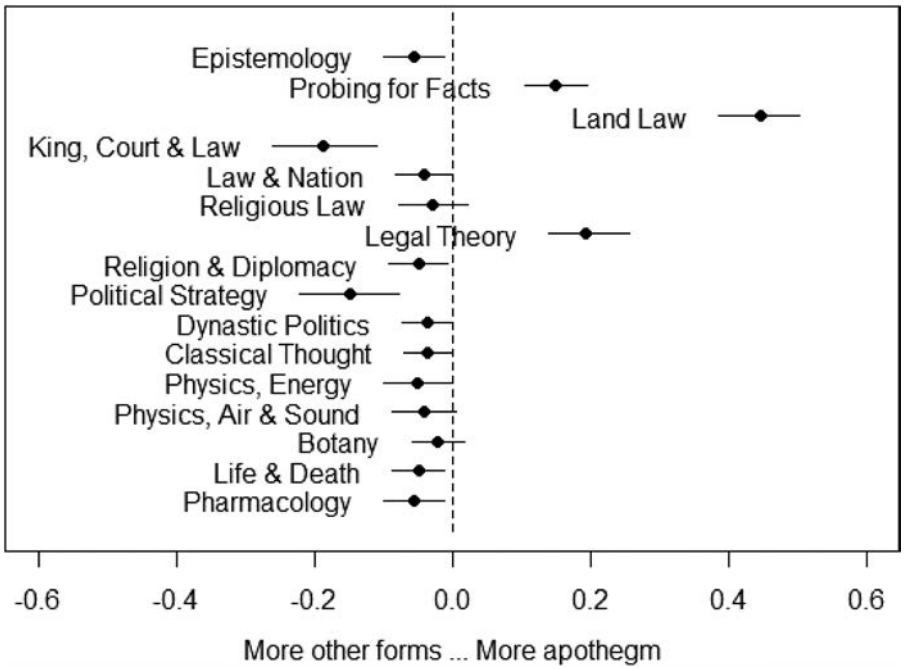

Fig. 15. The effect of form of communication, apothegm
Note: The figure shows the point estimate and 80% confidence interval of the mean difference in topic proportions for apothegms compared to all other forms of communication.

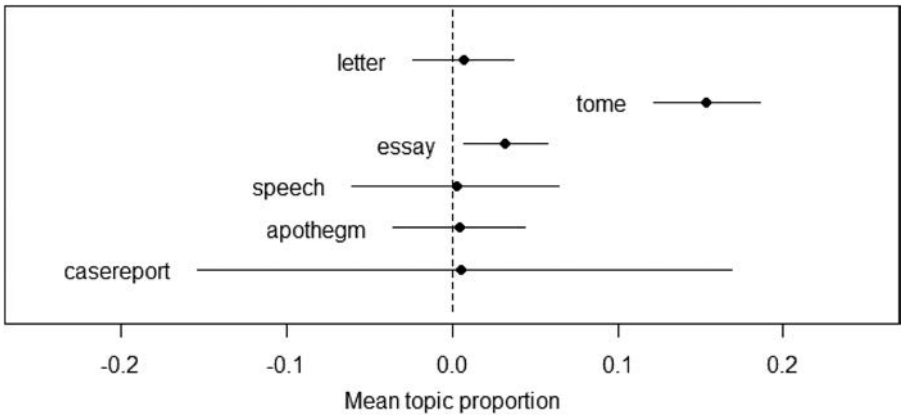

Fig. 16. The effect of form of communication on prevalence of Epistemology
Note: The figure shows the point estimate and 80% confidence interval of the mean topic proportions of Epistemology for each type of form of communication.

result in rapid change in prevailing social beliefs and customs (Kuran, 1995), it is not surprising that elements of Baconianism diffused earlier in England than elsewhere. Combined with Mokyr's (2005, 2006, 2010, 2016) compelling case for the subsequent influence of Baconianism on technological progress, our findings provide evidence in support of a direct connection between England's common-law culture in the 17th century and England's early economic development in the 18th century, with Bacon's work providing the most important link.

Therefore, the observed association between sophisticated political institutions and early economic development in 18th century England need not necessarily reflect a causal effect of the former on the latter. Instead, both better political institutions and sustained economic growth could have emerged as a consequence of England's many years of experience with the common-law and the culture that this experience generated. The influence of that culture can help explain both the institutional and the economic development. Perhaps, the unique precocity of England in both aspects of development was due to the common-law culture that had developed over many centuries and suffused English society. After all, Bacon himself, while enormously important in communicating the vast potential of the ideas within that culture, viewed truth as "the daughter of time, not of authority".

