**Online Appendix Table One**

*Industrial Relations Systems*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Britain | Denmark | Netherlands | Sweden |
| Labor-market regime | Pluralist | Macrocorporatist | Macrocorporatist | Macrocorporatist |
| Industrial relations regime (junctures) | Failed National Industrial Conference (1919) | The September Compromise (1899) | High Council of Labor (1919) | The December Compromise (1906);Central Arbitration Board (1920);Labor Court (1929);The Saltsjöbaden agreement (1938) |
| Peak employers’ association  | FBI (1916) but limited power | DO (1896) | VNW (1901) | SAF (1902) |
| ***Labor movement*** |  |  |  |  |
| Working class party  | Labour Party (1900) | Social Democratic Party (1871) | Social Democratic Party (1894) | Social Democratic Party (SAP, 1889) |
| Labor organization  | Sectoral unions | Trade Union Confederation (LO, 1898) | Dutch Association of Trade Unions (1906) | Trade Union Confederation (LO, 1898) |
| Unions | Craft | Craft | Industrial | Industrial |
| Major strikes and labor-market conflicts | 1897 | 1885 | 1903 Railway strike | Sundsvall (1879)Norberg (1891-1892)General strike (1902)General strike (1909)Ådalen (1931) |
| ***Education and skills*** |  |  |  |  |
| Public mass education  | 1870 (private church-affiliated schools since 1810s) | 1814 | 1806 | 1842 (Church-based education since 1686) |
| Folk high schools | N/A | Folkehøjskole (1844) | N/A | Folkhögskola (1868) |
| Vocational training | Market-based | Apprenticeship-based (social partners) | Apprenticeship-based (social partners) | School-based (state) |
| ***Government*** |  |  |  |  |
| Welfare state type (after 1945) | Liberal (residual) | Social-Democratic (universal) | Social-Democratic | Social-Democratic (universal) |
| Social welfare (pre-1945) | Limited unemployment insurance (1911); workers’ accident insurance (1897); old age insurance | Ghent-system for unemployment (1907); workers’ accident insurance (1897); invalidity and old age insurance (1891) | Ghent-system for unemployment (1917); workers’ accident insurance (1901); invalidity and old age insurance (1919) | Unemployment relief (1914); workers’ accident insurance (1901, 1916); old age insurance (1913) |
| Universal suffrage | 1918 (men)1928 (women) | 1849 (men)1915 (women) | 1917 (men)1919 (women) | 1909 (men)1919 (women) |

**Online Appendix Table Two**

*Theories of Industrial Relations Development*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Britain | Denmark | Netherlands | Sweden |
| ***Economic and structural factors*** |  |  |  |  |
| Population (1800) | 10,481,401 | 2,200,000 | About 2,000,000  | 2,347,000 |
| Economic StructuralDevelopment | Early developer | Late developer | Early developer | Late developer |
| Real GDP per capita in 2011$ (1800-1850) (Maddison Project) | Richer3,343 (1800)3,306 (1820)4,332 (1850) | PoorerNo data (1800)2,031 (1820) 2,817 (1850) | Richer4,184 (1800)3,006 (1820)3,779 (1850) | Poorer1,366 (1800)1,415 (1820)1,715 (1850) |
| ***Pre-industrial economic institutions*** |  |  |  |  |
| Serfs | No | Yes, abolished in 1788 | No | No |
| Abolished guilds | 1835, but earlier decline | 1857 | 1800, but earlier decline | 1846 |
| ***Social and political institutions*** |  |  |  |  |
| Stands (*ständer*, *stænder*) | No | Yes | No | Yes, abolished 1865/1866 |
| Religious cleavage | State church/evangelical | State church/evangelical | Catholic/Protestant | State church/evangelical |
| Politicalcameralism | No | Yes | No | Yes |
| Electoral/party system | FPTP (multiparty) | PR (multiparty) | PR (multiparty) | PR (multiparty) |

**Dictionary Word Lists**

We use online dictionaries to identify the words in our major theoretical categories, include all words unless the word has an ambiguous meaning. (For example, the English word “society” almost always refers to high society in the 18th and early 19th centuries; only later does it take on the social connotations that it has today.) We then translate the words, and words commonly used in centuries past. We are native speakers in all but Danish, but two of us are highly fluent in that language. Moreover, we have read quite a bit of fiction from the 18th and 19th centuries in most of our languages. Finally, we look at how individual words perform in the snippets, and choose words that had higher frequencies.

