



RESEARCH ARTICLE

# A macroscope of English print culture, 1530–1700, applied to the coevolution of ideas on religion, science, and institutions

Peter Grajzl<sup>1</sup>  and Peter Murrell<sup>2</sup> 

<sup>1</sup>Department of Economics, The Williams School of Commerce, Economics, and Politics, Washington and Lee University and CESifo, Munich, Germany and <sup>2</sup>Department of Economics, University of Maryland, College Park, MD, USA

**Corresponding author:** Peter Murrell; Email: [pmurrell@umd.edu](mailto:pmurrell@umd.edu)

(Received 1 January 2023; revised 25 March 2024; accepted 28 March 2024; first published online 14 May 2024)

## Abstract

We combine unsupervised machine learning and econometric methods to study England's print culture in the pivotal sixteenth and seventeenth centuries. Machine learning synthesizes the content of 57,863 texts comprising 83 million words into 110 topics. Topics include the expected, such as Natural Philosophy, and the unexpected, such as Baconian Theology. Timelines suggest that religious and political discourse gradually became less antagonistic and economic topics more prominent. The epistemology associated with Bacon was present in theological debates already before Bacon's epistemological contributions. Vector autoregression estimates provide insight into the coevolution of ideas on religion, science, and institutions. Innovations in religious ideas stimulated focus on science, especially at times when Puritanism was prominent in religious discourse. Neither science nor institutional thought evidence secularization. The Glorious Revolution and the Civil War did not spur debates on institutions nor did the founding of the Royal Society markedly elevate attention to science.

**Keywords:** cultural history; machine learning; coevolution; religion; science

“I need not tell the world what (to their cost) they know, That Souldiers by action, and Printers by promulgation, are the two great English Factors.”  
Captain John Randolph, 1643.

## Introduction

In English cultural history, the sixteenth and seventeenth centuries were especially fecund. This era solidified the importance of the book as the dominant vehicle of culture, vastly increasing the flow of vernacular texts and allowing a much deeper

penetration of the written word into a wide variety of English-speaking classes and cultural groups (Barnard and McKenzie 2002). “Retrospectively, the history of the book in Britain from 1557 to 1695 looks like a triumphalist progress in which a dominant Protestant vernacular culture, and an emergent canon of English literature, were steadily created and successfully displaced an earlier Latinate and Catholic world looking towards Europe” (Barnard 2002: 1). With religious considerations permeating every aspect of early-modern English society, religious works, particularly those expressing Puritan sentiments, provided an important element of the output of the expanding publishing sector (Collinson et al. 2002; Green and Peters 2002).

At the same time, the early-modern period saw the flourishing of experimental science, as exemplified in the publications of Bacon, Harvey, and Boyle and in the formation of the Royal Society, which itself played an important role in the emergence of scientific publishing (Johns 2002). And fundamental texts on political philosophy and institutional thought appeared, epitomized by the works of Harrington, Locke, and Hobbes. England in the sixteenth and seventeenth centuries thus witnessed a cultural flowering in a multiplicity of domains, with works made all the more accessible by the growing publishing industry. With accessibility, there was increasing scope for a fertile intermingling of ideas, including those about religion, science, and institutions.

But what were the central emphases in the burgeoning English print culture of the sixteenth and seventeenth centuries and when did those emphases emerge? To what extent did ideas about religion, science, and institutions coevolve and when were such coevolutionary processes especially important? In these processes, how consequential were specific historical events, such as the Civil War, the Glorious Revolution, and the formation of the Royal Society? These questions have long intrigued historians and social scientists (e.g., Hill 1965; Merton 1938; Wootton 2015) and lately have increased in importance in economics as a result of works that view culture as central in England’s early economic rise (see, e.g., McCloskey 2016; Mokyr 2016).

Traditional textual analysis constitutes one approach to tackling the above questions. However, providing a balanced summary of a very large corpus is not possible using only the close reading that forms the foundation of traditional methods. Instead, given recent availability of digitized corpora such as the Early English Books Online-Text Creation Partnership (hereafter EEBO-TCP), the application of computational and statistical techniques for analysis of text-as-data (see, e.g., Gentzkow et al. 2019; Grimmer et al. 2021; Livermore and Rockmore 2019) provides an eminently feasible route to systematic investigation of a large corpus of cultural works.

Indeed, aspects of early-modern English cultural history have recently been examined using computational and statistical methods. In particular, in an innovative contribution, Erikson (2021) applies computational methods in an analysis of 2353 EEBO-TCP economics-related texts to illuminate how the actions and writings of merchants led to greater prominence of economic ideas in the early-modern period.<sup>1</sup>

---

<sup>1</sup>More recently, Almelhem et al. (2023) undertake a quantitative study of English texts from 1500–1900 but do not specifically focus on the early-modern era – presumably because their corpus, drawn from Hathitrust Digital Library, is very limited for the pre-1700 period (see Almelhem et al. 2023: Figure B.1). No other existing contribution applies computational and statistical techniques to offer systematic quantitative insight into early-modern English culture.

In this paper, we also adopt a computational approach to the investigation of early-modern English cultural history. Our perspective is distinctly macroscopic. Unlike Erikson (2021), we do not limit ourselves to a subset of the available EEBO-TCP texts or focus on specific ideas. Rather, we paint a broad-brush picture of the central cultural emphases discernible in the entire sixteenth- and seventeenth-century EEBO-TCP corpus. We then use macro-econometric methods to study the interactions between three cultural aggregates, which reflect the sets of ideas about religion, science, and institutions discovered in the corpus.

To this end, we first use topic modeling, an unsupervised machine learning method (Grimmer et al. 2021), synthesizing the content of 57,863 EEBO-TCP texts into 110 topics. We thereby provide a quantitative digest of the key substantive emphases in sixteenth- and seventeenth-century English print culture.

We then examine patterns within the dataset for the 110 topics that constitute that digest. We first study temporal changes in attention to specific topics. Our investigation thereby generates new stylized facts about early-modern cultural change and the timing of important cultural developments. Among the latter, debate about the emergence and origins of epistemological ideas later associated with Francis Bacon has been especially prominent (e.g., Grajzl and Murrell 2019; Martin 1992: 164–71; Shapiro 2000: 107–12). Our data and analysis casts direct empirical light on this debate. Given that Bacon has been credited as providing a central impetus to the development of a “culture of growth” (Mokyr 2016), an understanding of the emergence of epistemological thought associated with him provides insight into the cultural roots of England’s early economic rise.

In our primary econometric exercise, we investigate the coevolution of ideas on religion, science, and institutions. We focus especially on the interplay between religion and science, a connection that has been the subject of long-standing interest among social historians (e.g., Barnes 2000; Shapin 1996; Wootton 2015).

One especially prominent theory is the “Merton thesis” (Merton 1938) – that the cultural sentiments imbued by Puritanism played a vital role in facilitating the development of modern science. According to Shapin (1996: 136), variants of that thesis are broadly accepted: “Much of what Merton then wrote about religious motives to science, and religious justifications for science, has passed into historical commonplace.”<sup>2</sup> But the support is far from universal. Shapin (1988) comments on the hostile reception among many historians in the decades after Merton’s original publication. Brooke (1990) views Merton’s thesis as controversial and Porter (2000) does not even mention religion when discussing the culture of science in the seventeenth century. Wootton (2015: 908) approvingly cites the work of one of the hostile critics, who reaches the conclusion that “in the story of the rise of science, therefore, religion is a peripheral concern” (Rabb 1965: 126).<sup>3</sup> These differences surely mean that “English society, strongly influenced by Calvinism and deeply

<sup>2</sup>Webster (1974: 16) summarizes the first four decades of research by historians directly on Merton’s thesis: “It appears that protestantism, or Calvinism in particular, provided a favorable environment for the development of experimental science.” See also Becker and Woessmann (2009), Cohen (1990), Harrison (2008), Mokyr (2009, 2016).

<sup>3</sup>Cohen (1990) reviews the views of both proponents and critics. For a recent dismissal of Merton from a different milieu, literary studies, see Price (2017: 18).

involved in the development of science and scientific institutions, constitutes an ideal test case in which to examine the relations between science and religion” (Webster 1974: 15). Moreover, a new opportunity for tests arises with the advent of promising methods offered by the conjunction of machine learning and econometrics.

We employ vector autoregression (VAR), which is a standard technique for modeling the dynamic relationships among variables over time (see, e.g., Stock and Watson 2001), much employed in empirical macroeconomics. Our VAR analysis also allows us to illuminate whether the interplay between religion and science also occurred in the comparatively less studied direction, from science to religion.<sup>4</sup> Additionally, we are able to provide insight into the timing of when the influence from religion to science was especially strong. For instance, both Merton (1938) and Mokyr (2016) in this respect emphasize the seventeenth century, but our analysis provides insights into whether any such influence occurred even earlier.

Finally, our VAR estimates allow us to cast new light on several major seventeenth-century events: the Civil War, the Glorious Revolution, and the formation of the Royal Society. In economics especially, the role of the Glorious Revolution has been contested. A long-held theory emphasized the view that the Glorious Revolution was a watershed in the history of English institutional development (e.g., Acemoglu and Robinson 2012; North et al. 2009; North and Weingast 1989). Recent research has challenged this perspective, revealing a more gradual evolution of English institutions than emphasized by earlier contributions (Henriques and Palma 2023; Hodgson 2017; Murrell 2017; Ogilvie and Carus 2014). We bring new empirical evidence to this debate, illuminating the extent to which the developments in 1688–1689 resonated in English print culture and specifically in the emphases on institutions.

## The corpus

### *The EEBO-TCP documents and their processing*

Processing print texts from before 1700 entails many challenges: the inscrutable fonts immune to optical character recognition (OCR), the chaotic orthography, the archaic inflections, and the appearance of untranslated Latin. TCP (2022) has solved the first of these problems for a wide-ranging set of texts by using manual keying. This is important because no current OCR software produces satisfactory output for sixteenth- and seventeenth-century texts. Therefore, no alternative machine-readable corpus of commensurate breadth and depth exists in a form that could underpin the type of inquiry into pre-1700 English culture that we undertake. We therefore begin with the corpus of 60,331 texts available from EEBO-TCP, addressing the problems of orthography, inflections, and foreign words with our

---

<sup>4</sup>For example, in commenting that “the willingness of the Protestant leaders to have reason and experience ‘test’ all religious beliefs” and that “religious conceptions were . . . definitely integrated with sentiments basic to the contemporary science and philosophy: there was throughout a reciprocal interaction,” Merton (1938: 469, 434) allows for, but does not focus on, the flow of ideas from science to religion. Harrison (2008: 266) is explicit on this point: “It can be said that the close relations between science and religion during the seventeenth and eighteenth centuries resulted not only in the transmission of epistemic authority from religion to science, but also in the transfer of some of the properties of the new inductive sciences to religion.”

own Python programs. This subsection provides an overview of the steps we took in processing that data. Appendix A in Supplementary material provides details.

We removed all EEBO-TCP-inserted formatting symbols to begin with versions of the texts that were as close to the originals as possible. We then assigned a year of publication to each text using the information provided by EEBO-TCP. A very small number of texts could not be dated and were discarded.

We converted the non-standardized orthography that was common before the eighteenth century into standard modern orthography. Older-style inflections were modernized. We translated on a word-by-word basis those words that could not be found in a modern English dictionary and were readily identified as Latin. We then dropped documents that contained either an especially small number of words or an uncharacteristically high share of words that could not be matched to any word in the English dictionary even after the processing.

