

# Enlightenment Ideals and Belief in Progress in the Run-up to the Industrial Revolution

## A Textual Analysis

Ali Almelhem  
World Bank

Austin Kennedy  
U. Colorado

Murat Iyigun  
U. Colorado

Jared Rubin  
Chapman U.

Foundations of Long-run Prosperity  
Hoover Institution  
March 2024

# Language and Long-Run Cultural Change

Cultural norms and **cultural change** are important determinants of why some countries are rich and others are not

- ▶ Deep roots: Alesina, Giuliano, and Nunn (2013); Giuliano and Nunn (2021); Schulz et al. (2019); Enke (2019)
- ▶ The “Great Divergence”: Greif and Tabellini (2017); Bisin et al. (2023)
- ▶ Britain: Mokyr (2016); McCloskey (2006, 2010, 2016); Koyama and Rubin (2022)

# Language and Long-Run Cultural Change

Cultural norms and **cultural change** are important determinants of why some countries are rich and others are not

- ▶ Deep roots: Alesina, Giuliano, and Nunn (2013); Giuliano and Nunn (2021); Schulz et al. (2019); Enke (2019)
- ▶ The “Great Divergence”: Greif and Tabellini (2017); Bisin et al. (2023)
- ▶ Britain: Mokyr (2016); McCloskey (2006, 2010, 2016); Koyama and Rubin (2022)

The issue: how do we obtain finely grained data regarding how **culture changes over time**?

# Language and Long-Run Cultural Change

Cultural norms and **cultural change** are important determinants of why some countries are rich and others are not

- ▶ Deep roots: Alesina, Giuliano, and Nunn (2013); Giuliano and Nunn (2021); Schulz et al. (2019); Enke (2019)
- ▶ The “Great Divergence”: Greif and Tabellini (2017); Bisin et al. (2023)
- ▶ Britain: Mokyr (2016); McCloskey (2006, 2010, 2016); Koyama and Rubin (2022)

The issue: how do we obtain finely grained data regarding how **culture changes over time**?

There is much cultural information **embedded in language**

- ▶ Erikson (2021); Chen (2013); Michalopoulos and Xue (2021); Galor, Özak, and Sarid (2020); Giorcelli, Lacetera, and Marinoni (2022)
- ▶ Reviews: Grimmer and Stewart (2013); Gentzkow, Kelly, and Taddy (2019)

# What Can Language Tell Us?

Language can tell us much about gender norms, religious norms, which occupations societies value, attitudes towards risk, etc.

# What Can Language Tell Us?

Language can tell us much about gender norms, religious norms, which occupations societies value, attitudes towards risk, etc.

This makes language an incredibly useful tool for [cross-society comparisons](#)

# What Can Language Tell Us?

Language can tell us much about gender norms, religious norms, which occupations societies value, attitudes towards risk, etc.

This makes language an incredibly useful tool for [cross-society comparisons](#)

Yet, analyzing a cross-section of language tells us little about [when](#) language changed

# What Can Language Tell Us?

Language can tell us much about gender norms, religious norms, which occupations societies value, attitudes towards risk, etc.

This makes language an incredibly useful tool for [cross-society comparisons](#)

Yet, analyzing a cross-section of language tells us little about [when](#) language changed

Timing of change is especially important for the study of long-run growth, which has been fraught with “efflorescences” and reversals of fortune



# Changes in the English Language as Seen in Books

Data on cultural change, as embedded in language, can be extracted from the corpus of **volumes** printed since the spread of the printing press in the 15th century

# Changes in the English Language as Seen in Books

Data on cultural change, as embedded in language, can be extracted from the corpus of **volumes** printed since the spread of the printing press in the 15th century

In this paper, we seek to understand how the English language changed (on certain dimensions) over the period 1500–1900 by applying text analysis techniques to the entire corpus that has been digitized (173,031 unique volumes)

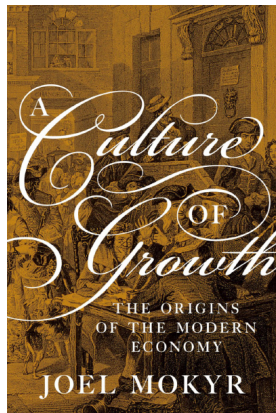
# Changes in the English Language as Seen in Books

Data on cultural change, as embedded in language, can be extracted from the corpus of **volumes** printed since the spread of the printing press in the 15th century

In this paper, we seek to understand how the English language changed (on certain dimensions) over the period 1500–1900 by applying text analysis techniques to the entire corpus that has been digitized (173,031 unique volumes)

There are thousands of ways language changed in this time period. Are any of those **economically relevant**? Specifically, can we detect any changes in language that may have been relevant for **Britain's economic takeoff**?

# The Industrial Enlightenment

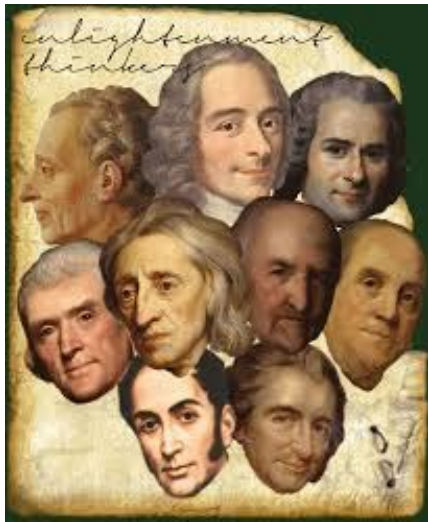


Mokyr (2016) makes the case that one of the keys to the “Industrial Enlightenment” was a progress-oriented view of science

- ▶ Especially Francis Bacon and Isaac Newton
- ▶ Pan-European: Descartes, Spinoza, Leibniz, Kepler, Galileo
- ▶ Upper-tail human capital as an important precursor of industrialization (Squicciarini and Voigtländer 2015)