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jce.2018.10.004.

References

Acemoglu, D., Robinson, J.A., 2012. Why Nations Fail: The Origins of Power, Prosperity, and Poverty. Crown Publishers, New York, NY.
Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., Zhuravskaya, E., 2015. Radio and the rise of the nazis in prewar Germany. *Q. J. Econ.* 130 (4), 1885–1939.
Alesina, A., Giuliano, P., 2015. Culture and institutions. *J. Econ. Lit.* 53 (4), 898–944.
Algan, Y., Cahuc, P., 2010. Inherited trust and growth. *Am. Econ. Rev.* 100 (5), 2060–2092.
Bacon, F., 1859. Valerius Terminus: Of the Interpetation of Nature. In: Spedding, J., Ellis, R.L., Heath, D.D. (Eds.), The Works of Francis Bacon Vol. III Longman, London, UK.
Bacon, F., 2014. *Novum Organum Or True Suggestions for the Interpretation of Nature*. Edited by Devey, J., Project Gutenberg EBook #45988. https://www.gutenberg.org/files/45988/45988-h/45988-h.htm. Accessed February 2018.
Baker, J., 2004. The Common Lawyers and the Chancery: 1616. In: Boyer, A.D. (Ed.), *Law, Liberty, and Parliament: Selected Essays on the Writings of Sir Edward Coke*. Liberty Fund, Indianapolis, pp. 254–281.
Bandiera, O., Hansen, S., Prat, A., Sadun, R., 2017. "CEO Behavior and Firm Performance." Harvard Business School Working Paper 17-083.
Barro, R.J., McCleary, R.M., 2003. Religion and economic growth across countries. *Am. Sociol. Rev.* 68 (5), 760–781.
Beck, T., Demirguc-Kunt, A., Levine, R., 2003. Law and finance: why does legal origin matter? *J. Compar. Econ.* 31 (4), 653–675.
Becker, S., Boeckh, K., Hainz, C., Woessmann, L., 2016. The empire is dead, long live the empire! Long-run persistence of trust and corruption in the bureaucracy. *Econ. J.* 126 (590), 40–74.
Bellstam, G., Bhagat, S., and Cookson, J.A. 2016. "A text-based analysis of corporate innovation." SSRN Working Paper No. 2803232.
Benjamin, E.B., 1965. Bacon and tacitus. *Class. Philol.* 60 (2), 102–110.
Berrios, G.E., 2012. The 19th-Century Nosology of Alienism: History and Epistemology. In: Kendler, K.S., Parnas, J. (Eds.), Philosophical Issues in Psychiatry II: Nosology. Cambridge University Press, Oxford, UK, pp. 101–117.
Besley, T., Montalvo, J.G., Reynal-Querol, M., 2011. Do educated leaders matter? *Econ. J.* 121 (554), F205–F227.
Blaydes, L., Grimmer, J., McQueen, A., 2018. Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds. Journal of Politics 80 (4), 1150–1167.
Blei, D.M., 2012. Probabilistic topic models. *Commun. ACM* 55 (4), 77–84.
Blei, D.M., Lafferty, J.D., 2007. A correlated topic model of science. *Annal. Appl. Stat.* 1 (1), 17–35.
Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.* 3, 993–1022.
Bowles, S., Gintis, H., 2011. A Cooperative Species: Human Reciprocity and Its Evolution.
Brender, A., Drazen, A., 2013. Elections, leaders, and the composition of government spending. *J. Pub. Econ.* 97 (C), 18–31.
Cardwell, K.W., 1990. Francis Bacon, Inquisitor. In: Sessions, W.A. (Ed.), *Francis Bacon's Legacy of Texts*. AMS Press, New York, NY, pp. 269–289.
Clark, A., 1898. 'Brief lives', Chiefly of Contemporaries, Set Down By John Aubrey, Between the Years 1669 & 1696. At the Clarendon Press, Oxford, UK.
Coquillette, D.R., 1992. Francis Bacon. Stanford University Press, Stanford, CA.
Coquillette, D.R., 2004. The Anglo-American Legal Heritage, Introductory Materials, Second Edition. Carolina Academic Press, Durham, NC.
Coşgel, M.M., Miceli, T.J., Rubin, J., 2012. The political economy of mass printing: legitimacy and technological change in the ottoman empire. *J. Comp. Econ.* 40 (3), 357–371.
Cromartie, A., 2006. The Constitutionalist Revolution: An Essay On the History of England, 1450-1642. Cambridge University Press, Cambridge, UK.
Damaška, M.R., 1986. The Faces of Justice and State Authority. A Comparative Approach to the Legal Process. Yale University Press, New Haven, CT.