The ensuing corpus was imported into R and further processed using the `textProcessor` function. We thereby converted all words to lower case, applied the Porter (1980) stemming algorithm, and removed standard English stop words, numbers, words with fewer than three characters, words included in only one document, and punctuation. After the resultant processing, the final corpus consisted of 57,863 documents containing 83,337,912 letter-based strings (i.e., words).

### **Selection issues**

The EEBO-TCP project began with lists of works contained in prominent catalogs that “trace the history of English thought from the first book printed in English in 1475 through to 1700” (TCP 2022). The combined catalog comprises more than 125,000 works for which facsimiles of texts are available. Due to resource and other constraints, the texts that EEBO-TCP actually processed amounted to approximately one half of those. The project’s aim was simply “to key as many different works – as much different text – as possible” (TCP 2022). EEBO-TCP implemented the project in a manner in which the individual preferences of project partners, staff, and editors had a negligible impact on text selection.<sup>5</sup>

But, of course, the EEBO-TCP corpus does not provide a random sample of English culture. Much culture was not committed to print, with only a minority of the population in early-modern England literate.<sup>6</sup> Many texts will have been lost, with survival depending on how much subsequent generations valued the texts. TCP (2022) also focused on first editions. The EEBO-TCP corpus is thus best viewed as capturing new developments rather than reflecting the stock of texts in use at any juncture: our analysis is more likely to reflect the production of print culture than the consumption of print culture.

<sup>5</sup>Paul Schaffner (pers. comm., February 23, 2023), a TCP production manager, described the project as one in which “personal preferences of any kind had virtually no influence on selection. . . . [The] selector’s job was to go through [the tracking database] picking unique works in English, picking the earliest copy (assuming it was complete), avoiding Latin, and so forth. There was simply no room to introduce personal preference into this mechanical and tedious task. This was not a ‘craft’ operation; it was a ‘production’ shop.”

<sup>6</sup>Nevertheless, much popular culture was published with the purpose of being read to the illiterate (see, e.g., Collinson et al. 2002) or to provide compendia.

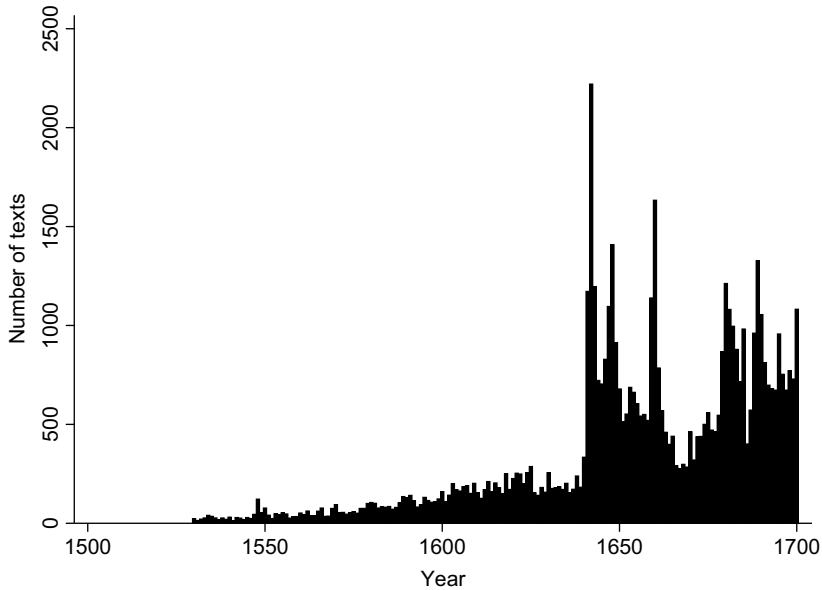


Figure 1. The distribution of EEBO-TCP texts over time.

Figure 1 shows the distribution over time of the texts in our corpus. Two features stand out. First, the growth from 1475 to 1640 reflects the growing importance of print. Second, the years with a large number of texts (e.g., 1642, 1660, 1689) are momentous ones in English history (the beginning of the English Civil War, the Restoration, and the resolution of the Glorious Revolution). None of the findings we reach in this paper simply reflect the number of texts published in any specific year.

### Producing a machine learning digest of English print culture

Our first objective is to generate a machine learning digest of the complete EEBO-TCP corpus. In this step, our unit of analysis is a document, a self-standing published work.<sup>7</sup> We use topic modeling, the standard natural-language-processing approach when the aim is to infer the core emphases in a large corpus (see, e.g., Grimmer et al. 2022, 2021).<sup>8</sup>

<sup>7</sup>We did not break documents into smaller chunks. As stressed by Grimmer et al. (2022: 50), “Splitting up a single text into many documents can substantially increase or decrease the computational cost of fitting a model. It also may make the model more or less statistically efficient.” Importantly, most EEBO-TCP texts do not have readily-identifiable chapters or even paragraphs that would provide natural break points. Thus, splitting EEBO-TCP documents into smaller chunks would introduce ad-hoc breaks between the chunks, an approach that is, in our context, unlikely to improve model fit.

<sup>8</sup>Another natural-language-processing method that is gaining popularity in the analysis of texts, word embeddings, emphasizes semantics and fine-grained representation at the word level. As such, word embeddings would have been a less suitable method than topic modeling for accomplishing the goals of this paper.

Topic modeling is an unsupervised machine learning method (see, e.g., Burkov 2019). The algorithm identifies topics by leveraging patterns in word use across the corpus documents. Each document is conceptualized as a mixture of all topics, while topics are probability distributions over the corpus vocabulary. We estimate a structural topic model (henceforth STM; Roberts et al. 2014, 2016a), a variant of topic modeling that directly incorporates document-level metadata (e.g., publication year) into the estimation to aid topic identification. To implement the estimation, we use R's *stm* package (Roberts et al. 2019).

### *Choosing the number of topics*

Before estimation, the number of topics must be chosen. There exists no universally agreed-upon methodology for making this decision (see, e.g., Grimmer et al. 2022: Ch. 13; Wang et al. 2019: 258–59). We first estimated a series of STMs by varying the number of topics between 10 and 200. We then examined standard measures of goodness-of-fit such as held-out likelihood and size of residuals (see, e.g., Roberts et al. 2016b; Taddy 2012; Wallach et al. 2009). The model with 110 topics fit the data well, with further increases in the number of topics producing only modest gains. We also directly compared the 110-topic model with models featuring fewer and more topics. None of the alternative models dominated the 110-topic model on standard criteria (see, e.g., Airoldi and Bischof 2016; Weston et al. 2023): semantic coherence (measuring the internal consistency of the topics) and exclusivity (measuring whether topics can be easily distinguished). The decision to use a 110-topic model was confirmed when we contrasted the output of that model with the output of models with different numbers of topics by reflecting on the ease of interpreting topics and distinguishing between them.<sup>9</sup>

### *Interpreting the estimated topics*

To interpret and name the estimated topics, we examined both the word-stems most highly associated with each topic and those documents that featured the topic most prominently.<sup>10</sup> Our interpretation and naming of the estimated topics therefore also incorporated an element of the close reading typical of conventional text analysis.

Appendix B in Supplementary material describes the content and justifies the assigned names for the 110 topics. Here, we briefly illustrate our process of topic interpretation and naming by using only one example. For the pertinent topic, the top word-stems (those most associated with the topic) emphasize logical connectives: upon, yet, though, thus, mean, inde [thereunto], even, impli [imply]. The top documents (those featuring this topic in the greatest proportion) often contain the word experiment. Nearly all top documents focus on religion using logical arguments and emphasizing inductive learning from facts: biblical, historical, or personal. The top document states that “if a man were but well read in the story

---

<sup>9</sup>We also implemented the Lee and Mimno (2014) algorithm that provides a non-statistical perspective on the number of topics. Running the algorithm for 10 different seeds, the average number of suggested topics was 97 but the maximum was 109, suggesting that 110 topics would be sufficient.

<sup>10</sup>We routinely examined the top 40 documents, but some topics required reading a greater number.

and various passages of his life, he might be able to make an experimental divinitie of his own. He that is observant of Gods former dealings and dispensations towards him, may be thence furnished with a rich treasury of experience against all future conditions.” Accordingly, we name the topic Baconian Theology.<sup>11</sup>

Via this process we were able to readily identify the ideas underlying all 110 estimated topics. This is, in itself, a verification of the quality of text preprocessing, the choice of the number of topics, and the applicability of STM. The topics include both the innocuous and familiar-sounding ones, as well as more unusual ones, all clearly dictated by our estimates. Thus, Salvation via Faith competes for attention with Lusty Entertainments; Emotional Relationships contrasts with Expressing Loving & Loathing; Deductive Theology appears alongside Baconian Theology. The most prevalent topic is an inward-looking one, Self-Reflection, accounting for 4.32 percent of the corpus, while the second most prevalent is other-directed Petitions, Protests, & Proposals, occupying 3.42 percent.

Table 1 lists the names of all topics together with the percentages of the corpus occupied by each. These names, being necessarily brief, might not convey the full meaning of the topic. Appendix B in Supplementary material provides more detailed information on each topic.

### **Grouping topics into broader themes**

The 110 estimated topics constitute one window into nearly two centuries of cultural change. Another perspective emerges from even more aggregation. We grouped the topics into a smaller number of broader themes. Following recent contributions faced with an analogous task (see, e.g., Gennaro and Ash 2022; Grajzl and Murrell 2023), we assigned topics to themes manually, based on our own understanding of the topics.<sup>12</sup>

In grouping topics into themes, we did not pre-commit to a fixed number of themes but rather allowed the number to emerge from the aggregation process. We

<sup>11</sup>We use the adjective “Baconian” as an indicator of the general style and content of ideas, not to signify that Bacon directly contributed. The term Baconian Theology was not used at the time, nor is it used by historians today to characterize ideas existing at that time. This term has been used to describe a strain of thought in American Presbyterianism that arose in the nineteenth century and that is similar to our topic (Holifield 2003). Natural theology was commonly used during the sixteenth- and seventeenth-centuries and by historians through the ages to refer to modes of thought combining elements of scientific and theological musings. We chose not to use natural theology for this topic because that term encompasses a broader area of thought than our Baconian Theology, including purely deductive theology, which is another of our topics. Physico-theology is much closer to our Baconian Theology. But experiment and induction are not central to physico-theology (Blair and von Greyerz 2020) whereas they are central in our Baconian Theology. Finally, the apogee of the development of Baconian Theology in our data occurs in the 1650s, whereas physico-theology only begins to stir in the 1650s (Calloway 2015).

<sup>12</sup>An alternative approach to grouping topics into themes would have been to use a clustering algorithm and cluster topics based on their document-level co-occurrence (see, e.g., Erikson 2021: Ch. 1). We chose not to pursue this route for two key reasons. First, substantively related topics do not need to co-occur at the document level. For example, in our data, Authority Relationships and Emotional Relationships, both clearly about relationships, do not co-occur at the document level: the pertinent pairwise correlation coefficient equals  $-0.12$ . Second, our methodological perspective is based on employing machine learning when necessary (creating a 57,863-by-110 topic-document matrix from 83,337,912 letter-based strings) but using judgment and existing knowledge where practical (aggregating 110 topics into fewer themes).