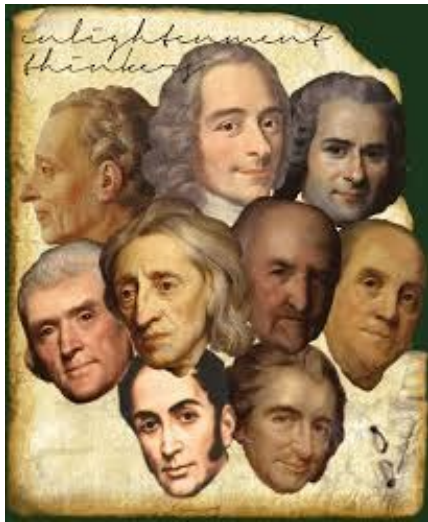
The idea: science and our understanding of the natural world could be used to improve the lot of humankind

## A Culture of Growth



The progress-oriented view of science was pan-European: a “culture of growth” was supported by the social norms of elite intellectuals that fostered the free flow, dissemination, and discussion of new ideas across Europe

## A Culture of Growth



The progress-oriented view of science was pan-European: a “culture of growth” was supported by the social norms of elite intellectuals that fostered the free flow, dissemination, and discussion of new ideas across Europe

Cultural values **in combination with Britain’s abundance of skilled craftsmen and artisans** (Mokyr 2009; Kelly, Mokyr, and Ó Gráda 2023) made its industrialization possible

## Evidence of a Culture of Growth?

Mokyr (2016) provides an overwhelming amount of qualitative evidence to support these claims, citing numerous works and quotations. The evidence is abundant and convincing. Yet, there are two margins for which the evidence could be stronger

## Evidence of a Culture of Growth?

Mokyr (2016) provides an overwhelming amount of qualitative evidence to support these claims, citing numerous works and quotations. The evidence is abundant and convincing. Yet, there are two margins for which the evidence could be stronger

1. Mokyr's evidence regards the language of the “human capital elite”. But the “Industrial Enlightenment” was advanced by artisans and those with ties to industry. How did the language aimed at—and sometimes produced by—this population change over time?



## Evidence of a Culture of Growth?

Mokyr (2016) provides an overwhelming amount of qualitative evidence to support these claims, citing numerous works and quotations. The evidence is abundant and convincing. Yet, there are two margins for which the evidence could be stronger

1. Mokyr's evidence regards the language of the “human capital elite”. But the “Industrial Enlightenment” was advanced by artisans and those with ties to industry. How did the language aimed at—and sometimes produced by—this population change over time?
2. Qualitative evidence by construction cannot account for the hundreds of thousands of works produced in this period

## Evidence of a Culture of Growth?

Mokyr (2016) provides an overwhelming amount of qualitative evidence to support these claims, citing numerous works and quotations. The evidence is abundant and convincing. Yet, there are two margins for which the evidence could be stronger

1. Mokyr's evidence regards the language of the “human capital elite”. But the “Industrial Enlightenment” was advanced by artisans and those with ties to industry. How did the language aimed at—and sometimes produced by—this population change over time?
2. Qualitative evidence by construction cannot account for the hundreds of thousands of works produced in this period

This paper addresses these issues by employing textual analysis methods to the universe of all digitized printed volumes published in Britain and written in English between 1500 and 1900

# What We Do

## Part I

We analyze textual data of all volumes published in Britain and written in English between 1500 and 1900, gathered from the Hathitrust Digital Library

- ▶ Comprised of digital scans and optical character recognition (OCR) output
- ▶ Accounting for duplicates and volumes that cannot be read via OCR, this yields 173,031 unique volumes

# What We Do

## Part I

We analyze textual data of all volumes published in Britain and written in English between 1500 and 1900, gathered from the Hathitrust Digital Library

- ▶ Comprised of digital scans and optical character recognition (OCR) output
- ▶ Accounting for duplicates and volumes that cannot be read via OCR, this yields 173,031 unique volumes

Topic modeling yields 60 topics based on words that frequently co-exist with each other and are unique from other topics

# What We Do

## Part I

We analyze textual data of all volumes published in Britain and written in English between 1500 and 1900, gathered from the Hathitrust Digital Library

- ▶ Comprised of digital scans and optical character recognition (OCR) output
- ▶ Accounting for duplicates and volumes that cannot be read via OCR, this yields 173,031 unique volumes

Topic modeling yields 60 topics based on words that frequently co-exist with each other and are unique from other topics

We develop an algorithm to determine the top three sets of unique topics, which clearly relate to three categories: science, religion, and political economy

# What We Do

## Part II

Each volume in the data set is given a weight for science, religion, and political economy, depending on their content as construed in the topic model

# What We Do

## Part II

Each volume in the data set is given a weight for **science**, **religion**, and **political economy**, depending on their content as construed in the topic model

Each volume also receives an **industrial score**, based on the presence of words associated with industry from the indexes of industrial manuals

# What We Do

## Part II

Each volume in the data set is given a weight for **science**, **religion**, and **political economy**, depending on their content as construed in the topic model

Each volume also receives an **industrial score**, based on the presence of words associated with industry from the indexes of industrial manuals

Each volume receives an **progress sentiment score**, derived from dictionaries and thesauruses, based on the presence of progress-oriented words contained in the volume



## Preview of the Findings

1. At some point in the 18th century, the language of **science** became secularized: there was little overlap of **science** and **religion** topics within volumes
  - ▶ There was a shared language for **science** and **political economy** volumes
  - ▶ There was a shared (though different) language for **religion** and **political economy** volumes

## Preview of the Findings

1. At some point in the 18th century, the language of **science** became secularized: there was little overlap of **science** and **religion** topics within volumes
  - ▶ There was a shared language for **science** and **political economy** volumes
  - ▶ There was a shared (though different) language for **religion** and **political economy** volumes
2. The language of **science** started to become more progress-oriented in the mid-18th century and this persisted throughout the period ...