de Montmorency, J.E.G., 1905. Francis Bacon. *J. Soc. Comp. Legis.* 6 (2), 263–283.
Djankov, S., Glaeser, E., La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2003. The new comparative economics. *J. Comp. Econ.* 31 (4), 595–619.
Drinker Bowen, C., 1993. Francis Bacon, The Temper of a Man. Fordham University Press, New York, NY.
Fligstein, N., Brundage, J. S., and Schultz, M. 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, S., 2001. Francis Bacon and the Transformation of Early-Modern Philosophy. Cambridge University Press, Cambridge, UK.
Gentzkow, M., Kelly, B.T., and Taddy, M. 2017. "Text as data." NBER Working Paper No. 23276.
Grajzl, P., Murrell, P., 2016. A Darwinian theory of institutional development two centuries before Darwin. *J. Econ. Behav. Org.* 131 (A), 346–372.
Guerriero, C., 2016. Endogenous legal traditions and economic outcomes. *J. Comp. Econ.* 44 (2), 416–433.
Guiso, L., Sapienza, P., Zingales, L., 2006. Does culture affect economic outcomes? *J. Econ. Perspect.* 20 (2), 23–48.
Guiso, L., Sapienza, P., Zingales, L., 2016. Long term persistence. *J. Eur. Econ. Assoc.* 14 (6), 1401–1436.
Gutmann, M.P., Klancher Merchant, E., Roberts, E., 2018. 'Big data' in economic history. *J. Econ. Hist.* 78 (1), 268–299.
Hansen, S., McMahon, M., 2016. Shocking language: understanding the macroeconomic effects of central bank communication. *J. Intern. Econ.* 99, S114–S133.
Hansen, S., McMahon, M., Prat, A., 2018. Transparency and Deliberation Within the FOMC: A Computational Linguistics Approach. *Quarterly Journal of Economics* 133 (2), 801–870.
Harkness, D.E., 2007. The Jewel House: Elizabethan London and the Scientific Revolution. Yale University Press, New Haven, CT.
Hathi Trust. n.d. https://www.hathitrust.org. Accessed January-May 2016.
Havens, E., 2002. 'Of common places, or memorial books': an anonymous manuscript on commonplace books and the art of memory in seventeenth-century England. *Yale Univ. Lib. Gaz.* 76 (3/4), 136–153.
Hill, C., 1997. Intellectual Origins of the English Revolution Revisited. Oxford University Press, New York, NY.
Holdsworth, W.S., 1927. Bacon. The Cambridge Law Journal 3. pp. 10–12.
Holdsworth, W.S., 1938. Some Makers of English Law. Cambridge University Press, London, UK.
Huang, A., Lehavy R., Zang A., and Zheng. R., 2016. "Analyst information discovery and interpretation roles: a topic modeling approach." University of Michigan Ross School of Business Working Paper No. 1229.
Hughes, J., 1986. The Vital Few: The Entrepreneur and American Economic Progress. Oxford University Press, New York, NY.
Internet Archive. n.d. https://archive.org. Accessed January-May 2016.
Jardine, L., 1974. Francis Bacon, Discovery and the Art of Discourse. Cambridge University Press, London, UK.
Jiang, J., Yang, D.L., 2016. Lying or believing? Measuring preference falsification from a political purge in China. *Comp. Polit. Stud.* 49 (5), 600–634.
Jonas, H., 1984. The Imperative of Responsibility: In Search of an Ethics for the Technological Age. University of Chicago Press, Chicago, IL.
Jones, B.F., Olken, B.A., 2005. Do leaders matter? National leadership and growth since world war II. *Q. J. Econ.* 120 (3), 835–864.
Kocher, P.H., 1953. Science and Religion in Elizabethan England. The Huntington Library, San Marino, CA.
Kocher, P.H., 1957. Francis Bacon on the science of jurisprudence. *J. Hist. Ideas* 18 (1), 3–26.
Kuran, T., 1987. Preference falsification, policy continuity and collective conservatism. *Econ. J.* 97, 642–665.
Kuran, T., 1989. Sparks and prairie fires: a theory of unanticipated political revolution. *Pub. Choice* 61 (1), 41–74.
Kuran, T., 1995. Private Truths, Public Lies. The Social Consequences of Preference Falsification. Harvard University Press, Cambridge, MA.
La Porta, R., Lopez-de-Silanes, F., Shleifer, A., 2008. The economic consequences of legal origins. *J. Econ. Lit.* 46 (2), 285–332.