\* Labor words around which snippets are built include:

worker guild craftsman journeyman apprentice farmer peasant serf mechanic labour

Arbejder/arbeider lav håndværk svend Lærling landmand bonder liveg mekanik arbeidskraft/arbejdskraft

Arbeider/gild/ambachtsman/gezel/leerling/boer/landman/lijfijg/monteur/arbeid

Arbet/Skrå/hantverk/Gesäll/Lärling/Bond/husbond/Liveg/mekanik/arbetskraft

\* Cooperation words include:

agreement arbitration bargaining coalition collaboration collective compromise cooperation coordination negotiation pact settlement unanimous unity confederation federation union

aftale voldgift forhandling coalition samarbejde fælles krompromis medvirkning samordning overenskomst forlig ordning enstem enhed forbund forening fagforening

afspraak/bemiddel/onderhandelen/coalitie/partnerschap/collectief/compromis/samenwerk/cöordinatie/overeenkomst/verdrag/regeling/eensgezind/eenheid/verbond/federatie/unie

Avtal/medling/förhandling/koalition/samarbet/gemensam/kompromiss/samverkan/samordning/överenskomm/uppgör/förlikning/enhäl/enig/förbund/förening/fackförening

\* Skills words include:

skill, ability, capacities, qualification, formation, school, teach

Færdigheder evne dygtihed kvalifikation dannelse skole lær

vaardig/vermog/kundig/bekwam/vorming/schol/ler

skick/Förmåg/färd/kvalifikation/bildning/skol/lær

\* Government words include:

nation government ministry authority law legal illegal court judgment judge council commission committee public municipality parish king kingdom crown throne

stat, regering, ministeriet, myndihed, lov, gyld, ulov, ret, vudering, dom råd, kommission, udvalg, offentlig, kommune, sogn, kong, rige, krone, trone,

stat/regering/ministerie/overheid/wet/gerecht/illegaal/rechtbank/vonnis/rechter/rad/commissie/bestur/publiek/gemeent/gemeenschap/koning/koninkrijk/kron/tron

Stat/Regering/Departement/mynd/lag/tillåtet/Olag/rätt/Dom/domm/Råd/kommission/utskott/

offent/Kommun/sock/Konung/Rik/Kron/Tron

\* Market words include:

market sell buy exchange supply demand price cost trade commerce

Marked/sælg/køb/udveksl/udbud/efterspørgsel/pris/kost/handl/handel

markt/verkop/kop/uitwissel/aanbod/vrag/prijs/kost/handel/commercie

Markn/Sälj/Köp/byt/Utbud/Efterfrågan/Pris/Kostn/Handl/Handel

**Computational Text Analysis Methods**

We calculate the frequencies of labor words appearing in the entire corpus for each country, and the frequencies of words associated with our theoretical concepts within snippets of text surrounding our labor words. To calculate word frequencies, we stem the words in the corpora and remove stop words. We construct snippets of 50 words around all of the labor words. The snippets stemmed and filtered; however, the findings are not affected when the snippets include stop words, nor are they significantly affected if we choose snippets of 200 words instead of 50 words around the education words. We use difference of proportion tests to evaluate statistical significance.

Scholars using machine learning or natural language processing analyses in political science typically use either supervised learning techniques or unsupervised learning techniques. Supervised learning techniques may be used to classify works by organizing them into categories or to scale works by determining the position of a document on a linear scale, i.e. from left to right. For example, Hopkins and King (2010) developed a “bag of words” technique with which they hand-code texts into a set of exhaustive categories, and used the output to estimate the proportion of documents into each category. Their goal was to obtain an accurate estimate of document categories and to represent unstructured text as structured variables that can be analyzed statistically. Scholars also use unsupervised-learning techniques such as probabilistic topic models. Rather than beginning with keywords, it is possible to ask the computer to identify the main themes or topics in a collection of documents. A topic is a “distribution over a fixed vocabulary,” and documents have multiple topics, each with their own particular distribution of topics (Blei).

