Technological Forecasting & Social Change 149 (2019) 119759

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

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

Big data insights into social macro trends (1800–2000): A replication study

Steffen Roth[a,b,*], Peter Schwede[c], Vladislav Valentinov[d], Krešimir Žažar[e], Jari Kaivo-oja[f,g]

[a] La Rochelle Business School, Excelia Group, 102 Rue de Coureilles, 17000 La Rochelle, France
[b] Kazimieras Simonavičius University, Dariaus ir Girėno gatvė 21, 02189 Vilnius, Lithuania
[c] Independent Coder, Dresden, Germany
[d] Leibniz Institute of Agricultural Development in Central and Eastern Europe, Theodor-Lieser-Straße 2, 06120 Halle an der Saale, Germany
[e] University of Zagreb, Department of Sociology, Ul. Ivana Lučića 3, 10000, Zagreb, Croatia
[f] Kazimieras Simonavičius University, Dariaus ir Girėno gatvė 21, 02189 Vilnius, Lithuania
[g] Turku School of Economics, Rehtorinpellonkatu 3, 20500 Turku, Finland

ARTICLE INFO

Keywords:
Big data
Social theory
Social change
Social macro trends
Google ngrams

ABSTRACT

Seeking to advance a big data approach to social theory, Roth et al. (2017) applied the Google Ngram Viewer to explore the way the evolution of the function systems of the modern society is reflected in the Google Books corpus. The authors produced a highly counterintuitive finding that the modern Western societies cannot be adequately described as capitalist. In order to respond to the controversies raised by this finding, the present article replicates Roth et al. (2017) study while using a superior plotting software that allows to control the risk that keyword strength can be biased due to the neglect of keyword quantity. Covering the English-, French-, and German-language corpora, the present replication effort has confirmed the existence of distinct trends exhibited by the individual function systems, such as secularization, the persistent dominance of the political system, and the relatively lesser role of the economic system. These results are largely consistent with those of Roth et al. (2017) and thus lend credence to the authors' sceptical assessment of the validity of the capitalist semantics. The article concludes by pleading for the routinization of big data-driven checks of the modern social theories.

1. Introduction

A recent series of articles in Technological Forecasting and Social Change have discussed impacts and consequences of the digital transformation on social theorizing and the analysis of social change (Guy, 2019; Karakilic, 2019; Ossenwarde, 2019; Palmås, 2019; Rivero, 2019; Roth, 2019; Wenzel and Will, 2019). Despite differences in theoretical or paradigmatic orientation, all authors agree that the digital transformation is changing our concepts of society as well as the way social theories are produced or tested. "Rather than writing manuscripts by hand and submitting them via mail, we use word processors and submit manuscripts online; rather than going to libraries to search for relevant literature, we use online databases that make most of our papers readily available; etc. Especially the latter aspect has rendered the importance of communication in the production and re-creation of academic fields perhaps more salient than ever before" (Wenzel and Will, 2019, p. 1). Rivero (2019, p. 33) notes that age of digitally transformed social theory creates new options for theory debugging. And even critical scholars such as Ossenwarde (2019, p. 25) hold that "(s)ocial theory in general does not necessarily reject technology and information capitalism: in fact, technology, including Windows, Apple and Google tools, may well be used to generate new social theoretic insights".

A striking and counter-intuitive contribution of digitally generated social theoretic insight has been in questioning the widespread assumption about the capitalist nature of the modern Western societies. The recent big data research on social macro trends not only does not support this assumption (Roth et al., 2017) but even explicitly challenges it (Roth et al., 2018). Whereas some of the findings of these authors lend credence to the intuitively accepted social macro trends, such as the secularization or the politicization of societies in the 19th or 20th century, respectively, the "no capitalism" diagnosis is bound to raise eyebrows. Moreover, Roth et al. (2017) research framework has drawn criticism for relying on the allegedly anthropomorphic and biomorphic "global brain" metaphor (Pålmas, 2019, p. 107).