Larsen, V. H. and Thorsrud., L. A. 2015. "The value of news." Unpublished manuscript.
Licht, A.N., Goldschmidt, C., Schwartz, S.H., 2007. Culture rules: the foundations of the rule of law and other norms of governance. *J. Comp. Econ.* 35 (4), 659–688.
Little, A.T., 2017. Propaganda and credulity. *Games Econ. Behav.* 102, 224–232.
Losee, J., 2001. A Historical Introduction to the Philosophy of Science. Oxford University Press, Cambridge, UK.
Lucas, C., Nielsen, R.A., Roberts, M.E., Stewart, B.M., Storer, A., Tingley, D., 2015. Computer-assisted text analysis for comparative politics. *Polit. Anal.* 23 (2), 254–277.
Mahlerbe, M., 1996. Bacon's Method of Science. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 75–98.
Martin, J., 1992. Francis Bacon, the State, and the Reform of Natural Philosophy. Cambridge University Press, Cambridge, UK.
Mimno, D., Wallach, H.M., Talley, E., Leenders, M., McCallum, A., 2011. Optimizing semantic coherence in topic models. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pp. 262–272.
Moby Word Lists. 2002. http://www.gutenberg.org/ebooks/3201. Accessed April 28, 2016.
Mokyr, J., 2005. The intellectual origins of modern economic growth. *J. Econ. Hist.* 65 (2), 285–381.
Mokyr, J., 2006. The Great Synergy: the European Enlightenment as a Factor in Modern Economic Growth. In: Dolfsma, W., Soete, L. (Eds.), *Understanding the Dynamics of a Knowledge Economy*. Edward Elgar, Cheltenham, UK Chapter 1.
Mokyr, J., 2010. The Contribution of Economic History To The Study of Innovation and Technical Change: 1750-1914. In: Bronwyn, H.H., Rosenberg, N. (Eds.), *Handbook of the Economics of Innovation Vol. 1*. Academic Press, Burlington, MA, pp. 11–50.
Mokyr, J., 2016. A Culture of Growth: The Origins of the Modern Economy. Princeton University Press, Princeton, NJ.
MorphAdorner. 2013. http://morphadorner.northwestern.edu. Accessed April 26, 2016.
Murrell, P., 2017. Design and evolution in institutional development: the insignificance of the English Bill of Rights. *J. Comp. Econ.* 45 (1), 36–55.
Nenner, H., 1977. By Colour of Law: Legal Culture and Constitutional Politics in England, 1660-1689. University of Chicago Press, Chicago.
North, D., Weingast, B., 1989. Constitutions and commitment: the evolution of institutions governing public choice in seventeenth-century England. *J. Econ. Hist.* 49 (4), 803–832.
Nunn, N., 2008. The Long-term Effects of Africa's Slave Trades. *Q. J. Econ.* 123 (1), 139–176.
Nunn, N., 2012. Culture and the historical process. *Econ. Hist. Develop. Reg.* 27 (S1), 108–126.
Peltonen, M., 1992. Politics and science: Francis Bacon and the true greatness of states. *Histor. J.* 35 (2), 279–305.
Peltonen, M., 1996a. Introduction. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 1–24.
Peltonen, M., 1996b. Bacon's Political Philosophy. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 283–310.
Peltonen, M., 2004. Bacon; Francis, Viscount St Alban (1561-1626). Oxford Dictionary of National Biography. Oxford University Press, Oxford.
Perseus. n.d. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059. Accessed May-December 2016.
Powell, D.X., 1996. Why was Sir Francis Bacon impeached? The common lawyers and the chancery revisited:1621. History 81 (264), 511–526.
Project Gutenberg. n.d. https://www.gutenberg.org. Accessed January-May 2016.
Roberts, M.E., Stewart, B.M., Tingley, D., 2016c. Navigating the Local Modes of Big Data: The Case of Topic Models. In: Alvarez, R.M. (Ed.), *Computational Social Science, Discovery and Prediction*. Cambridge University Press, New York, NY, pp. 51–97.
Roberts, M.E., Stewart, B.M., Tingley, D., 2016b. STM: R package for structural topic models. *J. Stat. Softw.* forthcoming.