**Table 1.** Topics and themes, with expected document-level prevalences

<i>History</i> (2.53)	<i>Geography</i> (1.35)	<i>Religion</i> (22.04)
Ecclesiastical History (0.23)	Voyages (0.83)	Holy Days (0.45)
Monarchs' Lives (0.84)	Africa & Asia (0.12)	Allaying Christian Doubt (1.38)
Biblical History (0.55)	Western Europe Surveyed (0.17)	Official Prayer (1.57)
Analyzing Ancient Thought (0.24)	Chorography (0.23)	Establishing Correct Doctrine (0.8)
French & Iberian History (0.31)	<i>Literature</i> (9.77)	Salvation via Virtue (1.63)
British Isles History (0.13)	Early Literature, Secular (0.16)	Papacy (0.25)
Roman History (0.23)	Early Literature, Spiritual (0.45)	Salvation via Faith (0.68)
<i>Institutions</i> (12.64)	Historical Romance (0.39)	Old Testament (0.29)
Constitutional Rules (0.74)	Jokes & Mishaps (2.29)	Attacking False Doctrine (2.37)
Ordinances (2.72)	Lusty Entertainments (1.95)	Apocalyptic Theology (0.63)
Parliamentary & Court Procedure (1.05)	Romance & Comedy Drama (0.69)	Counter-Reformation Scholarship (0.7)
Scottish Political Acts (0.88)	French Romance (0.29)	Catholics Under Protestant Rule (0.38)
Local Law (0.99)	Chivalric Literature (0.14)	Religious Love & Hope (1.43)
Scottish Law (0.17)	Poetic Laments (1.73)	Baconian Theology (1.81)
Autonomous Church Governance (0.84)	Comedy, Satire, Epigrams (0.59)	Transubstantiation (0.46)
Dignitary Law (0.07)	Political Poems (1.09)	Sin, Damnation, & Repentance (1.71)
Royal Proclamations (3.16)	<i>Science</i> (3.14)	Dissent, Schism, & Toleration (1.21)
Lawfulness (1.37)	Mathematics (0.17)	Reasonable Religious Discourse (1.61)

(Continued)

Table 1. (Continued)

Compilations of Laws (0.12)	Astronomy & Astrology (0.32)	Deductive Theology (1.61)
Hierarchical Church Governance (0.53)	Natural Philosophy (0.36)	Christian Mysticism (0.63)
<i>Miscellany</i> (3.02)	Botany (0.05)	Holy Lives (0.44)
Art Catalogs (0.48)	Zoology (0.41)	<i>Skills</i> (10.22)
Names Listed (1.19)	Medical Science (0.27)	Commercial Aids (0.59)
Listing City Facts (0.4)	Chemistry (0.24)	Equine Learning (0.17)
Not English (0.28)	Anatomy (0.09)	Military Training (0.53)
Non-Translated Latin (0.67)	Pharmacology (0.22)	Scholarly Learning Guides (0.6)
<i>Philosophy</i> (5.37)	Physics (0.26)	Practical Geometry (0.23)
Republicanism (0.96)	Diet & Health (0.49)	Student & Practitioner Law (0.35)
Moral Philosophy: Interests (1.68)	Diseases & Cures (0.26)	Cookery (0.29)
Moral Philosophy: Passions (1.18)	<i>Politics</i> (18.27)	Using Numbers (1.07)
Moral Philosophy: Virtues (0.77)	Military Campaigns (2.28)	Industrial Arts (0.39)
Deductive Reasoning (0.78)	Defending Monarchs (0.84)	Legal Practice Aids (0.44)
<i>Relationships</i> (11.66)	Continental International Relations (0.56)	Bibliographical Practice (0.64)
Family Matters (0.69)	High Crimes (1.21)	Improving Soil & Its Products (0.53)
Authority Relationships (1.23)	Allegiance & Resistance (1.14)	Rural Recreations (0.1)
Emotional Relationships (2.32)	Fearing Catholicism (2.02)	Catechismal Compilations (1.66)
Expressing Loving & Loathing (0.77)	Economic Lobbying (2.02)	Christian Mental Exercises (1.19)
Self-Reflection (4.32)	Turkish International Politics (0.25)	Practicing Christianity (1.44)

(Continued)

**Table 1.** (Continued)

Obloquy & Encomium (2.33)	Asserting Parliamentary Powers (2.88)
	Political Uses of Religion (1.65)
	Petitions, Protests, & Proposals (3.42)

Notes: The table lists the 110 STM-estimated topics (non-italicized) as interpreted by the authors. (Appendix B provides the lists of keywords most prominently associated with each topic and a detailed justification for each topic name.) The topics have been grouped into 11 themes (italicized). The numbers in parentheses are expected document-level prevalences, computed as simple (non-weighted) report-level means of the STM-estimated topic and theme prevalences, all expressed in percentages.

strove to create themes that were broad enough to achieve a further reduction in dimensionality and narrow enough to resonate with major areas of cultural inquiry. In this process, we assigned each topic to one theme only, using the criterion of the most natural fit. Classifications were straightforward for the overwhelming majority of topics (e.g., Sin, Damnation, & Repentance to religion; Chemistry to science; Constitutional Rules to institutions). However, culture is a seamless web and thus a small set of topics lie close to the boundaries between themes. For example, Autonomous Church Governance and Hierarchical Church Governance were close to the religion and institutions themes. Ultimately, given their emphasis on organization under the law rather than on religion, we allocated these topics to institutions.<sup>13</sup>

The resultant 11 themes differ considerably with regard to the number of included topics and the proportion of the corpus occupied (see Table 1). Unsurprisingly, religion accounts for a large number of topics (20) and a large proportion of the corpus (22 percent). Among religious topics, a number reflect areas of religious thought to which Puritans made major contributions (e.g., Apocalyptic Theology; Salvation via Faith; Baconian Theology; Christian Mental Exercises; Deductive Theology; Self-Reflection). Others reflect more general religious controversies in which Puritans participated in debate (Transubstantiation; Salvation via Virtue).

## Characterizing cultural evolution temporally

In this section we provide insights into how attention to various areas of cultural discourse changed over time. To this end, we first describe the construction of the pertinent time series. We then offer a set of observations on particularly interesting episodes identified in the corresponding timelines. The resulting observations provide a unique macroscopic overview of the evolution of a variety of cultural domains.

### *Constructing the time series of attention to topics and themes*

The core output of our topic model is a  $57,863 \times 110$  topic-document matrix. An element of that matrix,  $\theta_{idt}$ , gives the estimated prevalence of topic  $i$  in document  $d$  published in year  $t$ . To construct annual time series of attention to the topics, we merge the topic-document matrix with metadata on publication year and document length and compute mean yearly attention to each topic. We use weighted means, weighting those documents with a greater number of words more heavily.<sup>14</sup> Let  $w_{dt}$  denote the number of words in document  $d$  published in year  $t$ . The attention to topic  $i$  in year  $t$  is then:

<sup>13</sup>None of the central qualitative findings in this paper depend on the chosen classification.

<sup>14</sup>The weighting of document-level relative topic prevalences with document-level word counts addresses the fact that documents are of varying lengths. A simple (unweighted) average of the relative prevalences of a given topic in different documents for a particular year would effectively discount the importance of the longer documents, thereby underestimating the true attention in the corpus to the pertinent cultural ideas. This weighting is also necessary because some long documents are compilations, for example collections of sermons, and therefore contain more than one contribution highlighting the pertinent topic.

$$\Psi_{it} = \frac{\sum_{d \in D_t} w_{dt} \theta_{idt}}{\sum_{d \in D_t} w_{dt}}, \quad (1)$$

where  $D_t$  is the set of all documents published in year  $t$ .  $\Psi_{it}$  captures the proportion of the corpus occupied by topic  $i$  in year  $t$ ,  $i = 1, \dots, 110$ ,  $t = 1530, \dots, 1700$ .<sup>15</sup> Figures containing the timelines for all  $\Psi_{it}$  appear in Appendix C in Supplementary material.

Timelines for the themes are constructed using the same principles. For each theme  $m$  in Table 1, the average attention to that theme in year  $t$  is:

$$\Theta_{mt} = \frac{\sum_{d \in D_t} w_{dt} \sum_{i \in S_m} \theta_{idt}}{\sum_{d \in D_t} w_{dt}}, \quad (2)$$

where  $S_m$  is the set of topics comprising theme  $m$ ,  $m = 1, \dots, 11$ ,  $t = 1530, \dots, 1700$ . Appendix D in Supplementary material provides figures depicting the resultant timelines.

### **Some stylized facts about cultural change gleaned from the topic timelines**

For a particular topic, periods of rising attention to that topic indicate increasing importance of the corresponding ideas in the cultural discourse. In contrast, times of declining attention suggest a waning interest in the applicable ideas in print culture. Examination of the topic timelines therefore offers a path to establish stylized facts about cultural change. Below, we highlight a subset of the facts that are, in our view, the most intriguing ones.<sup>16</sup>

First, Baconian Theology appears already in the 1570s (i.e., when Bacon was still a teenager) and rises in importance over the next 80 years (Figure 2(a)). Therefore, even before Bacon's epistemological contributions, theological debates had featured ideas that were later to be associated with Bacon, particularly the importance of experiments or experience to pursue induction. Our results thus suggest that Bacon's name provided a particularly convenient label for a set of ideas that were already part of the existing culture: his writings served as an important "coordination device" for later thinkers (Mokyr 2016: 73). Hence, our findings suggest a view of Bacon as very much a product of his times, consistent with Harkness (2007) and Grajzl and Murrell (2019). Interestingly, in contrast to the increasing attention to Baconian Theology in the seventeenth century, the more conventional and austere Deductive Theology steadily wanes from the close of the sixteenth century onward (Figure 2(b)).

Second, among other religious topics as well as institutional and political topics touching upon religion, there is clear evidence of a shift toward a less antagonistic

<sup>15</sup>In this period, each year has a significant number of documents.

<sup>16</sup>Importantly, many features of the topic timelines resonate with well-known historical events, a feature that lends credibility to our approach of examining the topic timelines to generate stylized facts about early-modern cultural change. For example, attention to Constitutional Rules rises in the 1550s (the tumultuous Tudor period ending in the Elizabethan Religious Settlement of 1559), 1640s (the Civil War era) and 1680s (the decade culminating in the Glorious Revolution). Parliamentary & Court Procedure receive above-average attention immediately before the Bill of Rights (1689) and the Act of Settlement (1701). Turkish International Politics peaks with the siege of Vienna (1683). See Appendix C.

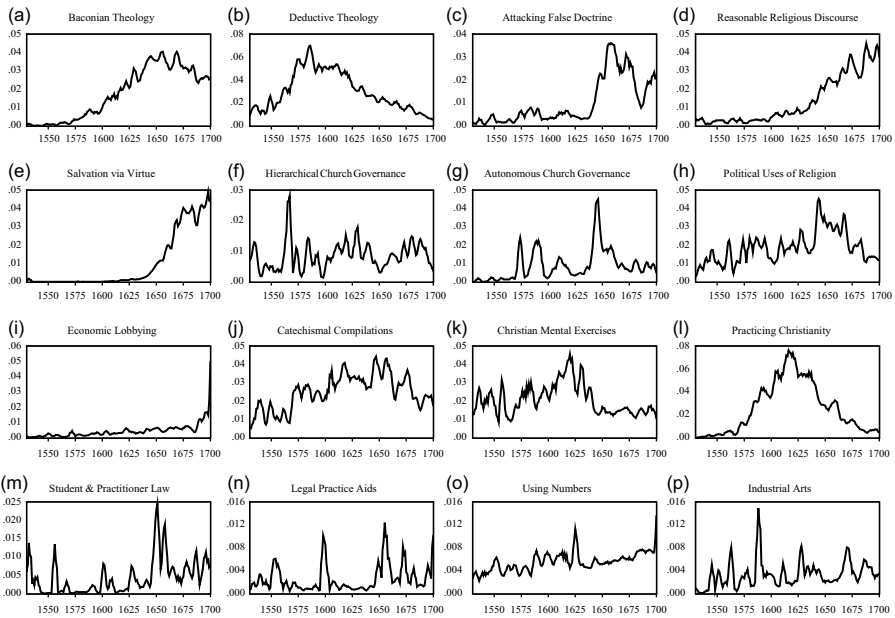


Figure 2. Selected topic timelines.

form of debate as the seventeenth century progresses. For example, Attacking False Doctrine, which characterizes the beliefs of others with venom and hatred, declines, while a topic that captures debate conducted in non-antagonistic tones, Reasonable Religious Discourse, increases (Figure 2(c) and (d)). Salvation via Virtue, which emphasizes the importance of good works such as charity, rises continually from the mid-seventeenth century (Figure 2(e)). Baconian Theology, also an expression of reasoned views in looking for evidence about how to interpret the world and live a good life, rises to prominence in the second half of the seventeenth century (Figure 2(a)). Hierarchical Church Governance peaks earlier than Autonomous Church Governance and the prevalence of each declines (Figure 2(f) and (g)), indicating that the long struggles between the corresponding sets of ideas became less important. Political Uses of Religion, which expresses judgment of political classes from a religious stance, similarly declines after 1650 (Figure 2(h)).