There are some methodological problems with supervised learning techniques. If the labeled set is not a random sample from the population, then it may not reflect the larger population. Moreover, it is difficult to find the true model (Hopkins and King 2010). Thus, scholars also use unsupervised-learning techniques such as probabilistic topic models. Rather than beginning with keywords, the computer is asked to identify the main themes or topics in a collection of documents. A topic is a “distribution over a fixed vocabulary,” and documents have multiple topics, each with their own particular distribution of topics (Blei). There are probabilities of topics appearing in each document and of words appearing in each topic. With processes such as the latent Dirichlet allocation, it is possible to specify the number of topics in advance, use a “topic model algorithm” to infer the hidden topic structure, and compute the distribution of topics that best captures the collection of words in a document. Machine learning defines a joint probability distribution over observed and hidden random variables. One therefore computes the conditional distribution of the hidden variables, given the distribution of the observed variables. This conditional distribution is the “posterior distribution.” In LDA, the observed variables are the words, and the hidden variables are the topic structure. The aim is therefore to infer the hidden topic structure from the documents, which is called computing the posterior distribution (Blei).

In that way, Catalinac uses a probabilistic topic model, the Latent Dirichlet Allocation approach, in her analysis of topics in Japanese electoral manifestos before and after the 1994 electoral reform. She explores whether electoral reform causes candidates to go from targeting the median voter (expected under majoritarian systems) to groups of voters (expected with proportional representation). Rather than specifying topics in advance, she uncovers the topics and estimates the probability that each document covers the topic. She then creates a term-document matrix: words in rows, document identifiers in columns, frequencies in cells. The word frequencies allow her to make inferences about the topics and to create estimates of probabilities that the topic will appear. She bases her assessment of particularistic versus programmatic topics on whether policies benefitted large or small groups (with the particularistic obviously more geared to small groups). She calculated the mean percentage of discussions of particularistic and programmatic goods, and compared the means before and after the reforms. She also applied “wordfish,” which uses word frequencies to infer the locations of a document on an ideological scale (Slapin & Protesch).

We use supervised learning techniques because the differences between the documents we compared were preset and unambiguous. We did not need to scale or to sort the documents, as they had already been sorted into classifications. They constituted corpora of literature in different countries (the cross-national comparison) or works written at different points in time (the temporal differences), and our task was to compare the characteristics of these documents along dimensions that we hypothesized from assumptions about the cultural differences that might give rise to diverse types of education systems. Because the differences among sets of documents to be compared were so rigid, we felt confident in using word frequencies. Moreover, because we were analyzing literature and not tweets or posts, there was likely to be a huge number of topics that had nothing to do with education.

We controlled for possible issues with the structure of the languages. It might be, for example, that a larger number of words are nouns in one language than another, or that the tendency to form compound words in the non-English languages might alter the findings. We therefore compared the English word “give” to the Danish “giver” to gain added confidence that the findings did not merely reflect linguistic differences. Moreover, word frequencies vary in different directions across categories, and this provides added confidence that we were not simply observing linguistic differences.

**Readership**

A set of questions evaluating the impact of literature is clearly concerned with readership. At a minimum, the political class responsible for making public policy should have read literature, if fiction matters to perceptions of education reform. Yet to the extent that public opinion shapes policymaking in pre-democratic regimes, a broader swath of the middle class and even the working class should be exposed to these works for the argument to hold. Except for Britain, our corpora include all available full-text files. For Britain, we chose files that appeared on online lists of works.

Fortunately, recent scholars of digital humanities have undertaken remarkable work tracing the reading of texts, and there is ample evidence that fiction played a huge role in shaping the public’s perception of politics in the 18th and 19th centuries. Our own goals (and capacities) were more modest: as political scientists, our focus was on the possible impacts of cultural artifacts on policy outcomes rather than on the patterns of use of cultural artifacts.

One author of this paper is conducting broader research that focuses on two types of readers. First, I explore how novels shaped intellectual elites’ views of politics and political reforms. Elite reading patterns provide insight into how fiction influences policy decisions in authoritarian regimes, as literature was a site of political struggle and public opinion among elites. The literary community debated the grand issues in salons and taverns on a daily basis, and this debate within the “republic of letters” led to social cohesion and shared values among the reading public. It facilitated debate among authors and spread their views to the broader reading public (Keen, 1999, 29-30). In short, Keen suggests that by the 1780s and 1790s, literature became the “single most effective means by which people could engage each other in a rational debate whose authority all governments would be compelled to recognize.” Thus literature was a kind of “group project” for projecting the groups’ interests onto the public consciousness so that “relations of power would give way to questions of morality” (Keen 1999, 33). William Godwin reflected these views in 1793 when he wrote: “Few engines can be more powerful, and at the same time more salutary in their tendency, than literature. Without enquiring for the present into the cause of this phenomenon, it is sufficiently evident in fact, that the human mind is strongly infected with prejudice and mistake. The various opinions prevailing in different countries and among different classes of men upon the same subject, are almost innumerable; and yet of all these opinions only one can be true. Now the effectual way for extirpating these prejudices and mistakes seems to be literature” (cited in Keen, 1999, 28). Dissenters were particularly pronounced within this group (Keen, 1999 38). Romanticism at the end of the 18th century was a huge political project and, for Shelly, poets were de facto legislators (Keen 1999, 27).