The ambition of this article is, therefore, to take a second look at the big data challenges of social theory by replicating Roth et al. (2017) study. The present replication effort will not only dispose of the global

* Corresponding author at: La Rochelle Business School, Excelia Group, 102 Rue de Coureilles, 17000 La Rochelle, France.
E-mail addresses: roths@excelia-group.com (S. Roth), pschwede@posteo.de (P. Schwede), valentinov@iamo.de (V. Valentinov), kzazar@ffzg.hr (K. Žažar), jari.kaivo-oja@utu.fi (J. Kaivo-oja).

https://doi.org/10.1016/j.techfore.2019.119759
Received 24 June 2019; Received in revised form 24 August 2019; Accepted 23 September 2019

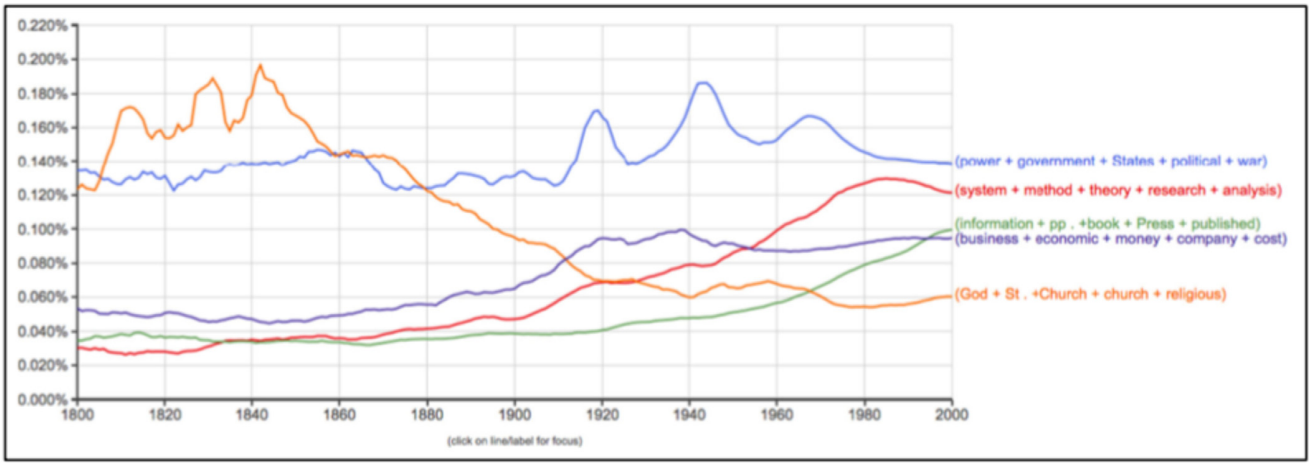

Fig. 1. Combined occurrence frequencies of the five most frequent keywords for political system (blue), economy (violet), religion (orange), mass media (green), and science (red) in the English language Google Books corpus (1800–2000) (Roth et al., 2017, p. 312). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

brain metaphor but will also use an improved methodology. Whereas Roth et al. (2017) relied on the Google Books corpus as a powerful big data source for the analysis of social macro trends in the 19th and 20th century, the validity of their findings is considerably limited by search-term limits inflicted by Google's default plotting software, the Google Ngram Viewer. In our replication study, we seek to overcome this limitation using a different procedure. Most importantly, we shall utilize a different plotting software that allows to control the risk that search-term strength can be overrated due to the neglect of search-term quantity. Thus, the new procedure can take account of the possibility that even if capitalist keywords are not the strongest ones, they might still be the most prevalent keywords in the longer tail of the corpus.

After double-checking the validity of the spectacular "no capitalism" result, we shall re-examine the trends in the relative importance of function systems, such as politics, economy, and religion. If the function systems of politics and economy are found to be mutually independent, this finding would pose a further challenge not only to the Marxist and critical schools of thought but also to a considerable number of mainstream social theories. On this basis, we shall conclude with an outlook on how big data research might be useful for the assessment of the adequacy of (certain pars-pro-toto) definitions of society (as capitalist society or information society) and the corresponding social theories.