Third, after 1688, there is a large rise in Economic Lobbying (Figure 2(i)), a finding resonating with Erikson's (2021) emphasis on the growing importance of economic ideas in the early-modern period. At the same time, within skills, those associated with religion (Catechismal Compilations; Christian Mental Exercises; Practicing Christianity) decline in the latter half of the seventeenth century, while those especially relevant to commerce increase: Student & Practitioner Law, Legal Practice Aids, Using Numbers, and Industrial Arts (Figure 2(j)–(p)).

Viewed as a whole, these patterns suggest that the eighteenth-century relative calm that facilitated economic progress was anticipated by cultural changes that appeared in the latter part of the bellicose seventeenth century: our timelines evidence a turn to a more measured religious and political debate, as well as an

increasing emphasis on economic ideas and skills relevant to commercial matters. The cultural origins of the eighteenth-century “nation of shopkeepers” can be clearly seen in the seventeenth century.

### The coevolution of ideas about religion, science, and institutions

We next turn to the investigation of the coevolution of ideas on religion, science, and institutions. To this end, we first use our STM estimates to construct the pertinent time series. We then lay out a plausible empirical model and show how that model can be analyzed using standard econometric techniques.

#### The time series

We construct annual time series of attention to the three themes of interest: religion, science, and institutions. With each theme defined as an aggregate of our STM topics (see Table 1), we must first decide whether to form aggregates of total attention or relative attention to the pertinent topics. Whereas total attention captures the total number of words devoted to a topic in a given year, relative attention reflects the proportion of words. We choose total attention because it better reflects the notion of expansion, and thus flow, of culture.<sup>17</sup> Moreover, since we are not interested in cultural accumulation that arises purely from a larger population, we measure total attention in per capita terms. Finally, to moderate the influence of outliers, we take the natural logarithm of per capita attention. Thus, we use the following measures of attention to religion, science, and institutions in year  $t$ :

$$relig_t = \ln\left(\frac{\Theta_{religion,t} \sum_{d \in D_t} w_{dt}}{pop_t}\right) \quad (3a)$$

$$sci_t = \ln\left(\frac{\Theta_{science,t} \sum_{d \in D_t} w_{dt}}{pop_t}\right) \quad (3b)$$

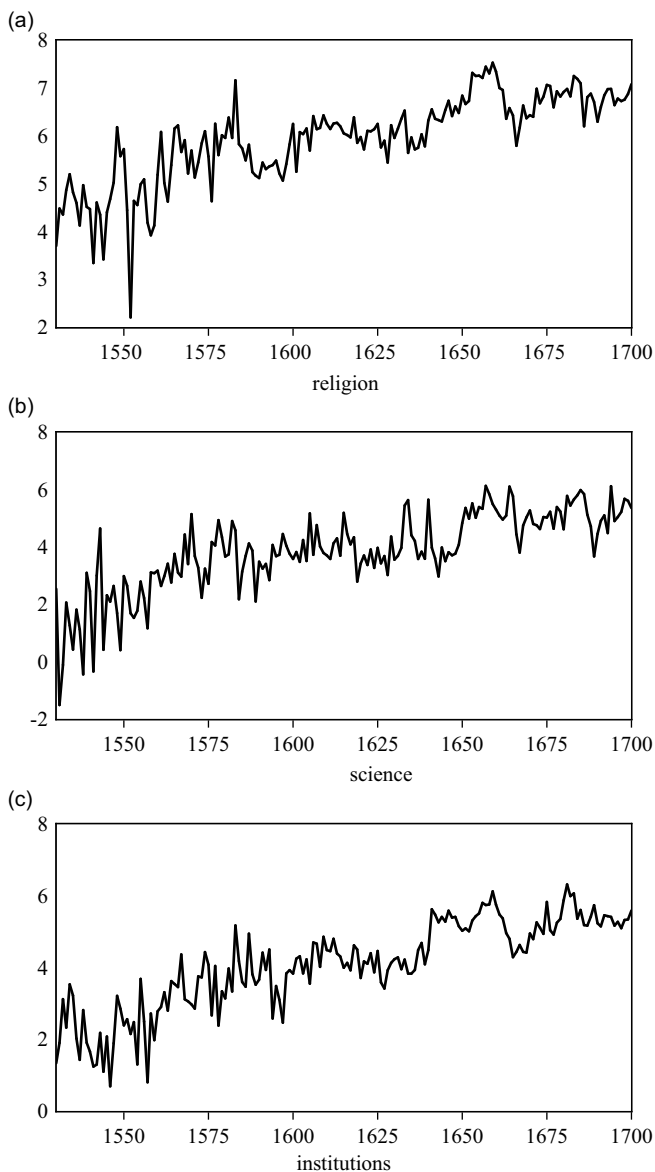
$$inst_t = \ln\left(\frac{\Theta_{institutions,t} \sum_{d \in D_t} w_{dt}}{pop_t}\right), \quad (3c)$$

where  $\Theta_{mt}$ , the average attention to theme  $m \in \{religion, science, institutions\}$  in year  $t$ , is defined in expression (2).  $w_{dt}$  is the number of words in document  $d$  published in year  $t$ , and  $D_t$  is the set of all documents published in year  $t$ .  $pop_t$  is England's population in year  $t$ .<sup>18</sup>

Figure 3 plots the three time series defined by (3a), (3b), and (3c). Table E.1 in Appendix E in Supplementary material presents the descriptive statistics.

<sup>17</sup>If our time series captured relative attention, then a change in the value of a theme might merely reflect a change in attention to another theme, as opposed to a substantive change in attention to the theme under consideration.

<sup>18</sup>See Broadberry et al. (2015) for the data on population, which is measured in millions.



**Figure 3.** The time series of logged per capita attention to religion, science, and institutions.

### **Empirical model**

Our goal is to investigate the interrelated dynamics of ideas on religion, science, and institutions. In particular, we wish to allow the possibility of two-way effects between each pair of the series  $relig_t$ ,  $sci_t$ , and  $inst_t$  defined in (3a)–(3c). We therefore study the behavior of the three-variable vector



$$\mathbf{y}_t = \begin{bmatrix} \text{relig}_t \\ \text{sci}_t \\ \text{inst}_t \end{bmatrix}. \quad (4)$$

Importantly, the value of each element of  $\mathbf{y}_t$  plausibly depends on its own past values as well as on past values of the other elements of  $\mathbf{y}_t$ . Specifically, the “normal” process of reaction and counter-reaction of ideas in publications makes  $\mathbf{y}_t$  a function of  $\mathbf{y}_{t-1}$ ,  $\mathbf{y}_{t-2}$ ,  $\dots$ . This reflects the notion that heightened attention in one area of print culture at time  $t$  might change the amount of attention to all areas after  $t$ .

But  $\mathbf{y}_t$  is affected by more than this process of lagged reaction and counter-reaction. There are shocks in the form of new ideas that are not a product of the normal process of response to past publications. These are one-time, idiosyncratic or “abnormal,” changes in  $\mathbf{y}_t$  reflecting extrinsic factors: a divine revelation might cause speculation in theology; discovery of a new continent opens up study of new scientific phenomena; an unexpected development in caselaw spurs debate on new constitutional structures.

Finally, there are secular changes in all elements of  $\mathbf{y}_t$ . In sixteenth- and seventeenth-century England, the overall attention in publications to ideas on religion, science, and institutions rose steadily (see Figure 3) due to a variety of factors. These include improvements in printing technology and an increasingly thriving vernacular print culture (Barnard 2002), population growth (Merton 1938: 570–5), improved access to ideas via international commerce (Palma 2016: fn. 25), an expanding influence of the merchant class (Erikson 2021), and early preindustrial economic development (Crafts and Mills 2017).

The structure described above fits a vector autoregressive (VAR) model, which captures the intertwined dynamics of multiple time series (see, e.g., Kilian and Lütkepohl 2017; Sims 1980; Stock and Watson 2001). A cornerstone of macroeconomic methods, VAR has been employed productively by economic historians.<sup>19</sup>

We posit the following structural VAR:

$$\mathbf{A}\mathbf{y}_t = \mathbf{\Gamma}_0 + \sum_{i=1}^3 \mathbf{\Gamma}_i \mathbf{y}_{t-i} + \mathbf{D}t + \mathbf{u}_t \quad (5)$$

where  $\mathbf{y}_t$  is defined above.  $\mathbf{\Gamma}_0$  is a  $3 \times 1$  vector of constants. The  $\mathbf{\Gamma}_i$ ,  $i \in \{1, 2, 3\}$ , are  $3 \times 3$  matrices of coefficients. The model with three lags was selected on the basis of conventional lag length criteria and tests (see, e.g., Kilian and Lütkepohl 2017).<sup>20</sup>  $t$  is a linear time trend, capturing the above-noted secular trends in  $\mathbf{y}_t$ .  $\mathbf{D}$  is the corresponding  $3 \times 1$  vector of coefficients. Importantly, the inclusion of the time trend (implicitly detrending the data) implies that our empirical analysis illuminates

<sup>19</sup>See, for example, Ben Zeev et al. (2017), Grajzl and Murrell (2023), Jalil (2015), Kenny et al. (2021), Møller and Sharp (2014), and Quinn and Roberds (2019). A subset of recent empirical contributions in economic history instead estimates the effects of interest (typically impulse-responses) using local projections while relying on data that provides a source of exogenous variation; see the subsection on identification assumptions below for references. Importantly, if the underlying economic model is the same, VAR and local projections estimate the same impulse-responses (Plagborg-Møller and Wolf 2021).

<sup>20</sup>The sequential modified likelihood-ratio, final prediction error, Akaike information criterion, and Hannan-Quinn information criterion tests univocally dictate the choice of three lags. Only the Schwarz information criterion suggests a model with a single lag. We thus chose the model with three lags.

the variability in attention to each of the three cultural domains around their long-run trends. Finally,  $\mathbf{u}_t$  is a  $3 \times 1$  vector of orthogonal structural shocks, with  $E(\mathbf{u}_t \mathbf{u}_t') = \mathbf{I}_3$ .