Second, we explore how mass readers were exposed to the cultural assumptions of fictional works. For example, did these cultural assumptions play a role in how the masses thought about industrial institutions and democratic reforms in the 19th century? Of course, if the reading of classic works remained within the elite class, literature may well have an impact on political reforms but be less influential on the development of the mentality of working class people.

In continental Europe, the revolution to read and educate in the vernacular––still a topic in early Danish literature in the 1700s––began in the 16th century with the Protestant reformation. Similarly, Henry VIII wanted all English books to be printed in London (Gilmont, 2003, 215-217). Reading was elevated in Elizabethan England, but the Protestant regime and civil war caused elites to fear literacy among the common man. The Glorious Restoration in 1688 was associated with diminished literacy. Yet the Society for Promoting Christian Knowledge was formed in 1699 to establish charity schools (Altick, 1954, 30-1). The reading public started to expand steadily in the early 18th century. Feather (1988, 90-91) suggests that 50-60% of men were functionally literate by the middle of the 1700s. Watt’s estimate is lower, based on the fact that only one in eleven individuals read newspapers readers in the mid-1700s, but that readership continued to expand to the end of the century. Two groups of people with more limited means were likely to be readers: apprentices and household servants (Watt, 36, 47). The works read during the 18th century were mostly religious ones, but authors such as Defoe and Richardson crossed genres (Watt 50), and there is evidence that the middle and increasingly the working classes engaged with the leading fiction of the times. Keen believes that there was already a sizable working class readership by the late 18th century (Keen, 1999, 37). Certainly, by the 1850s, the reading public fully encompassed a mass audience, made up of between 5 and 6 million people (Altick, 1954, 4-6). Uncle Tom’s Cabin, the biggest literary phenomenon ever, sold 150,000 copies in the first six months (Altick, 1954, 6). Some believe that the Victorian writers were read by a more unified audience, compared with the fragmented readership for 20th century authors. Altick believes that publishers delivered cheaper fare to the working class, but some works such as Dickens’ *Household Words* were very widely read (Altick, 1954, 17-20). Indeed, in the mid-19th century, working class autodidacts were a veritable movement and authors such as Dickens were like the rock stars of their times (Cordner). Just think of the dock workers who upon Dickens’ arrival home from America shouted, “What happened to Little Nell?”

The centrality of reading is reflected in book sales. It is beyond our scope to lay out book sales for all of the volumes in the corpora; however, the data for leading novels are instructive. Altick estimates the English population at 6-7 million in 1750. Fielding published 6,500 copies of *Joseph Andrews* in 13 months, and sold 5,000 copies of *Amelia* in the first week (Altick, 1954, 49). The low sales in part reflected the high prices of books. However, that changed in 1774, when the publisher John Bell created a much cheaper version on coarse paper that was the forerunner of the reprinted six-penny volumes. John Cooke followed suit with a series of British classics (Altick 1954, 54).

*Robinson Crusoe* was written in 1719, before the huge growth in the book trade that began around 1730, when first editions could total up to 10,000 copies (Feather 1988, 90-91). Yet eleven editions of *Robinson Crusoe* were issued by 1759 (“Editions of Robinson Crusoe in English,” 1936, 22). Readership of the book expanded to encompass a broader cross-section of the public when the novel was reprinted in the *Original London Post* (Watt, 42).

The original issue of *David Copperfield* reached total sales of 25,000, with an additional 83,000 copies of the penny version sold in three weeks in 1871. Dickens’ publishers had sold 4,239,000 works by 1882 in England alone (Atlick, 1957, 384.) Hardy’s *Jude the Obscure* sold 20,000 copies in the first three months (Altick, 1986, 238). Readership also expanded through the reprinting of cultural artifacts. Ryan Cordell and David Smith have developed a website allowing readers to trace the circulation of “viral” texts through reprinting during the 19th century (see also Matthew Jockers).