## Preview of the Findings

1. At some point in the 18th century, the language of **science** became secularized: there was little overlap of **science** and **religion** topics within volumes
  - ▶ There was a shared language for **science** and **political economy** volumes
  - ▶ There was a shared (though different) language for **religion** and **political economy** volumes
2. The language of **science** started to become more progress-oriented in the mid-18th century and this persisted throughout the period ... however, works of “pure” **science** were largely neutral with respect to progress-oriented language. The most progress-oriented works were at the nexus of **science** and **political economy**

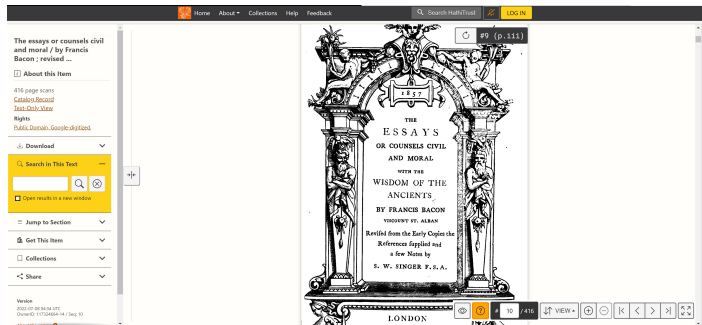
## Preview of the Findings

1. At some point in the 18th century, the language of **science** became secularized: there was little overlap of **science** and **religion** topics within volumes
  - ▶ There was a shared language for **science** and **political economy** volumes
  - ▶ There was a shared (though different) language for **religion** and **political economy** volumes
2. The language of **science** started to become more progress-oriented in the mid-18th century and this persisted throughout the period ... however, works of “pure” **science** were largely neutral with respect to progress-oriented language. The most progress-oriented works were at the nexus of **science** and **political economy**
3. Beginning in the mid-18th century, works with higher industrial scores were more progress-oriented; this result is strongest for works related to industrialization that were at the nexus of **science** and **political economy**

## Data from the Hathitrust Digital Library (HDL)

Data set covers 173,031 unique  
works published in Britain and  
written in English in the HDL  
over the period 1500–1900

Each volume modeled as a “bag  
of words”



# Data Extraction

A Latent Dirichlet Allocation model reduces the corpus to a  $D \times T$  matrix, with  $D$  volumes and  $T = 60$  topics

- ▶  $T = 60$  chosen through model selection process that minimizes average perplexity (i.e., how probable some new unseen data is)

## First 4 Topics

---

---

1 - paint pictur artist music engrav painter colour

2 - town road church build built river stone

3 - franc pari french loui madam duke count

4 - church christian christ bishop holi paul doctrin

# Categorization

The goal is to find categories of topics that have a high relative importance in the corpus and are distinct from each other

For each volume, each topic has a weight representing its occurrence in the volume (weights sum to one per volume)

Use these weights to determine how often two topics co-exist in the same volume, yielding  $\frac{60!}{2!(60-2)!} = 1770$  topic-pair weights

# Categorization

Use topic-pair weights to identify the most frequently occurring and distinct topic-pairs for all volumes

We define a **category** as three topics; the categories we use for categorization have the highest topic-pair weights among all three topics and are distinct from other categories

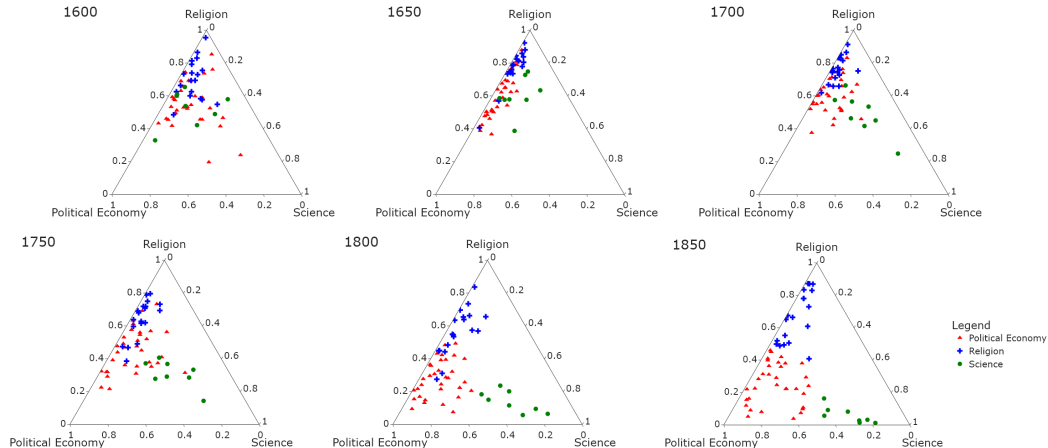
This process yields three categories which can clearly be categorized as **science**, **religion**, and **political economy**



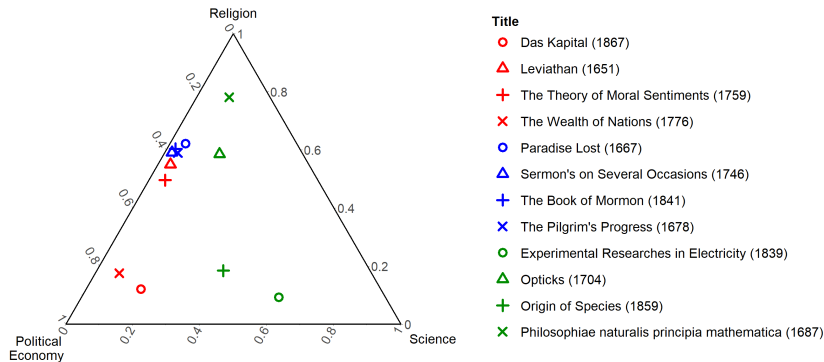
# Three Categories

Category	Topics and associated root words
Political Economy	33 - law lord show public evid opinion fact
	34 - govern nation polit parliament constitut war parti
	47 - trade amount labour money price cent increas
Religion	4 - church christian christ bishop holi paul doctrin
	12 - god christ lord thi faith holi sin
	52 - hath fame religion men shew virtu likewis
Science	7 - fig water iron engin pressur steam electr
	8 - acid solut heat carbon water sulphur iron
	41 - line angl equal equat sin sun plane

# Topics by Category, 1600–1850



# Categorization of Famous Books



“When I wrote my treatise about our system, I had an eye upon such principles as might work with considering men, for the belief of a deity, and nothing can rejoice me more than to find it useful for that purpose.”