### Identification assumptions

In model (5), the  $3 \times 3$  coefficient matrix  $\mathbf{A}$  captures how shocks occurring in one domain can immediately affect attention in all domains. Thus, the model (5) allows each element of  $\mathbf{y}_t$  to depend on the contemporaneous values of the other elements of  $\mathbf{y}_t$ . This possibility means that the parameters of (5) are not identified without additional assumptions. However, under reasonable assumptions, estimates of those parameters can be derived from ordinary least squares (OLS) estimates of the following:

$$\mathbf{y}_t = \mathbf{A}^{-1} \mathbf{\Gamma}_0 + \mathbf{A}^{-1} \sum_{i=1}^3 \mathbf{\Gamma}_i \mathbf{y}_{t-i} + \mathbf{A}^{-1} \mathbf{D}t + \mathbf{e}_t \quad (6)$$

where  $\mathbf{e}_t \equiv \mathbf{A}^{-1} \mathbf{u}_t$ .

Our approach to making the requisite assumptions rests on short-run restrictions, a bedrock of the VAR literature in macroeconomics (see, e.g., Caldara and Iacoviello 2022; Christiano et al. 1999; Ramey 2016).<sup>21</sup> Specifically, we assume that (A1) shocks to  $sci_t$  or  $inst_t$  do not contemporaneously impact  $relig_t$  and (A2) shocks to  $inst_t$  do not contemporaneously affect  $sci_t$ . In Appendix F in Supplementary material, we provide a detailed justification of these assumptions in light of the nature of our data and the workings of English society in the time period under consideration. But we do not claim to provide the last word on these issues. Issues of causality among multiple endogenous variables are notoriously difficult to solve, as attested by modern-day macroeconomics (Nakamura and Steinsson 2018). Our approach – the first ever applied to the coevolution of religion, science, and institutions in a historical setting – offers one route to do so within a viable estimation framework.

Given assumptions (A1) and (A2), justified in Appendix F in Supplementary material,  $\mathbf{A}$  in (5) and, consequently,  $\mathbf{A}^{-1}$  in (6) are lower triangular and the residuals from the reduced-form VAR in expression (6) can be expressed as:

$$\mathbf{e}_t \equiv \begin{bmatrix} e_t^{relig} \\ e_t^{sci} \\ e_t^{inst} \end{bmatrix} = \begin{bmatrix} \tilde{a}_{11} & 0 & 0 \\ \tilde{a}_{21} & \tilde{a}_{22} & 0 \\ \tilde{a}_{31} & \tilde{a}_{32} & \tilde{a}_{33} \end{bmatrix} \begin{bmatrix} u_t^{relig} \\ u_t^{sci} \\ u_t^{inst} \end{bmatrix}. \quad (7)$$

With this form, estimates of the structural parameters of (5) can be derived from OLS estimates of (6). Importantly, one can readily obtain estimates of  $\mathbf{u}_t = \mathbf{A} \mathbf{e}_t$ , the vector of structural shocks. As we stress below, these estimated shocks add new

<sup>21</sup>An alternative approach would be based on, first, identifying plausibly exogenous variation via the collection of external data and, second, using such variation to assess the effects on outcomes of interest (Nakamura and Steinsson 2018). For examples of innovative identification strategies of this kind in economic history, see Brzezinski et al. (2022), Jordà et al. (2022), and Palma (2022). Although this alternative approach has many advantages, it can only be implemented if one is successful in finding additional time series that capture truly exogenous variation. We do not know of data that could play this role for attention to ideas on religion, science, and institutions.

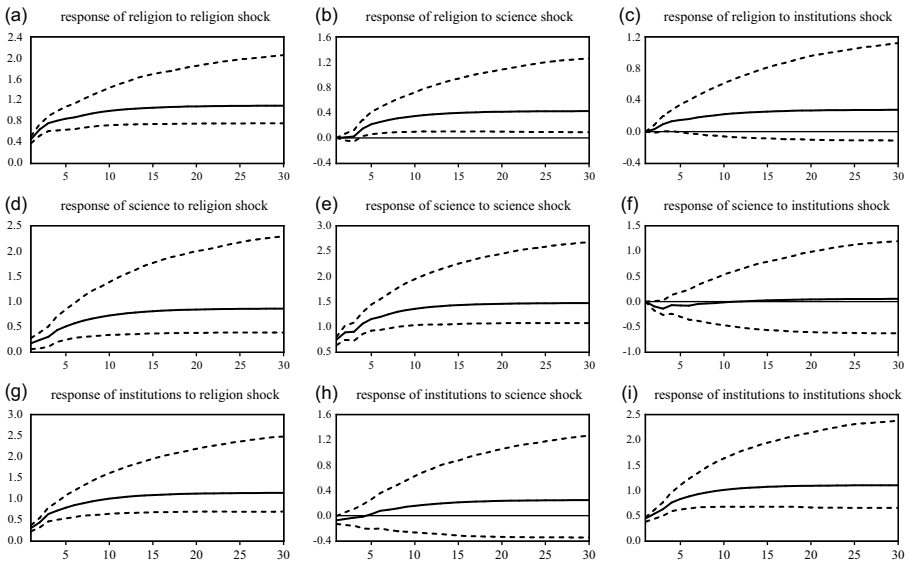


Figure 4. The accumulated impulse-responses.

information to the historical record on the occurrence of intellectual developments that our estimates suggest were not simply the product of reactions to earlier intellectual developments.

### *Impulse-responses: a first look at coevolutionary dynamics*

We first investigate the coevolution of ideas about religion, science, and institutions using impulse response functions (IRFs). The IRFs summarize the average expected response of each element of  $y_{t+s}$  at time horizon  $s \geq 0$  following a one-time shock to one specific element of  $y_t$ . We model the initial change as a one-time, one-standard-deviation structural shock (i.e., a shock of typical size) that elevates the attention to the pertinent theme (i.e., raises the value of the applicable element of  $y_t$ ), a manifestation of an exogenous “innovation” in culture. We estimate cumulative IRFs in attention to each of the three areas over a 30-year period, reporting 90-percent confidence intervals.<sup>22,23</sup>

Figure 4 summarizes the results. In each of the nine subfigures, the horizontal axis shows the number of years since the shock. The vertical axis measures the

<sup>22</sup>Confidence-intervals are computed using Kilian’s (1998) bootstrap method. The use of 90 percent is more conservative than is typical in the macroeconomics literature, which usually uses one-standard-deviation confidence intervals.

<sup>23</sup>In Appendix G we present the full set of impulse-responses obtained using the method of local projections (LP), an alternative approach to estimating the impulse-responses. To generate the LP-estimated impulse-responses, we use the same identification assumptions as in the VAR reported here. Appendix G shows that the VAR-based and the LP-based impulse-responses paint a qualitatively consistent picture on the coevolution of cultural ideas.

cumulative change in the pertinent series at each horizon, expressed in proportions since our series are measured in natural logs (see expressions (3a)–(3c)). For interpretation of Figure 4, it is helpful to know that a shock to  $relig_t$  causes an immediate jump in  $relig_t$  (i.e., increases attention to ideas about religion) of 45 percent. Similarly, a shock to  $sci_t$  immediately increases  $sci_t$  by 75 percent and a shock to  $inst_t$  immediately elevates  $inst_t$  by 44 percent.

The subfigures along the main diagonal (Figure 4(a), (e) and (i)) capture the extent to which a shock in one domain spurs subsequent developments in the same domain. The corresponding IRFs are positive and statistically significant at all horizons: innovations in a given cultural domain do not dissipate but rather lead to long-lasting elevated attention to that same domain. For example, following a shock to  $sci_t$ , the 30-year effect on  $sci_t$  is twice the size of the immediate rise in  $sci_t$  (Figure 4(e)). This effect corresponds to the mechanism alluded to in Merton's (1938: 436–69) theorizing, that “a fixed order must prevail in the appearance of scientific discoveries; each discovery must await certain prerequisite [scientific] developments.”

The subfigures off the main diagonal show how innovations (shocks) in one cultural domain affect future developments in the other two domains. Most importantly, innovations in ideas on religion spur strong responses in attention to the other two cultural domains (Figure 4(d) and (g)). The 30-year-effect of a shock to  $relig_t$  on  $sci_t$  is a 109 percent rise in  $sci_t$  (Figure 4(d)). Thus, religious thought was central in spurring developments in ideas about science as theorized by Merton (1938: 440): “It was precisely Puritanism which built a new bridge between the transcendental and human action, thus supplying a motive force for the new science.”

Innovations in ideas about science elevate attention to ideas about religion (Figure 4(b)): the 30-year-effect of a shock to  $sci_t$  on  $relig_t$  is a 43 percent rise in  $relig_t$ . The coevolution between ideas on religion and ideas on science was therefore a “reciprocal interaction” (Merton 1938: 434), but especially strong from religion to science (Merton 1938; Shapin 1996).

Innovations in ideas about religion also stimulate ideas about institutions: the 30-year-effect of a shock to  $relig_t$  on  $inst_t$  is a 28 percent increase in  $inst_t$  (3(g)). This finding resonates with the recent scholarship emphasizing the general relevance of religious ideas for long-run institutional development.<sup>24</sup>

In contrast, innovations in ideas about institutions lead to a relatively small and statistically insignificant increase in attention to ideas about religion (Figure 4(c)). There is no effect of innovations in ideas about institutions on attention to ideas about science (Figure 4(f)).<sup>25</sup>

<sup>24</sup>For example, Kuran (2010) and Rubin (2017) emphasize the importance of religious ideas for long-run comparative institutional and economic development. Dittmar and Seabold (2021) document a positive relationship between Protestant ideas in print and city-level institutional change in German-speaking Europe in the fifteenth and sixteenth centuries.

<sup>25</sup>Recall that the institutions theme does not reflect developments in caselaw, the primary source of legal developments in England at this time.

### The temporal incidence of shocks and relevance of revolutions

Shocks to  $y_t$  (the  $u_t$  in (5)) are one-time changes that cannot be predicted within the normal process of evolution and coevolution. They might be due to acts of genius (Newton), a political event (Civil War), or a monarch changing religion (Henry VIII). An understanding of the timing of the shocks provides insight into the times when cultural evolution took (or did not take) an unexpected turn relative to what would have been predicted by the model's "normal" dynamics.

Figure 5 plots the temporal paths of five-year centered moving averages of the estimated shocks to  $relig_t$ ,  $sci_t$ , and  $inst_t$ . The vertical axis measures the magnitude of the shocks. The absolute value of the typical moving average displayed in Figure 5 is approximately one-third.

The largest positive shocks to  $relig_t$  (Figure 5(a)) occur in the years 1547–1549, 1565–1567, 1579–1583, and 1650–1659. During these periods, attention to religion evidences positive shocks that are about three times as large as typical shocks during 1530–1700. The first period (1547–1549) is at the beginning of Edward VI's reign, when a more radical form of Protestantism was being introduced. The second period (1565–1567) is in Elizabeth's first decade, when Calvinist ideas were being strongly promoted by those outside the established church.<sup>26</sup> The third period (1579–1583) coincides with a time when Elizabeth was facing internal and external challenges from Catholics, with Puritan ideas finding more favor.<sup>27</sup> The fourth period (1650–1659), the Interregnum when Puritans controlled the country, evidences the largest amount of unusual attention to religion.