Roberts, M.E., Stewart, B.M., Airoldi, E.M., 2016a. A model of text for experimentation in the social sciences. *J. Am. Statist. Assoc.* 111 (515), 988–1003.
Roberts, M.E., Stewart, B.M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S., Albertson, B., Rand, D.G., 2014. Structural topic models for open-ended survey responses. *Am. J. Polit. Sci.* 58 (4), 1064–1082.
Ross, R.J., 1998. The commoning of the common law: the renaissance debate over printing English law,1520-1640. *Univ. Penn. Law Rev.* 146 (2), 323–461.
Rossi, P., 1968. Francis Bacon: From Magic to Science. The University of Chicago Press, Chicago, IL.
Rossi, P., 1996. Bacon's Idea of Science. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 25–46.
Rubin, J., 2014. Printing and protestants: an empirical test of the role of printing in the reformation. *Rev. Econ. Stat.* 96 (2), 270–286.
Sagasti, F., 2000. The twilight of the Baconian age and the future of humanity. Futures 32, 595–602.
Sargent, R.M., 1996. Bacon as an Advocate for Cooperative Scientific Research. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 146–171.
Schinke, R., Greengrass, M., Robertson, A.M., Willett, P., 1996. A stemming algorithm for Latin text databases. *J. Document.* 52, 172–187.
Serjeantson, R., 2014. "Francis Bacon and the 'interpretation of nature' in the late renaissance. Isis 105 (4), 681–705.
Sgarbi, M., 2012. The Aristotelian Tradition and the Rise of British Empiricism. Logic and Epistemology in the British Isles (1570-1689). Springer, New York, NY.
Shapiro, B.J., 1980. Sir Francis Bacon and the mid-seventeenth century movement for law reform. *Am. J. Legal Hist.* 24 (4), 331–362.
Shapiro, B.J., 2000. A Culture of Fact: England, 1550-1720. Cornell University Press, Ithaca, NY.
Simonds, R.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. Center for Medieval and Early Renaissance Studies, Binghamton, NY.
Snider, A., 1991. Bacon, legitimation, and the 'origin' of restoration science. *Eighteenth Cent.* 32 (2), 119–138.
Stone, L., 1964. "The educational revolution in England, 1560-1640." *Past and Present*. 28(1), 41–80.
Strauss, L., 1952. Persecution and the Art of Writing. The Free Press, New York, NY.
Tabellini, G., 2008. Institutions and culture. *J. Eur. Econ. Assoc.* 6 (2-3), 255–294.
Taddy, M.A., 2012. On estimation and selection for topic models. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics, pp. 1184–1193.
The Text Creation Partnership for Early English Books Online. 2014. http://www.textcreationpartnership.org.
Tingley, D., 2017. "Rising Power on the Mind." *International Organization*, 71:S1, S165–S188.
Tinkler, J.F., 1996. Bacon and History. In: Peltonen, M. (Ed.), *The Cambridge Companion to Bacon*. Cambridge University Press, Cambridge, UK, pp. 232–259.
Vickers, B., 1978. Francis Bacon. Longman, London, UK.
Vickers, B., 1992. Francis Bacon and the progress of knowledge. *J. Hist. Ideas* 53 (3), 495–518.
Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D., 2009. Evaluation methods for topic models. In: ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning, pp. 1105–1112.
Wheeler, H., 1983. The invention of modern empiricism: juridical foundations of Francis Bacon's philosophy of science. *Law Lib. J.* 76, 78–120.
Whitaker's Words. 2006. http://archives.nd.edu/whitaker/wordsoth.htm. Accessed May 15, 2016.
Wootton, D., 2015. The Invention of Science: A New History of the Scientific Revolution. Harper, New York, NY EPub Edition.
Yanagizawa-Drott, D., 2014. Propaganda and conflict: theory and evidence from the Rwandan genocide. *Q. J. Econ.* 129 (4), 1947–1994.
Zagorin, P., 1998. Francis Bacon. Princeton University Press, Princeton, NJ.