–Isaac Newton, discussing *Principia* in private correspondence

## Finding #1: Secularization of the Language of Science

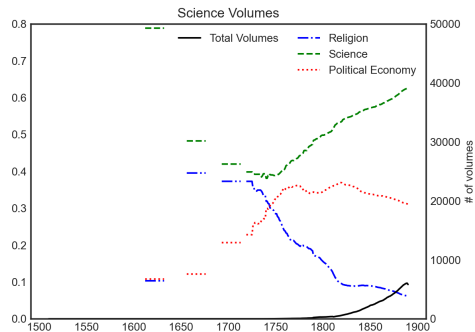
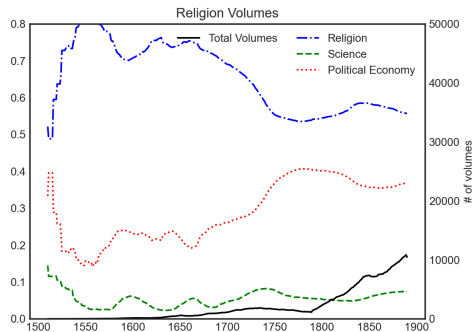
Beginning in the latter half of the 18th century, the language of **science** and **religion** became distinct from each other

- ▶ The frequency of topics that overlap with both **science** and **religion** start to thin out between 1750 and 1800

By contrast, there is a visible and steady shift in language that combines **religion** topics with **political economy** as well as **science** and **political economy**

# Relationship between Categories over time, within volumes

## Science and Religion volumes



## Sentiment Analysis

If changes in culture—as embedded in language—mattered for Britain's ultimate rise, volumes of **science** should have become more progress-oriented in the build-up to Industrialization

# Sentiment Analysis

If changes in culture—as embedded in language—mattered for Britain's ultimate rise, volumes of **science** should have become more progress-oriented in the build-up to Industrialization

We test this hypothesis with a sentiment analysis of all volumes, focusing on how sentiment changed at different times in different places in the simplex

## “Progress-Oriented” Sentiment

We employ dictionary techniques from the Natural Language Processing literature, which rely on lists of words (dictionaries) that comprises of synonyms for each category

- ▶ List of synonyms for each of our four sentiment categories from [www.thesaurus.com](http://www.thesaurus.com)
- ▶ Only include words in use prior to 1643 (year of Newton’s birth) according to OED
  - ▶ Results robust to include more recently used synonyms
- ▶ Manually reviewed each list to tag any words that may have double meanings or are related to religious or scientific language (e.g., “positive” could be a math or chemistry term)
- ▶ **Dictionary:** progress(ion), advance, amelioration, betterment, improve(ment), rise, stride
- ▶ Also created a new “dictionary” using a 17th century dictionary; results are very similar

$$\text{Sentiment}_i = \frac{\sum_{\ell \in L} w_{i,\ell}}{W_i}. \quad (1)$$

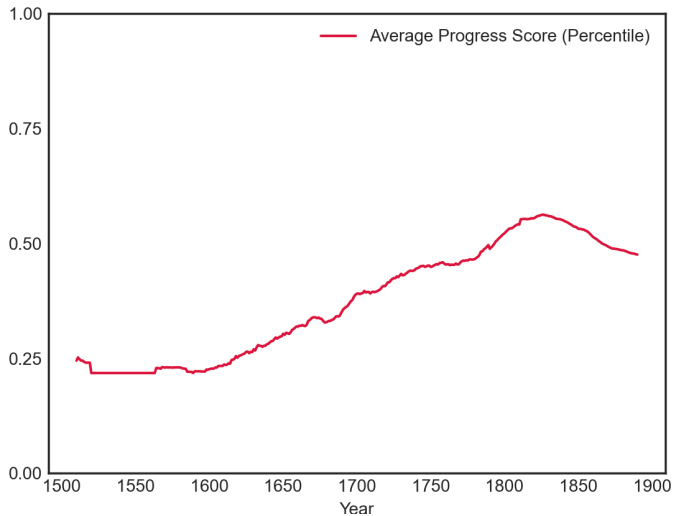
$w_{i,\ell}$ : count of word  $\ell$  in dictionary list  $L$  in volume  $i$

$W_i$ : total number of words in volume  $i$

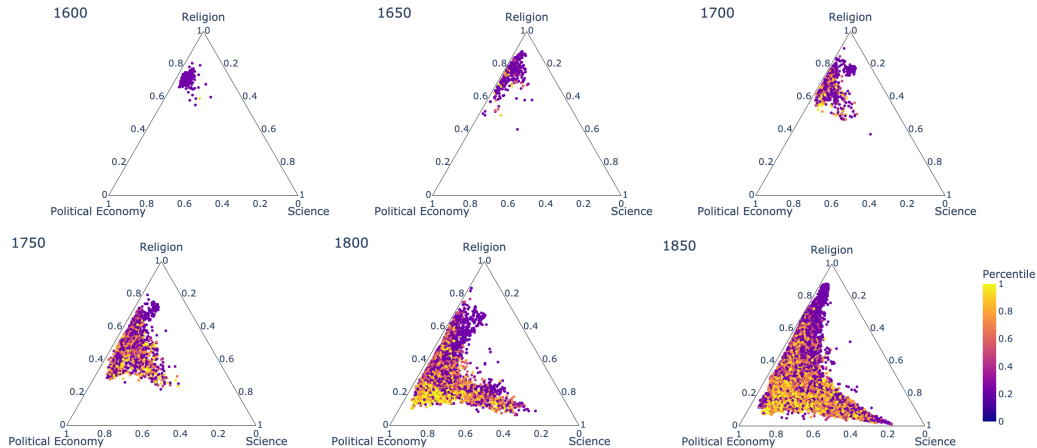


# Average Progress Score

Percentile, 1500–1900



# Volume Progress Sentiment, 1600–1850



# Industrial Scores

APPLEBY'S LIBRARY  
THE PA. STATE  
COLLEGE  
ILLUSTRATED HANDBOOK  
OF  
MACHINERY.

SECTION I.—PRIME MOVERS,

INCLUDING

FIXED, PORTABLE AND MARINE ENGINES, BOILERS,  
LOCOMOTIVES, STEAM LAUNCHES, HEATED AIR,  
GAS AND WATER ENGINES, TURBINES,  
AND WATER WHEELS,

WITH

PRICES, WEIGHTS, MEASUREMENTS, AND SOME DATA ON WORKING  
EXPENSES AND RESULTS OBTAINED.

BY

C. J. APPLEBY,  
(APPLEBY BROS., LIMITED)

WORKS: EAST GREENWICH.