2. Hypotheses scrutinising trend assumptions regarding functional differentiation

Roth et al. (2017) compared combined word frequency time-series plots for chunks of pertinent keywords in six languages as found in the Google Books corpus. As is well-known, the Google Books project has created a 'data lake' of over 25 million books and several hundreds of billions of words, thus creating what has been dubbed as "the largest online body of human knowledge" (Wilson-Lee, 2017). To facilitate the analysis of this huge dataset, Google has developed the Google Ngram Viewer, which is both an online search engine and a graphing tool that charts the time series of frequency of any search string ("ngram) as found in a specific language area. Roth et al. (2017) have focused on the English, Russian, German, Spanish, French, and Italian language areas between 1800 and 2000, period for which the data is most reliable.

The conceptual framework of Roth et al. (2017, p. 307) "analysis of social macro trends" (i.e., the secularisation, "economization", politicization, etc. of society) rests on Luhmann (1977, 1982) theory of functional differentiation, which refers to "the decomposition of the modern societies into function systems, such as the economy, law, politics, science, education and others" (Roth et al., 2018, p. 42). Functional differentiation is often regarded as an indispensable feature of modern society by scholars in different fields of research and expertise (Beck and Lau, 2005; Bergthaller and Shinko, 2011; Brier, 2007; Guy, 2013; Kjaer, 2010; La Cour, 2006; Laermans, 2007).

Against this backdrop, Roth et al. (2018, p. 42) argue that the transition from the middle age to modernity

"(M)ust have been associated with the rise to prominence of every function system, with the exception of religion, which has been losing traction over the course of modernisation. The notion of functional differentiation makes clear that the isolated observations of religious or economic trends do not suffice to prove or disprove that modern societies are adequately described as secularised or economy-biased. The observation of an increasing importance of the economy, for example, does not yet imply that the concerned society is actually dominated by the economy. In a similar way, the secularisation trend does not logically preclude that religion retains an important role. The question whether or not modern societies are on the whole characterised by overarching trends can only be decided through the overall comparison of the dynamics of all function systems".

According to the authors, this systems-theoretical framework allows for a comprehensive and comparative analysis of the importance of a total of ten different "function systems": politics, science, mass media, religion, economy, legal system, education, health, art, and sport.

For each of the above-mentioned language areas, the authors plotted time series for combinations of selected keywords that unambiguously referred to one of the function systems. The term "university", for example, was considered as ambiguous as it refers to both science (research) and education (teaching), whereas keywords such as money or theory were thought to refer unambiguously to economy or science, respectively.

As the employed Ngram Viewer can hardly process more than 35 search terms per query, the authors were limited to comparisons between the performances of five most frequent keywords per category. For example, the keyword chunk for science is: (system + method + theory + research + analysis). Five such chunks were then entered into the Google Ngram Viewer. The result of this procedure is depicted in Fig. 1:

Fig. 1 shows the combined word frequency time-series plots for the five most frequent economic, political, scientific, religious, and mass media-related keywords as found in the English language area (1800–2000). Word frequency is an established measure for word importance (Ophir, 2010; Bohannon, 2011; Kloumann et al., 2012). The relatively low performance of economy (as compared to politics and

science) is therefore striking in the context of a plot otherwise depicting trends that are congruent with established historical knowledge and assumptions on the importance of the respective "function systems" as well as the corresponding social macro trends (such as the secularisation of the Western societies or the visible interactions between the World Wars and the importance of politics). As the situation of the economy is similar in all language areas except for the French one, the authors conclude that their data do not support but rather challenge the idea an economy-dominated or capitalist society.

As it is our ambition to replicate and scrutinize Roth et al. methodology (2017) and capitalism-sceptical arguments (2018, 2019), we follow the authors' use of Luhmann's concept of functional differentiation as well as their choice to focus on a selection of ten function systems. For pragmatic reasons, however, we excluded sports from our analysis because the authors admitted that they had issues with identifying sufficiently significant and unambiguous sport keywords (see Roth et al., 2017, p. 311).