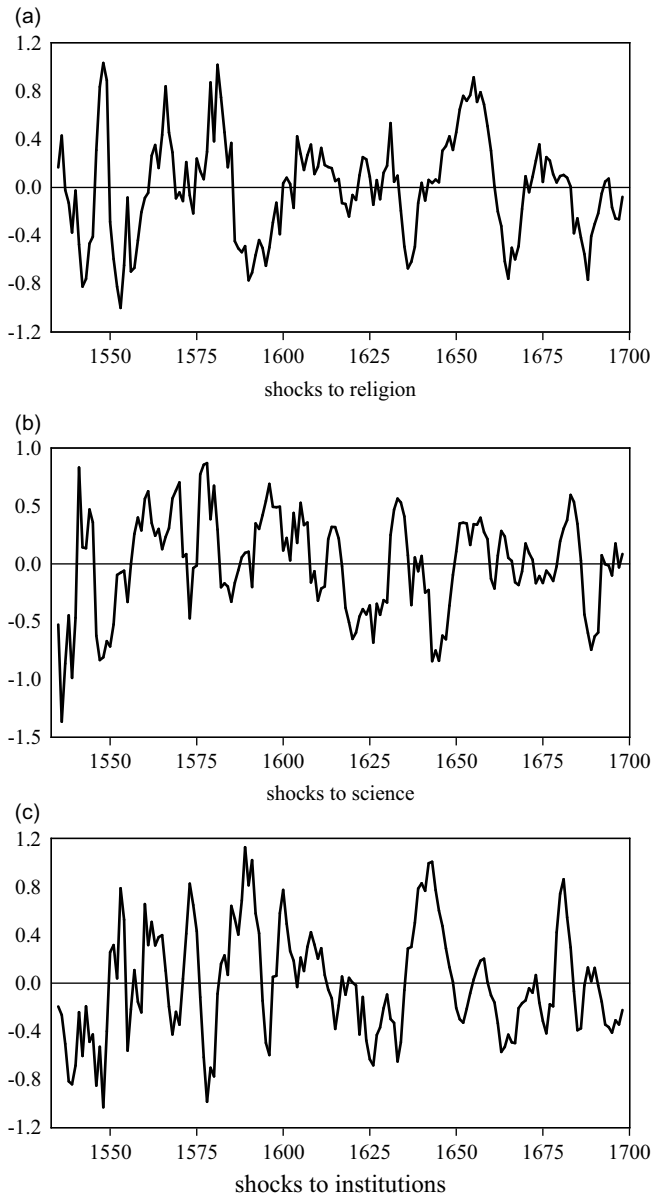
In the last four decades of the seventeenth century, the only unusually large shocks to  $relig_t$  are negative: in 1663–1668 and 1686–1689. During these times, attention to religion experiences negative shocks that are about two-and-a-half times the size of typical shocks occurring between 1530 and 1700. During the first of these periods (1663–1668), the country saw a backlash against the Puritan ideas that had been so prominent during the Interregnum, with pathways to the spread of such ideas largely blocked by the established church (MacCulloch 2004: 544). The latter period (1686–1689) coincides with the reign of an avowedly Catholic King, James II, who for a brief time threatened all flavors of Protestantism, discouraging attention to the corresponding ideas. Thus, echoing the conclusion of the previous paragraph, our religion variable clearly reflects the ebb and flow of the attention to Puritan ideas.

Large positive shocks to  $inst_t$  occur in the 1580s (Figure 5(c)). This period is a particularly fecund one in caselaw, the reports of which are not contained in the EEBO-TCP corpus, but the reverberations would certainly enter that corpus.<sup>28</sup> After that, there are two time periods featuring prominent positive shocks to  $inst_t$ : 1636–1647, the era before and during the Civil War; and 1679–1683, coinciding with the

<sup>26</sup>In 1553, a large group of Protestant clergy fled England upon the accession of Mary I, a Catholic monarch. Many spent their exile in Calvin's Geneva. They returned on the accession of Elizabeth and thereafter struggled to change the tenor of English Protestantism England. The Geneva Bible, in English produced by Calvinists, became available in 1560.

<sup>27</sup>As Marshall (2017: 738) notes, "At this moment of crisis, Elizabeth's Church needed the intellectual firepower of godly Protestantism."

<sup>28</sup>The observation about the fecundity of caselaw in the 1580s follows from our ongoing work in which we draw on a separate caselaw corpus. The relevant results are available on request to the authors.



**Figure 5.** The temporal incidence of structural shocks.

exclusion crisis that foreshadowed the Glorious Revolution. Our estimates are thus more consistent with the idea that debates on institutional development spurred revolutions than the notion that revolutions spurred debates on institutional development.

With regard to the Glorious Revolution (1688–1689), our evidence does not lend support to the North and Weingast (1989) view that it marked a critical juncture in

English institutional history. The years 1688–1700 do not feature the kind of unusually large shocks (positive or negative) to  $inst_t$  that one would expect to observe had the Glorious Revolution fundamentally impacted ideas on institutions. Moreover, with the period leading up to the Glorious Revolution witnessing large negative shocks to  $relig_t$  (1686–1689; Figure 5(a)) and large positive shocks to  $inst_t$  (1679–1683; Figure 5(b)), our evidence indicates that the struggles leading up to the Glorious Revolution were not centered on religion, but rather on institutional, and especially constitutional, matters.<sup>29</sup>

Finally, the era from 1558 to 1610 saw a series of comparatively large positive shocks to  $sci_t$  (Figure 5(b)), an observation that locates the stirring of an English scientific revolution earlier than is conventional (Wootton 2015). In contrast, the negative shocks to  $sci_t$  from 1618 to 1650 can be explained by the existence of deep religious and political cleavages. An emphasis on coping with deep divisions within the nation and the accompanying instability would have naturally led to a decline in attention to science.

### **What drove attention to science? The Royal Society and alternatives**

Did the founding of the Royal Society in 1660 elevate attention to science? And more generally, which shocks were particularly important as drivers of attention to science and when? To shed light on these questions, we investigate the  $sci_t$  series using historical decomposition, a VAR tool that allows us to apportion the overall fluctuation of  $sci_t$  among the effects of different types of shocks (i.e., current and past exogenous events). This tool combines information from the pertinent IRFs (Figure 4(d), (e) and (f)), which show how attention to science on average responded to specific types of shocks, and the timing of the different types of shocks (Figure 5).

Figure 6 presents a historical decomposition of the  $sci_t$  series for the 1650–1675 period, the era immediately preceding and following the 1660 founding of the Royal Society. In the figure, “total” indicates the total deviation of  $sci_t$  from the long-run mean predicted by the model. This deviation is then apportioned among three possible sources: past and current shocks to  $sci_t$  (dark gray line),  $inst_t$  (light gray), and  $relig_t$  (black).

The central implication of Figure 6 is that the founding of the Royal Society leaves no visible trace in our results. Specifically, even though the years 1660–1663 exhibit a somewhat higher than expected attention to science (the total is positive), this is by and large an effect of positive cumulative shocks to  $relig_t$  (the black line). These religious shocks, which on average elevated attention to science (Figure 4(d)), occurred especially during the Interregnum (Figure 5(a)), when Puritan sentiments were particularly strong. In contrast, shocks to  $sci_t$ , which on average boosted subsequent attention to science (Figure 4(e)), were of limited size throughout the 1650s and the early 1660s (Figure 5(b)). As such, cumulative shocks to  $sci_t$  had little effect on attention to science during 1660–1663 (the dark gray line). The same is true of the shocks to  $inst_t$  (the light gray line).

<sup>29</sup>The topics Constitutional Rules and Parliamentary & Court Procedure, both of which are part of the religion theme (see Table 1), have particularly high levels of attention during this time period. See Appendix C.

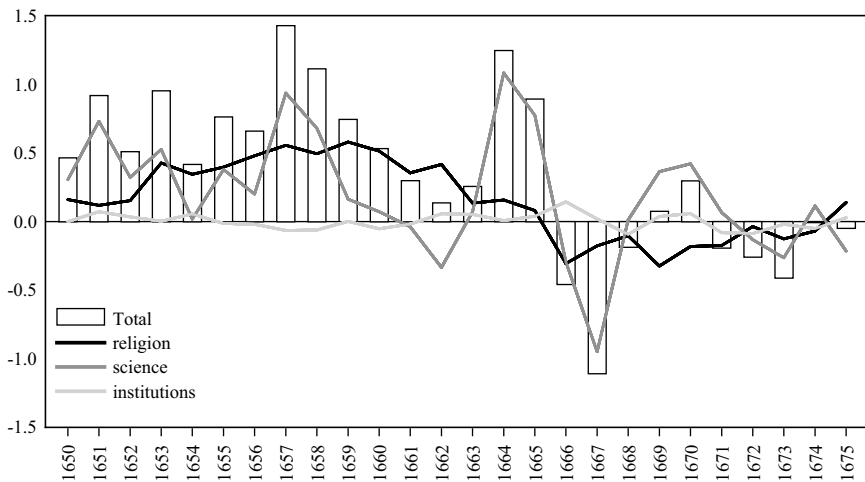


Figure 6. The historical decomposition of the science series, 1650–1675.

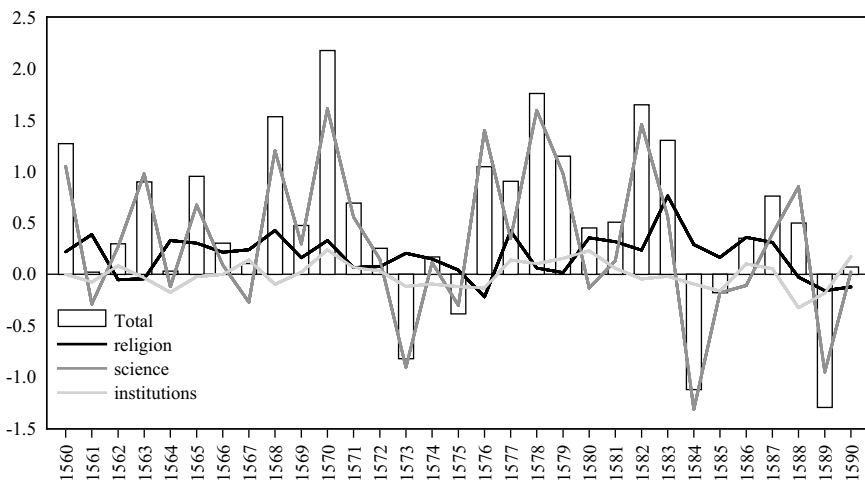


Figure 7. The historical decomposition of the science series, 1560–1590.

Our results are thus not consistent with the hypothesis that the founding of the Royal Society quickly led to an unusual outpouring of science.<sup>30</sup> Our estimates are more consistent with an interpretation that developments in religion during the Interregnum later heightened attention to ideas about science – ideas that were conceivably important for the founding of the Royal Society in 1660.

Figure 7 shows an analogous historical decomposition of the  $sci_t$  series for an earlier period, 1560–1590. The figure reveals that between 1564 and 1587, shocks to

<sup>30</sup>The original name was the College for the Promoting of Physico-Mathematical Experimental Learning. Importantly, our results should not be interpreted as implying that the Royal Society had no effect on the nature of scientific output, for example, on how science was done.