LONDON OFFICE: 48, CANNON STREET, E.C.

THIRD EDITION.—REVISED AND ENLARGED

LONDON:

E. & F. N. SPON, 125, STRAND.  
1880.

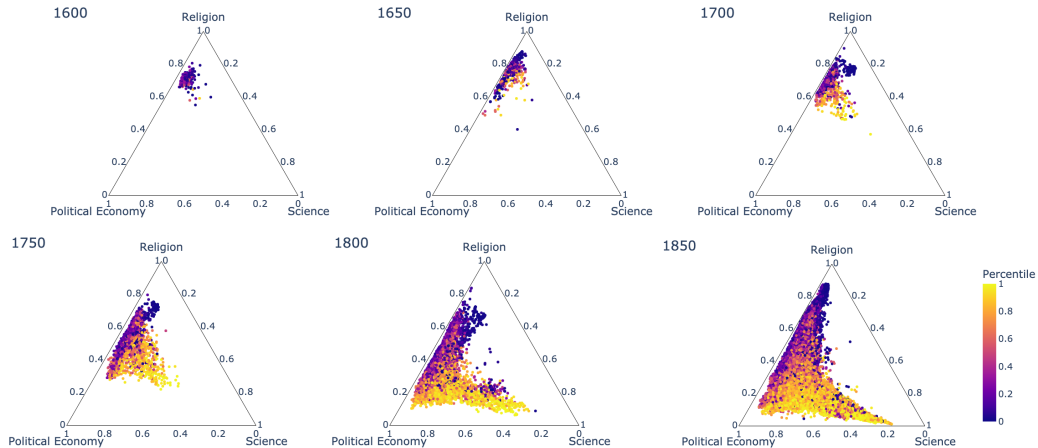
Digitized by Google

Digitized the detailed indexes of *Appleby's Illustrated Handbook of Machinery*, volumes 1–5

- ▶ Published between 1877 and 1903
- ▶ Provide schematics, mechanical details, measurements, prices, etc. for a wide range of industrial machines
- ▶ Volumes titled “prime movers” (volume 1), “hoisting machinery” (volume 2), “pumping machinery” (volume 3), “machine and hand tools” (volume 4), and “steam and electric plant” (volume 5)
- ▶ Cover all types of industrial machinery; indexes are extremely detailed

Derive list of industrial words, weighted by number of appearances in the indexes; remove all words not in use prior to 1643

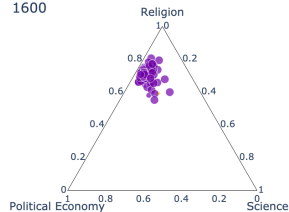
# Volume Industrialization, 1600–1850



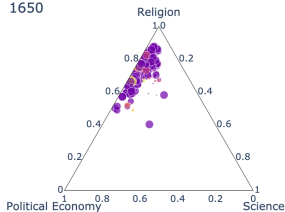
# Volume Sentiment, Bigger Circles = Lower Industry Score

1600–1850

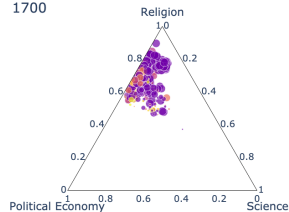
1600



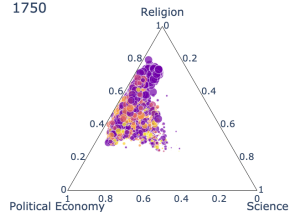
1650



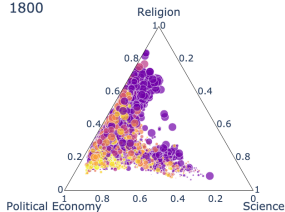
1700



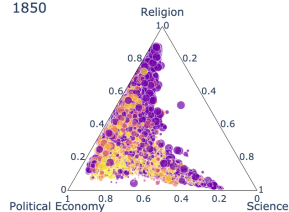
1750



1800



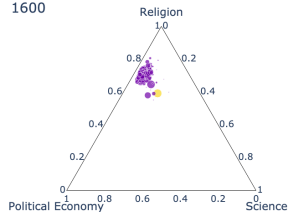
1850



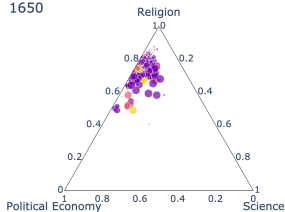
# Volume Sentiment, Bigger Circles = Higher Industry Score

1600–1850

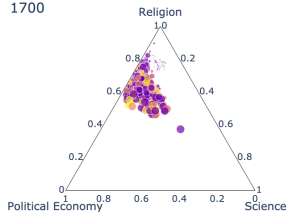
1600



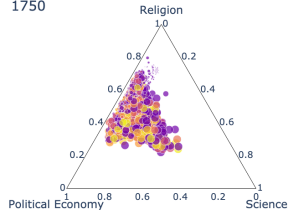
1650



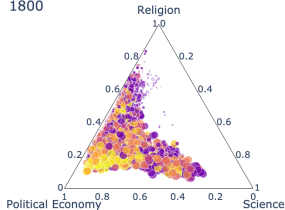
1700



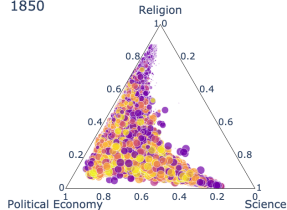
1750



1800



1850



# Industrialization and Progress

**Finding:** Volumes with high industrial scores were more progress-oriented by the latter half of the 18th century, and this is true in all parts of the simplex

- ▶ Volumes with high **science** scores that have low industrial scores were not progress-oriented at any point in the period
- ▶ Volumes with high **religion** scores were more likely to have low industrial scores and not be progress-oriented

# Industrialization and Progress

**Finding:** Volumes with high industrial scores were more progress-oriented by the latter half of the 18th century, and this is true in all parts of the simplex

- ▶ Volumes with high **science** scores that have low industrial scores were not progress-oriented at any point in the period
- ▶ Volumes with high **religion** scores were more likely to have low industrial scores and not be progress-oriented

There is a **compositional effect**: works with high industrial scores were less likely to use the language of **religion** and more likely to be at the nexus of **science** and **political economy**. This compositional effect in combination with the above finding entails that...