Specifically, we hypothesize that

1) We shall confirm the general idea and result of trends in functional differentiation. That is, we expect the importance of politics, religion, economy, science, and other "function systems" to feature changes over time.
2) We shall confirm the results indicating a trend of secularisation, particularly in the 19th century, in all language areas.
3) We shall confirm the significant politicisation trend in the 20th century, and we expect this trend to appear even more significant due to our enhanced methodology. We expect this result because, in their critical self-assessment, Roth et al. (2017, p. 315) conceded that their "approach systematically disfavored the stronger function systems that feature not only the more frequent but also simply more keywords in the word frequency lists. Particularly the dominance of the political system might therefore be even more pronounced if we had the means to trace the combined performances of all political keywords". Our methodology has been designed to address this issue.
4) Although Roth et al. (2017, p. 315) claim not to "have any evidence that the relative performance of the economy would be increased if we combined all economic keywords" (and not only the five most frequent ones), the previous result indicating the only marginal importance of economy, and hence the empirical basis of their capitalism-sceptical argument, will not be replicated.

3. Sample. The English, French, and German version of the Google Books corpus 2012

As Roth et al. (2017) did, we analysed the word frequencies of pertinent keywords as found in the July 2012 language-specific versions of the Google Books corpus (available at http://storage.googleapis.com/books/ngrams/books/datasetsv2.html, accessed 18.04.2019). Due to a different constellation of language skills among the present team of co-authors, however, we limited our sample to the English-, French-, and German-language corpora and abstained from replicating the results for the Spanish, Russian, and Italian language areas. We hold that a replication of the results of three out of six language areas will be enough to check the earlier results for systematic errors.

4. Procedure A programme architecture for comprehensive time series of functional differentiation

We used the English-, French-, and German-language word frequency lists Roth et al. (2017) had extracted from the Google Ngram corpora using Python. For the first 2000 ngrams within these lists of 10,000 words each, all the ngrams with an unambiguous reference to one of the function systems had already been identified and coded by the authors. We double checked and decided to adhere to Roth et al.'s coding scheme. We then designed a computer program that repeated the following steps for each language area (*$lang*):

1) For each letter *$k*, download the 1-gram frequency as found in the corresponding language area: http://storage.googleapis.com/books/ngrams/books/googlebooks-$lang-all-1 gram-20,120,701-*$k*.gz
2) Download the total counts of the http://storage.googleapis.com/books/ngrams/books/googlebooks-$lang-all-totalcounts-20120701.txt
3) For each ngram in each 1-gram frequency file, either replace the ngram with its "function system" category, as assigned in the language-specific word frequency list, or remove it if none is assigned
4) Divide the frequency of each ngram by the total counts of the same year
5) Compute the sums of frequencies for each year and each category
6) Plot a line chart with one line per category

The tools used for this procedure encompass *GNU Bash, GNU awk, mawk, GNU gzip, wget, pv, gnuplot*, and a customized *color brewer palette* for *gnuplot*. The full code of the corresponding program is available in the annex of this article.

5. Results

As a result of the above procedure, we were able to plot the combined frequencies of all the keywords for each "function system" as found within the sample of the 2000 most frequent ngrams per language area.

5.1. English language area

In the case of the English language area (see Fig. 2), religion and politics are the most important function systems in the first half of the 19th century before we observe a decline in religion starting in approximately 1850. Religion is ousted as the second most important system by the turn of the century and, for its part, trumped by science in the 1970s. There is a modest yet continuous rise in the importance of the mass media system, which outperforms education and ranks fourth as of the mid-1970s.