*Table One: Difference of Proportions Results*

*(Frequencies of Cooperation Words in Labor Snippets)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Period | British Percentage | Danish Percentage | Dutch Percentage | Swedish Percentage | Z Score | Significance P value |
| 1770-1820 | 0.0168 | 0.0853 |  |  | -3.129 | P=0.0017 |
| 1770-1820 | 0.0168 |  | 0.0441 |  | -3.889 | P=0.0001 |
| 1770-1820 | 0.0168 |  |  | 0.0406 | -3.22 | P=0.001 |
| 1820-1870 | 0.0251 | 0.0816 |  |  | -3.627 | P=0.0003 |
| 1820-1870 | 0.0251 |  | 0.0315 |  | -3.149 | P=0.0016 |
| 1820-1870 | 0.0251 |  |  | 0.0212 | -1.468 | P=0.142 |
| 1870-1920 | 0.0251 | 0.1286 |  |  | -7.956 | P=0.00 |
| 1870-1920 | 0.0251 |  | 0.0989 |  | -12.479 | P=0.00 |
| 1870-1920 | 0.0251 |  |  | 0.0559 | -5.993 | P=0.00 |

Note: The same British value for the last two periods is correct.

*Table Two: Difference of Proportions Results*

*Frequencies of Skill Only Words in Labor Snippets*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Period | British Percentage | Danish Percentage | Dutch Percentage | Swedish Percentage | Z Score | Significance Level |
| 1770-1820 | 0.0519 | 0.1384 |  |  | -3.383 | P=0.0007 |
| 1770-1820 | 0.0519 |  | 0.1544 |  | -5.566 | P=0.00 |
| 1770-1820 | 0.0519 |  |  | 0.1074 | -2.537 | P=0.01 |
| 1820-1870 | 0.0477 | 0.1074 |  |  | -2.542 | P=0.011 |
| 1820-1870 | 0.0477 |  | 0.0416 |  | 1.128 | P=0.259 |
| 1820-1870 | 0.0477 |  |  | 0.1481 | -8.466 | P=0.00 |
| 1870-1920 | 0.0343 | 0.1883 |  |  | -14.53 | P=0.00 |
| 1870-1920 | 0.0343 |  | 0.0469 |  | -2.102 | P=0.03 |
| 1870-1920 | 0.0343 |  |  | 0.1246 | -12.51 | P=0.00 |

*Table Three: Difference of Proportions Results*

*(Frequencies of Government Words in Labor Snippets)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Period | British Percentage | Danish Percentage | Dutch Percentage | Swedish Percentage | Z Score | Significance Level |
| 1770-1820 | 0.3116 | 0.767 |  |  | -10.409 | P=0.000 |
| 1770-1820 | 0.3116 |  | 0.565 |  | -6.92 | P=0.000 |
| 1770-1820 | 0.3116 |  |  | 0.926 | -10.538 | P=0.000 |
| 1820-1870 | 0.3932 | 0.847 |  |  | -15.744 | P=0.000 |
| 1820-1870 | 0.3932 |  | 0.4288 |  | -2.896 | P=0.004 |
| 1820-1870 | 0.3932 |  |  | 0.4749 | -5.362 | P=0.000 |
| 1870-1920 | 0.3192 | 0.4495 |  |  | -9.745 | P=0.000 |
| 1870-1920 | 0.3192 |  | 0.6318 |  | -17.062 | P=0.000 |
| 1870-1920 | 0.3192 |  |  | 0.5242 | -12.447 | P=0.000 |

*Table Four: Difference of Proportions Results*

*(Frequencies of Market Words in Labor Snippets)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Period | British Percentage | Danish Percentage | Dutch Percentage | Swedish Percentage | Z Score | Significance Level |
| 1770-1820 | 0.1867 | 0.1082 |  |  | 3.536 | P=0.0004 |
| 1770-1820 | 0.1867 |  | 0.3682 |  | -6.494 | P=0.000 |
| 1770-1820 | 0.1867 |  |  | 0.0987 | 2.927 | P=0.003 |
| 1820-1870 | 0.1935 | 0.0615 |  |  | 14.834 | P=0.000 |
| 1820-1870 | 0.1935 |  | 0.3041 |  | -4.723 | P=0.000 |
| 1820-1870 | 0.1935 |  |  | 0.074 | 9.757 | P=0.000 |
| 1870-1920 | 0.1635 | 0.1208 |  |  | 5.448 | P=0.000 |
| 1870-1920 | 0.1635 |  | 0.3261 |  | -10.441 | P=0.000 |
| 1870-1920 | 0.1635 |  |  | 0.0904 | 7.869 | P=0.000 |