# Industrialization and Progress

**Finding:** Volumes with high industrial scores were more progress-oriented by the latter half of the 18th century, and this is true in all parts of the simplex

- ▶ Volumes with high **science** scores that have low industrial scores were not progress-oriented at any point in the period
- ▶ Volumes with high **religion** scores were more likely to have low industrial scores and not be progress-oriented

There is a **compositional effect**: works with high industrial scores were less likely to use the language of **religion** and more likely to be at the nexus of **science** and **political economy**. This compositional effect in combination with the above finding entails that...

**Finding:** Beginning in the latter half of the 18th century, there was a **marked increase in progress-oriented sentiment in volumes at the science-political economy nexus**

## Results in the Context of the Industrial Enlightenment

Mokyr's (2009, 2016) “Industrial Enlightenment” hypothesis suggests that practical works of **science**—those related to **industry**—were becoming more progress-oriented in the build-up to the Industrial Revolution

One of our central findings supports this hypothesis: beginning in the late 18th century and continuing through 1900, volumes at the **science-political economy** nexus—those “practical” volumes key to the Industrial Enlightenment—were much more progress-oriented **if they had high industrial scores**

- ▶ **Science** volumes with low industry scores were not progress-oriented

## Results in the Context of the Industrial Enlightenment

Mokyr's (2009, 2016) “Industrial Enlightenment” hypothesis suggests that practical works of **science**—those related to **industry**—were becoming more progress-oriented in the build-up to the Industrial Revolution

One of our central findings supports this hypothesis: beginning in the late 18th century and continuing through 1900, volumes at the **science-political economy** nexus—those “practical” volumes key to the Industrial Enlightenment—were much more progress-oriented **if they had high industrial scores**

- ▶ **Science** volumes with low industry scores were not progress-oriented

Timing of our finding (early industrialization) is a bit later than that proposed by Mokyr (Enlightenment; pre-industrialization), perhaps reflecting a lag between language used by those with elite human capital and the rest of the population

## Concluding Thoughts

Beginning in the mid-18th century and continuing through 1900, volumes using the language of applied science (those at the science-political economy nexus) that were also related to industry were increasingly progress-oriented

We also find a “secularizaion” of science that emerged in the 18th century and became entrenched in the 19th century

## Concluding Thoughts

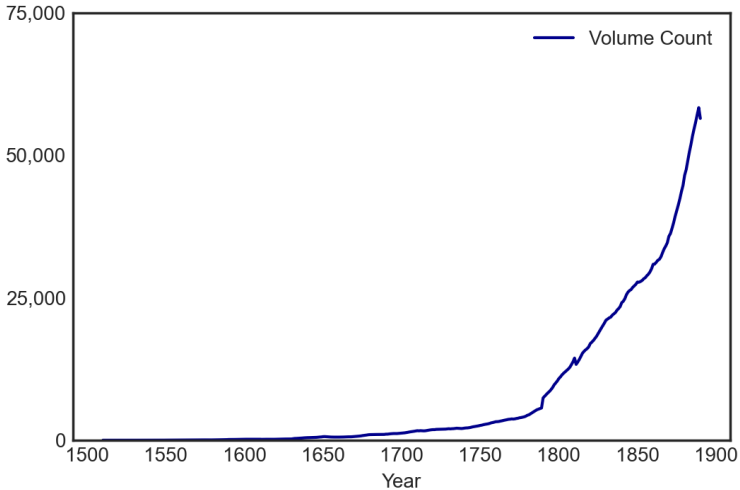
Beginning in the mid-18th century and continuing through 1900, volumes using the language of applied science (those at the science-political economy nexus) that were also related to industry were increasingly progress-oriented

We also find a “secularizaion” of science that emerged in the 18th century and became entrenched in the 19th century

Future work: what is the counterfactual? Were industrial, scientific volumes written in Spanish similarly progress-oriented (ex ante we would expect not)? How about works written in Dutch (ex ante we would expect so)?

# APPENDIX

# Distribution of Volumes, with 20-year smooth



## Data Processing

Each volume grouped by quarter-century of publication (to avoid issues of the meaning of language changing over time)

Data cleaned (e.g., “long-S” correction); stop words (e.g., the, of, and) and short words ( $< 3$  characters) removed

Ranked remaining words by term frequency-inverse document score (Blei and Lafferty 2009): a measure of informativeness that boosts the ranking of words that occur frequently in one volume and less frequently in all other volumes

$$tvidf_v = \frac{1 + \log n_v}{\log(D_v/D)} \quad (2)$$

$n_v$ : number of times term  $v$  appears in a volume

$D_v$ : number of volumes in which term  $v$  appears

$D$ : number of volumes



## Placement of Topics

Each topic receives a **science**, **religion**, and **political economy** score by summing the topic-pair shares of the topic and the topics in the categories

- ▶ e.g., Topic 1 **religion** score is the sum of the topic-pair shares for 1 and 4, 1 and 12, and 1 and 52

Divide the raw category scores for each topic over the sum of all three category scores to derive a category share for each topic

# Volume Classification

We can now classify individual volumes into the three categories

Take each volume and multiply the weight of each topic by the category coefficients for the corresponding topic and year

Yields **science**, **religion**, and **political economy** weights (summing to 1) for each volume in the corpus

Classifying each volume based on its highest weight yields 50,090 **science** volumes, 102,565 **political economy** volumes, and 14,124 **religion** volumes

# An Example of Progress-Oriented Language

17183.  
284

T H E

## *MOTION* of *FLUIDS*,

NATURAL and ARTIFICIAL;

In particular that of the

AIR and WATER,

In a familiar Manner, proposed and proved,  
by evident and conclusive EXPERIMENTS  
with many useful REMARKS.

D O N E

With such *Plainness* and *Perspicuity*, as  
that they may be understood by the UN-  
LEARNED.

For whose Sake there is added,

A Short EXPLANATION of such UNCOMMON  
TERMS, which in Treating on this SUBJECT could  
not, without Affectation, be avoided.

With plain DRAUGHTS

Of such EXPERIMENTS and MACHINES, which, by De-  
scription only, might not readily be comprehended.

---

By M. CLARE, A.M. F.R.S.