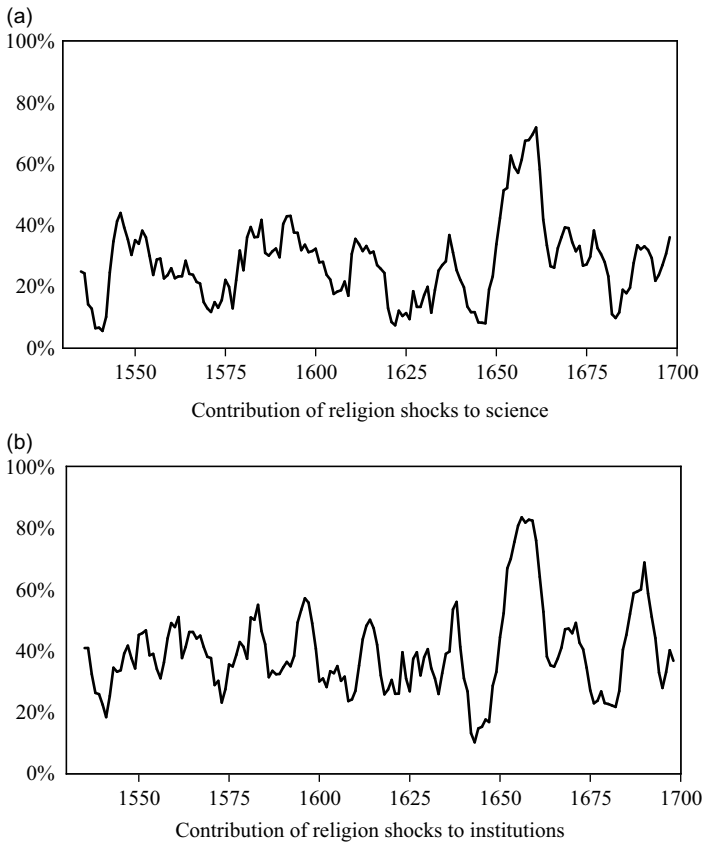


Figure 8. Importance for attention to science and institutions of shocks to ideas about religion.

$relig_t$  boosted attention to science in all but one year (the black line is above zero). As discussed in the previous subsection, during that time the large positive shocks to  $relig_t$ , which on average increased  $sci_t$  (Figure 4(d)), plausibly arose from the rising influence of Puritanism.

Figure 7 therefore provides further direct evidence in support of the Merton thesis: ideas about religion stimulated attention to ideas about science. However, Figure 7 also contextualizes Merton's theory by providing novel insights regarding the timing of the pertinent influence. Merton (1938: 414–8) himself, as well as Mokyr (2016: Ch. 13), focused on the role of religious sentiments in spurring the development of science in the seventeenth century. Our results indicate that these effects were particularly strong already in the last half of the sixteenth century.

### ***Was there secularization of science and institutional thought?***

Finally, we draw on our estimated historical decompositions to assess whether the strength of the influence of attention to religion on science and institutional thought changed over time. If the influence of religion waned over time, this would be evidence of secularization of thought in these two areas.

Figure 8 shows 5 year moving averages of the relative importance of cumulative shocks to religion as drivers of attention to science (part (a)) and institutions (part (b)). To construct the figure, we use the estimates of the historical decomposition for each of the  $sci_t$  and  $inst_t$  series. We then calculate the percentage share of the effects of cumulative shocks to  $relig_t$  in the sum of the absolute values of the contributions of all three types of cumulative shocks. This is a measure of the relative importance of the effects of cumulative shocks to  $relig_t$  on  $sci_t$ .

The key finding implied by Figure 8 is that innovations in religious thought were an important driver of attention to both science and institutions throughout the period 1530–1700. There is no evidence of any diminishing relevance over time of shocks in religion to debates on science and institutions. In particular, our findings are inconsistent with those of Almelhem et al. (2023: 3) who claim that, in England, “as early as 1600, and certainly by the mid-seventeenth century, there was little overlap of scientific and religious topics.”

### Concluding discussion

Our objective has been, first, to present a quantitative macroscope of English print culture during the pivotal sixteenth and seventeenth centuries and, second, to examine the coevolution of ideas on religion, science, and institutions. To this end, we analyzed a major corpus of early-modern publications, combining unsupervised machine learning methods for analysis of text-as-data with time-series econometrics.

Our analysis generated five main substantive findings. First, as the early-modern era unfolded, English religious and political discourse became less antagonistic and, in line with Erikson (2021), witnessed a rise in prominence of topics relevant to the economy and commerce. Within the economics discipline, the existing literature on early-modern England has highlighted institutional developments as the key to subsequent sustained economic progress (Acemoglu and Robinson 2012; Henriques and Palma 2023; North et al. 2009). Our inquiry provides a different perspective, suggesting that the eighteenth-century stability of the “nation of shopkeepers” was foreshadowed by important sixteenth- and seventeenth-century cultural changes.

Second, the epistemology that came to be associated with Bacon was present in theological debates already before Bacon’s epistemological contributions. Our analysis thus casts new light on an influential strand of recent research stressing the role that Francis Bacon played in defining a set of cultural beliefs that were conducive to later economic development (Mokyr 2005, 2016). Our evidence tends to support a view of Bacon as a “synthetic thinker” (Mokyr 2016: 74) or “an influential mouthpiece for a culture that already existed” (Grajzl and Murrell 2019: 112) rather than a fully fledged “intellectual innovator” (Mokyr 2016: 68).

Third, congruent with Merton’s (1938) seminal thesis, innovations in religious ideas stimulated attention to science, especially at times when Puritanism was prominent in religious discourse. We show that the link from religion to science was strong already in the latter half of the sixteenth century, that is, much earlier than stressed by Merton (1938: 414–6) or more recently by Mokyr (2016: Ch. 13), who both emphasized developments after the mid-seventeenth century. Moreover, we demonstrate that the interplay between religion and science in early-modern English print culture was bidirectional.

Fourth, neither science nor institutional thought evidence secularization as the seventeenth century closes. Our evidence is thus not consistent with recent analysis by Almelhem et al. (2023: 35) who use texts available with the Hathitrust Digital Library (HDI) to conclude that, in Britain, “the ‘secularization’ of science was entrenched from the beginning of the Enlightenment.” Importantly, our analysis utilizes the EEBO-TCP corpus, which focuses on the pre-1700 era. This is not true of the HDI corpus available to Almelhem et al., where the coverage of the pre-1700 period is sparse. Moreover, unlike Almelhem et al.’s conclusion, our finding is based on an explicit VAR model that naturally lends itself to the study of coevolutionary patterns (see, e.g., Grajzl and Murrell 2022).

Fifth, the Civil War and the Glorious Revolution did not spur debates on institutions nor did the founding of the Royal Society markedly elevate attention to science. Our empirical analysis thus provides a novel interpretation of major historical events during a crucial epoch of England’s cultural history. We show that a focus on institutional debates preceded the Civil War. Then, the emphasis on religion during the ensuing Interregnum led to an enhanced focus on science at the time of the formation of the Royal Society. In addition, attention to institutions did not noticeably change around the time of the Glorious Revolution. Thus, rather than supporting the view that 1688–1689 was a pivotal moment in English institutional history, our evidence from print culture echoes recent research emphasizing the gradual character of England’s institutional development (Henriques and Palma 2023; Hodgson 2017; Murrell 2017; Ogilvie and Carus 2014).

A core product of this research has been the creation of a machine learning digest of sixteenth- and seventeenth-century English culture, a self-standing dataset available to researchers for further exploration. Our construction and provision of data is analogous to many other data-producing exercises in social-science and cliometric research that combine large amounts of micro data to build a macro dataset that can be used as an input into further research (see, e.g., Broadberry et al. 2015). At the same time, the availability of our machine-learning-generated data allows others to easily reassess our findings by applying, for example, different assumptions about the construction of cultural aggregates, different identification assumptions, and, indeed, wholly different econometric approaches.

**Supplementary material.** To view supplementary material for this article, please visit <https://doi.org/10.1017/ssh.2024.17>

**Acknowledgments.** We thank Katherine Calloway for help with the history of ideas, Boragan Aruoba for econometric advice, Franz Klein for facilitating use of the BSWIFT cluster, Paul Schaffner for help in downloading and understanding the TCP corpus, two anonymous referees for crucial comments, and the Editors for help in polishing the paper.

## References

- Acemoglu, Daron, and James A. Robinson** (2012) *Why Nations Fail: The Origins of Power, Prosperity and Poverty*. New York, NY: Crown Business.
- Airoldi, Edoardo M., and Jonathan M. Bischof** (2016) “Improving and evaluating topic models and other models of text.” *Journal of the American Statistical Association* **111** (516): 1381–403.
- Almelhem, Ali, Murat Iyigun, Austin Kennedy, and Jared Rubin** (2023) “Enlightenment ideals and belief in science in the run-up to the Industrial Revolution: A textual analysis.” Unpublished manuscript.

- Barnard, John** (2002) "Introduction," in John Barnard and Donald F. McKenzie (eds.) with Maureen Bell, *The Cambridge History of the Book in Britain, Volume IV, 1557–1695*. Cambridge, UK: Cambridge University Press: 1–26.
- Barnard, John, and Donald F. McKenzie**, eds. with Maureen Bell (2002) *The Cambridge History of the Book in Britain, Volume IV, 1557–1695*. Cambridge, UK: Cambridge University Press.
- Barnes, Michael H.** (2000) *Stages of Thought: The Co-Evolution of Religious Thought and Science*. New York, NY: Oxford University Press.
- Becker, Sascha O., and Ludger Woessmann** (2009) "Was Weber wrong? A human capital theory of Protestant economic history." *Quarterly Journal of Economics* **124** (2): 531–96.
- Ben Zeev, Nadav, Joel Mokyr, and Karine van der Beek** (2017) "Flexible supply of apprenticeship in the British Industrial Revolution." *Journal of Economic History* **77** (1): 208–50.
- Blair, Ann, and Kaspar von Greyerz** (2020) "Introduction," in Ann Blair and Kaspar von Greyerz (eds.) *Physico-theology: Religion and Science in Europe, 1650–1750*. Baltimore, MD: Johns Hopkins University Press: 1–21.
- Broadberry, Stephen, Bruce M.S. Campbell, Alexander Klein, Mark Overton, and Bas van Leeuwen** (2015) *British Economic Growth, 1270–1870*. Cambridge, UK: Cambridge University Press.
- Brooke, John Hedley** (1990) "Science and religion," in R.C. Olby, G.N. Cantor, J.R.R. Christie and M.J.S. Hodge (eds.) *Companion to the History of Modern Science*. New York, NY: Routledge: 763–82.
- Brzezinski, Adam, Yao Chen, Nuno Palma, and Felix Ward** (2022) "The vagaries of the sea: Evidence on the real effects of money from maritime disasters in the Spanish empire." *Review of Economics and Statistics*, forthcoming. [https://doi.org/10.1162/rest\\_a\\_01223](https://doi.org/10.1162/rest_a_01223)
- Burkov, Andriy** (2019) *The Hundred-Page Machine Learning Book*. Quebec City: Andriy Burkov.
- Caldara, Dario, and Matteo Iacoviello** (2022) "Measuring geopolitical risk." *American Economic Review* **112** (4): 1194–225.
- Calloway, Katherine** (2015) *Natural Theology in the Scientific Revolution: God's Scientists*. New York, NY: Routledge.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans** (1999) "Monetary policy shocks: What have we learned and to what end?" in John B. Taylor and Michael Woodford (eds.) *Handbook of Macroeconomics*. Vol. 1. Amsterdam: Elsevier: 65–148.
- Cohen, I. Bernard** (1990) "Introduction: The impact of the Merton thesis," in I. Bernard Cohen (ed.) *Puritanism and the Rise of Modern Science, The Merton Thesis*. New Brunswick, NJ: Rutgers University Press: i–xiii.
- Collinson, Patrick, Arnold Hunt, and Alexandra Walsham** (2002) "Religious publishing in England 1557–1640," in John Barnard and Donald F. McKenzie (eds.) with Maureen Bell, *The Cambridge History of the Book in Britain, Volume IV, 1557–1695*. Cambridge, UK: Cambridge University Press: 29–66.
- Crafts, Nicholas, and Terence C. Mills** (2017) "Six centuries of British economic growth: A time-series perspective." *European Review of Economic History* **21** (2): 141–58.
- Dittmar, Jeremiah, and Skipper Seabold** (2021) "New media and competition: Printing and Europe's transformation after Gutenberg." Unpublished manuscript.
- Erikson, Emily** (2021) *Trade and Nation: How Companies and Politics Reshaped Economic Thought*. New York, NY: Columbia University Press.
- Gennaro, Gloria, and Elliott Ash** (2022) "Emotion and reason in political language." *Economic Journal* **132** (643): 1037–59.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy** (2019) "Text as data." *Journal of Economic Literature* **57** (3): 535–74.
- Grajzl, Peter, and Peter Murrell** (2019) "Toward Understanding 17<sup>th</sup> century English culture: A structural topic model of Francis Bacon's ideas." *Journal of Comparative Economics* **47** (1): 111–35.
- (2021) "A machine-learning history of English caselaw and legal ideas prior to the Industrial Revolution I: Generating and interpreting the estimates." *Journal of Institutional Economics* **17** (1): 1–19.
- (2022) "A macrohistory of legal evolution and coevolution: Property, procedure, and contract in early-modern English caselaw." *International Review of Law and Economics* **73**: 106113.
- (2023) "Of families and inheritance: Law and development in England before the Industrial Revolution." *Cliometrica* **17**: 387–432.
- Green, Ian, and Kate Peters** (2002) "Religious publishing in England 1640–1695," in John Barnard and Donald F. McKenzie (eds.) with Maureen Bell, *The Cambridge History of the Book in Britain, Volume IV, 1557–1695*. Cambridge, UK: Cambridge University Press: 67–94.

- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart** (2021) "Machine learning for social science: An agnostic approach." *Annual Review of Political Science* **24**: 395–419.
- (2022) *Text as Data. A New Framework for Machine Learning and the Social Sciences*. Princeton, NJ: Princeton University Press.
- Harkness, Deborah E.** (2007) *The Jewel House: Elizabethan London and the Scientific Revolution*. New Haven, CT: Yale University Press.
- Harrison, Peter** (2008) "Religion, the Royal Society, and the rise of science." *Theology and Science* **6** (3): 255–71.
- Henriques, António, and Nuno Palma** (2023) "Comparative European institutions and the little divergence, 1385–1800." *Journal of Economic Growth* **28**: 259–94.
- Hill, Christopher** (1965) *Intellectual Origins of the English Revolution Revisited*. Oxford, UK: Clarendon Press.
- Hodgson, Geoffrey M.** (2017) "1688 and all that: Property rights, the glorious revolution and the rise of British capitalism." *Journal of Institutional Economics* **13** (1): 79–107.
- Holifield, E. Brooks** (2003) *Theology in America: Christian Thought from the Age of the Puritans to the Civil War*. New Haven, CT: Yale University Press.
- Jalil, Andrew J.** (2015) "A new history of banking panics in the United States, 1825–1929: Construction and implications." *American Economic Journal: Macroeconomics* **7** (3): 295–330.
- Johns, Adrian** (2002) "Science and the book," in John Barnard and Donald F. McKenzie (eds.) with Maureen Bell, *The Cambridge History of the Book in Britain, Volume IV, 1557–1695*. Cambridge, UK: Cambridge University Press: 274–303.
- Jordà, Óscar, Sanjay R. Singh, and Alan M. Taylor** (2022) "Longer-run economic consequences of pandemics." *Review of Economics and Statistics* **104** (1): 166–75.
- Kenny, Seán, Jason Lennard, John D. Turner** (2021) "The macroeconomic effects of banking crises: Evidence from the United Kingdom, 1750–1938." *Explorations in Economic History* **79** (C): 101357. <https://doi.org/10.1016/j.eeh.2020.101357>
- Kilian, Lutz** (1998) "Small-sample confidence intervals for impulse response functions." *Review of Economics and Statistics* **80** (2): 218–30.
- Kilian, Lutz, and Helmut Lutkepohl** (2017) *Structural Vector Autoregressive Analysis*. Cambridge, UK: Cambridge University Press.
- Kuran, Timur** (2010) *The Long Divergence: How Islamic Law Held Back the Middle East*. Princeton, NJ: Princeton University Press.
- Lee, Moontae, and David Mimno** (2014) "Low-dimensional embeddings for interpretable anchor-based topic inference," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL: 1319–28.
- Livermore, Michael A., and Daniel N. Rockmore** (2019) *Law as Data: Computation, Text, and the Future of Legal Analysis*. Santa Fe, NM: Santa Fe Institute Press.
- Marshall, Peter** (2017). *Heretics and Believers: A History of the English Reformation*. New Haven, CT: Yale University Press.
- Martin, Julian** (1992) *Francis Bacon, the State, and the Reform of Natural Philosophy*. Cambridge, UK: Cambridge University Press.
- MacCulloch, Diarmaid** (2004) *The Reformation*. New York, NY: Viking Penguin Group.
- Merton, Robert K.** (1938) "Science, technology and society in seventeenth century England." *Osiris* **4**: 360–632.
- McCloskey, Deirdre N.** (2016) *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World*. Chicago, IL: The University of Chicago Press.
- Mokyr, Joel** (2005) "The intellectual origins of modern economic growth." *Journal of Economic History* **65** (2): 285–351.
- (2009) *The Enlightened Economy: An Economic History of Britain 1700–1850*. New Haven, CT: Yale University Press.
- (2016) *A Culture of Growth: The Origins of the Modern Economy*. Princeton, NJ: Princeton University Press.
- Møller, Niels Framroze, and Paul Sharp** 2014. "Malthus in cointegration space: Evidence of a post-Malthusian pre-industrial England." *Journal of Economic Growth* **19** (1): 105–40.
- Murrell, Peter** (2017) "Design and evolution in institutional development: The insignificance of the English Bill of Rights." *Journal of Comparative Economics* **45** (1): 36–55.

- Nakamura, Emi, and Jón Steinsson** (2018) "Identification in macroeconomics." *Journal of Economic Perspectives* **32** (3): 59–86.
- North, Douglass C., and Barry R. Weingast** (1989) "Constitutions and commitment: The evolution of institutions governing public choice in seventeenth-century England." *Journal of Economic History* **49** (4): 803–32.
- North, Douglass C., John J. Wallis, and Barry R. Weingast** (2009) *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. New York, NY: Cambridge University Press.
- Ogilvie, Sheilagh, and A.W. Carus** (2014) "Institutions and economic growth in historical perspective," in Philippe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth*. Vol. 2. Amsterdam: Elsevier: 403–513.
- Palma, Nuno** (2016) "Sailing away from Malthus: Intercontinental trade and European economic growth, 1500–1800." *Cliometrica* **10**: 129–49.
- (2022) "The real effects of monetary expansions: Evidence from a large-scale historical experiment." *Review of Economic Studies* **89** (3): 1593–627.
- Plagborg-Møller, Mikkel, and Christian K. Wolf** (2021) "Local projections and VARs estimate the same impulse responses." *Econometrica* **89** (2): 955–80.
- Porter, Martin F.** (1980) "An algorithm for suffix stripping." *Program* **14** (3): 130–7.
- Porter, Roy** (2000) *The Creation of the Modern World: The Untold Story of the British Enlightenment*. New York, NY: W.W. Norton & Company.
- Price, Bronwen** (2017) "Finding the genuine light of nature: Religion and science," in Andrew Hiscock and Helen Wilcox (eds.) *The Oxford Handbook of Early Modern English Literature and Religion*, Oxford, UK: Oxford University Press: 579–97.
- Quinn, Stephen, and William Roberds** (2019) "A policy framework for the Bank of Amsterdam, 1736–1791." *Journal of Economic History* **79** (3): 736–72.
- Rabb, Theodore K.** (1965) "Religion and the rise of modern science." *Past & Present* **31**: 111–26.
- Ramey, Valerie** (2016) "Macroeconomic shocks and their propagation," in John B. Taylor and Harald Uhlig (eds.) *Handbook of Macroeconomics*. Vol. 2. Amsterdam: Elsevier: 71–162.
- Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoidi** (2016a) "A model of text for experimentation in the social sciences." *Journal of the American Statistical Association* **111** (515): 988–1003.
- Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley** (2016b) "Navigating the local modes of big data: The case of topic models," in R. Michael Alvarez (ed.) *Computational Social Science, Discovery and Prediction*, New York, NY: Cambridge University Press: 51–97.
- (2019) "stm: R package for structural topic models." *Journal of Statistical Software* **91** (2): 1–40.
- Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand** (2014) "Structural topic models for open-ended survey responses." *American Journal of Political Science* **58** (4): 1064–82.
- Rubin, Jared** (2017) *Rulers, Religion, and Riches: Why the West Got Rich and the Middle East Did Not*. New York, NY: Cambridge University Press.
- Shapin, Steven** (1988) "Understanding the Merton thesis." *Isis* **79** (4): 594–605.
- (1996) *The Scientific Revolution*. Chicago, IL: University of Chicago Press.
- Shapiro, Barbara J.** (2000). *A Culture of Fact: England, 1550–1720*. Ithaca, NY: Cornell University Press.
- Sims, Christopher A.** (1980) "Macroeconomics and reality." *Econometrica* **48** (1): 1–48.
- Stock, James H., and Mark W. Watson** (2001) "Vector autoregressions." *Journal of Economic Perspectives* **15** (4): 101–15.
- Taddy, Matthew A.** (2012) "On estimation and selection for topic models," in *Proceedings of the 15th International Conference on Artificial Intelligence and Statistics*. La Palma, Canary Islands: JMLR WCP: 1184–93.
- The Text Creation Partnership [TCP]** (2022), [textcreationpartnership.org](https://textcreationpartnership.org) (accessed February 27, 2024).
- Wallach, Hanna M., Iain Murray, Ruslan Salakhutdinov, and David Mimno** (2009) "Evaluation methods for topic models," in *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal, Canada: ACM: 1105–12.
- Wang, Hongbin, Jianxiong Wang, Yafei Zhang, Meng Wang, and Cunli Mao** (2019) "Optimization of topic recognition model for news texts based on LDA." *Journal of Digital Information Management* **17** (5): 257–69.

- Webster, Charles** (1974) "Introduction," in Charles Webster (ed.) *The Intellectual Revolution of the Seventeenth Century*. Boston, MA: Routledge & Kegan Paul: 1–22.
- Weston, Sara J., Ian Shryock, Ryan Light, and Phillip A. Fisher** (2023) "Selecting the number and labels of topics in topic modeling: A tutorial." *Advances in Methods and Practices in Psychological Science* 6 (2): 1–13.
- Wootton, David** (2015) *The Invention of Science: A New History of the Scientific Revolution*. New York, NY: Harper. Kindle Edition.

**Peter Grajzl** is John F. Hendon Professor of Economics at Washington and Lee University. His research interests lie in comparative institutional economics, law and economics, and political economy. He has published on a wide range of topics pertaining to the emergence, the functioning, and the impact of different legal, political, and economic institutions, as well as cultural beliefs, in multiple parts of the world. His history-oriented research combines machine-learning and econometrics to examine legal, cultural, and economic development in pre-19th-century England ([grajzl@wlu.edu](mailto:grajzl@wlu.edu)).

**Peter Murrell** is professor of economics at the University of Maryland. Earlier research on the transition economies of Eastern Europe, the former USSR, and China led to deliberation on institutional development. This led naturally to research on the earliest successful development, early-modern England, which examines when and how legal, cultural, and economic development occurred, and how they interacted. Machine-learning methods are integral. Other research involves law and culture, on attitudes to law in Russia and on methods for contract enforcement ([pmurrell@umd.edu](mailto:pmurrell@umd.edu)).