---

L O N D O N:

Printed for EDWARD SYMON, at the Corner of  
*Pope's-Head Alley* in *Cornhill*: M,DCC,XXXV.

*The Motion of Fluids, Natural and Artificial*, a 1735 book by  
Martin Clare

A lengthy book on the science of fluid motion, including  
chapters on hydrostatic principles, gravity, cohesion, siphons,  
pumps, engines, and much more

# An Example of Progress-Oriented Language

17. 183.  
THE  
*MOTION* of *FLUIDS*,  
NATURAL and ARTIFICIAL;  
In particular that of the  
AIR and WATER,

In a familiar Manner, proposed and proved,  
by evident and conclusive EXPERIMENTS  
with many useful REMARKS.

D O N E  
With such *Plainness* and *Perspicuity*, as  
that they may be understood by the UN-  
LEARNED.

For whose Sake there is added,  
A Short EXPLANATION of such UNCOMMON  
TERMS, which in Treating on this SUBJECT could  
not, without Affectation, be avoided.

With plain DRAUGHTS  
Of such EXPERIMENTS and MACHINES, which, by De-  
scription only, might not readily be comprehended.

---

By M. CLARE, A.M. F.R.S.

---

L O N D O N:

Printed for EDWARD SYMON, at the Corner of  
*Pope's-Head Alley* in *Cornhill*: M,DCC,XXXV.

Subtitle: In particular that of Air and Water, In a familiar  
Manner, proposed and proved, by evident and conclusive  
Experiments with many useful Remarks. *Done with Plainness  
and Perspicuity, as that they may be understood by the  
Unlearned*

# An Example of Progress-Oriented Language

17. 183.  
284

T H E

**MOTION of FLUIDS,**

NATURAL and ARTIFICIAL;

In particular that of the

**AIR and WATER,**

In a familiar Manner, proposed and proved,  
by evident and conclusive EXPERIMENTS  
with many useful REMARKS.

D O N E

With such *Plainness* and *Perspicuity*, as  
that they may be understood by the UN-  
LEARNED.

For whose Sake there is added,

A Short EXPLANATION of such UNCOMMON  
TERMS, which in Treating on this SUBJECT could  
not, without Affectation, be avoided.

With plain DRAUGHTS

Of such EXPERIMENTS and MACHINES, which, by De-  
scription only, might not readily be comprehended.

---

By M. CLARE, A.M. F.R.S.

---

L O N D O N:

Printed for EDWARD SYMON, at the Corner of  
*Pope's-Head Alley* in *Cornhill*: M,DCC,XXXV.

Preface: The young Philosopher may be assisted hereby, in his first Searches after truth: Besides which Advantage, his Mind will be better prepared for receiving Lectures in Natural and Experimental Philosophy; which, with proper Encouragement, might easily be introduced into Societies, and *made of singular Use and Benefit to Mankind.*

## Two Issues with these Data

1. Most of the digitized volumes are from the 19th century. Training the algorithm on the entire corpus may bias results, given that language and sentiment changes over time

## Two Issues with these Data

1. Most of the digitized volumes are from the 19th century. Training the algorithm on the entire corpus may bias results, given that language and sentiment changes over time
  - ▶ **Response # 1:** Drop pre-1650 books (results are similar)
  - ▶ **Response # 2:** Re-trained the algorithm using equal draws of data from each quarter-century.  
We are currently re-doing the entire analysis using the re-trained algorithm (this takes a while!)

## Two Issues with these Data

1. Most of the digitized volumes are from the 19th century. Training the algorithm on the entire corpus may bias results, given that language and sentiment changes over time
  - ▶ **Response # 1:** Drop pre-1650 books (results are similar)
  - ▶ **Response # 2:** Re-trained the algorithm using equal draws of data from each quarter-century.  
We are currently re-doing the entire analysis using the re-trained algorithm (this takes a while!)
2. There were many more than 173,031 volumes published between 1500 and 1800. What was digitized may be a biased sample



## Two Issues with these Data

1. Most of the digitized volumes are from the 19th century. Training the algorithm on the entire corpus may bias results, given that language and sentiment changes over time
  - ▶ **Response # 1:** Drop pre-1650 books (results are similar)
  - ▶ **Response # 2:** Re-trained the algorithm using equal draws of data from each quarter-century. We are currently re-doing the entire analysis using the re-trained algorithm (this takes a while!)
2. There were many more than 173,031 volumes published between 1500 and 1800. What was digitized may be a biased sample
  - ▶ **Response:** We scraped the catalogue of the English Short Title Catalog ( $\approx 343$ k volumes from 1500–1800). The ESTC includes meta-data (importantly, the subject of the work)

## Two Issues with these Data

1. Most of the digitized volumes are from the 19th century. Training the algorithm on the entire corpus may bias results, given that language and sentiment changes over time
  - ▶ **Response # 1:** Drop pre-1650 books (results are similar)
  - ▶ **Response # 2:** Re-trained the algorithm using equal draws of data from each quarter-century. We are currently re-doing the entire analysis using the re-trained algorithm (this takes a while!)
2. There were many more than 173,031 volumes published between 1500 and 1800. What was digitized may be a biased sample
  - ▶ **Response:** We scraped the catalogue of the English Short Title Catalog ( $\approx 343$ k volumes from 1500–1800). The ESTC includes meta-data (importantly, the subject of the work)
  - ▶ **Results:** no biases in the HDL data with respect to scientific works, although the HDL data has relatively fewer religious works and (slightly) relatively more political economy works
  - ▶ Biases can mostly be accounted for by the ESTC containing ephemera  $\longrightarrow$  if biases exist in HDL data, they are small and in the direction of under-counting religious documents  $\longrightarrow$  our results are an **upper bound** on progress-oriented sentiment for religious works

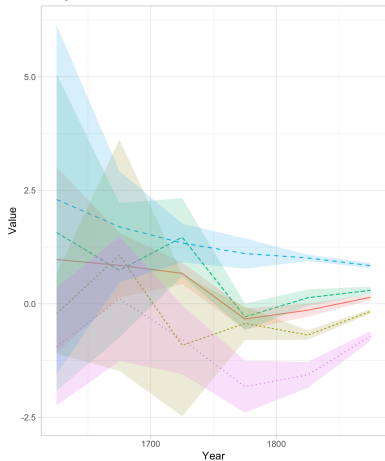
## Regression Analysis

We confirm these observations with a regression analysis (as an accounting exercise), estimating the following equation:

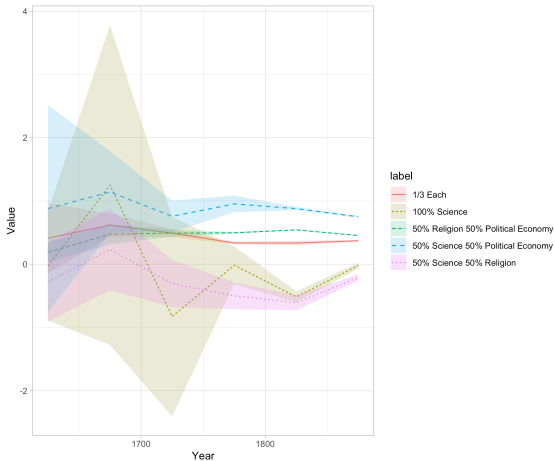
$$\begin{aligned} \textit{Sentiment}_{v,t} = & \alpha_1 + \alpha_2 \textit{Science}_v + \alpha_3 \textit{PolitEcon}_v + \alpha_4 \textit{Science}_v \times \textit{PolitEcon}_v \\ & + \alpha_5 \textit{Science}_v \times \textit{Religion}_v + \alpha_6 \textit{Religion}_v \times \textit{PolitEcon}_v + \lambda_t \textit{Interactions}_{v,t} + \lambda_t + \varepsilon_{v,t} \end{aligned} \quad (3)$$

# Marginal Effects (left) and Predicted Values (right)

**A** Marginal Effects



**B** Predicted Values



## Volume Sentiment Findings

As with topics, the language of individual volumes show a distinction between religious language and scientific language beginning in the latter half of the 18th century

## Volume Sentiment Findings

As with topics, the language of individual volumes show a distinction between **religious** language and **scientific** language beginning in the latter half of the 18th century

Volumes published along the **political economy-science** axis became **increasingly progress-oriented over time**, especially in the 18th century (before flattening out)

## Placebo Test: Optimistic Sentiment

It is possible that our analysis has picked up sentiment that is not necessarily more progress-oriented, but is more broadly optimistic in nature

These are distinct concepts, and they have significant implications for the theory we are testing

## Placebo Test: Optimistic Sentiment

It is possible that our analysis has picked up sentiment that is not necessarily more progress-oriented, but is more broadly optimistic in nature

These are distinct concepts, and they have significant implications for the theory we are testing

The idea espoused in Mokyr (2016) is that the key cultural change associated with the Enlightenment was in how our understanding of the world could be used to **improve the lot of humankind**, not that people spoke of science in “happier” terms

Yet, optimistic language is close enough to progress-oriented language that a change in the former could lead to spurious correlations regarding the latter



## Optimism Dictionary

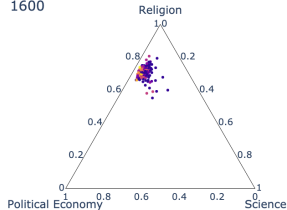
We address this issue by creating an “optimism dictionary,” using the same method as with the progress dictionary, and re-run the analysis

### Optimism Dictionary

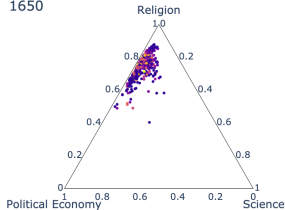
optimistic	optimism	anticipation	assurance
assured	calmness	cheer	cheerful
cheerfulness	cheering	confidence	confident
easiness	elation	encouraged	encouragement
enthusiasm	exhilaration	expectant	happiness
happy	hopeful	hopefulness	hoping
idealism	idealistic	merry	promising
rosy	sanguine	sanguineness	sureness
trust	trusting	utopian	

# Volume Optimism Sentiment, 1600–1850

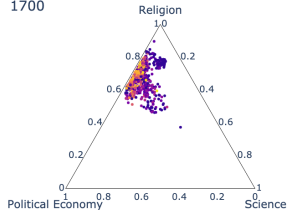
1600



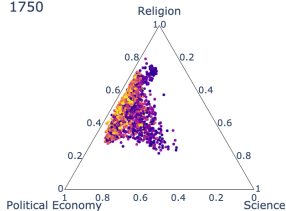
1650



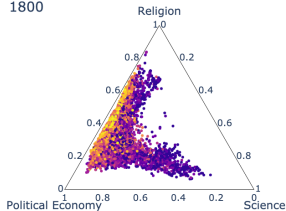
1700



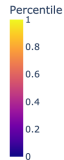
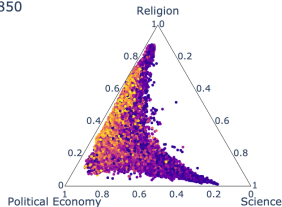
1750



1800



1850



# The Language of Industrialization

Mokyr's hypothesis is more specific than what we have shown above: it relates specifically to how new techniques translated to **industry**

Volumes related to **industrial production** should have been particularly progress-oriented; this is especially true of volumes related to **science**

This is testable! Much like we derived sentiment scores, we derive **industrial scores** for each volume

## Top 10 Industrial Words

Word/Prefix	Count
crane	51
electr	42
weight	37
rope	27
cost	27
water	25
machin	24
coal	23
iron	22
steel	21

Industrial score calculated by multiplying the count of each word in a volume by its corresponding weight, summed across all words with positive industrial weights (normalized by dividing by the total length of the volume)

## Industrialization within Volumes

Across all time periods, volumes using industrial terminology appear most commonly on the science-political economy nexus

# Industrialization within Volumes

Across all time periods, volumes using industrial terminology appear most commonly on the **science-political economy** nexus

Classifying volumes by their industrial score permits a test of the “Industrial Enlightenment” hypothesis

- ▶ According to this hypothesis, views on applied, industrial pursuits using scientific principles became much more progress-oriented in the build-up to Britain’s industrialization
- ▶ Indicates that industrial volumes should have been most progress-oriented (at some point during the Enlightenment) when they were related to **science**