In comparing our results, we find that the dominance of religion in the first part of the 19th century is less pronounced than was suggested by the previous results. Moreover, the previous study suggested that science overtook economy as early as the early stage of the cold war, whereas our research suggests that this turn towards science did not occur before the 1970s. The post-1900 rise of economy is also more significant than in the previous case. Moreover, in our case, the distances between politics, science, and economy are less pronounced by the end of the sample period, and the rise of the information age is much less significant. On the other hand, some of the most basic story lines are the same for both studies. These story lines, which include the declining importance of religion, an overall dominant position of politics, significant interactions between the world wars and politics, a moderate importance of economy in the 19th century, and a rise in the importance of science, particularly in the second half of the 20th century, are the same as those in the previous study.

5.2. French language area

In the French language, politics is by far the most dominant function system throughout the entire sample period (see Fig. 3). Unlike the case of the English and the German language areas, politics' main adversary in the early 19th century is not religion but the legal system, which continues to play the role of the second most important system until it is taken over by science in approximately 1940. A little later, economy

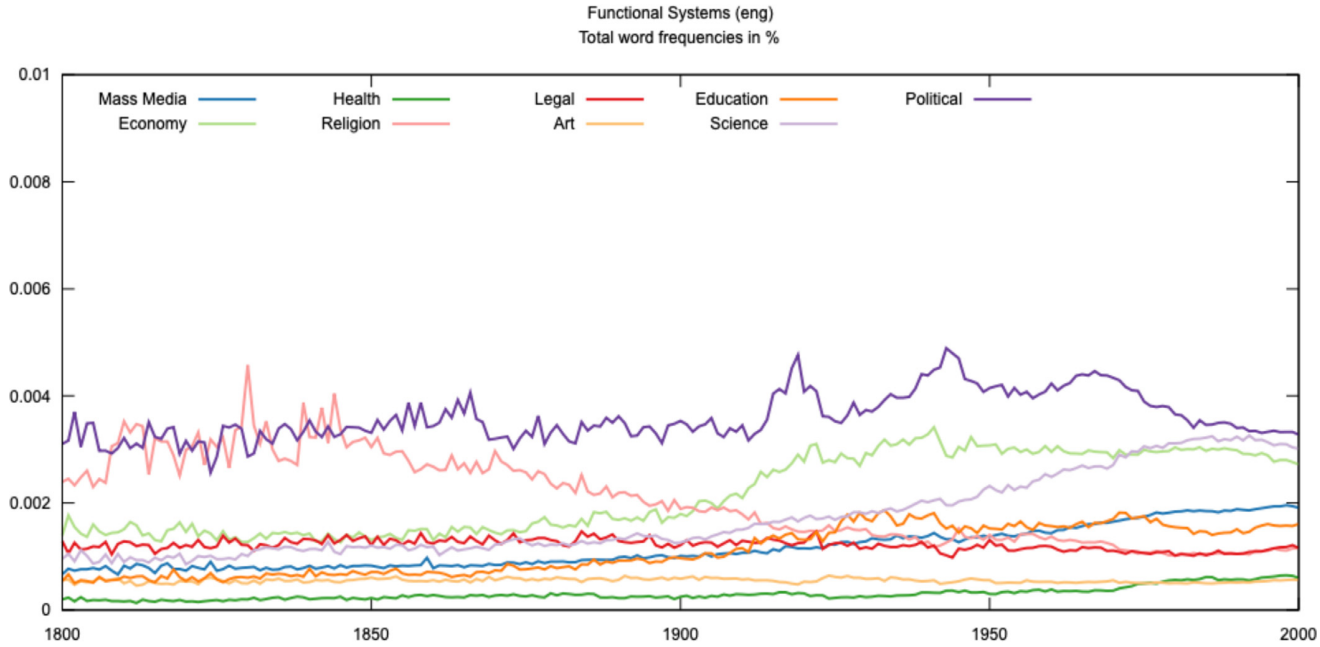

Fig. 2. Combined occurrence frequencies of all the "function system"-specific keywords as found in the English language Google Books corpus (1800–2000).

also overtakes the legal system. The decline of religion is the least significant in all three of the language areas, and the rise of the mass media system is very modest. The most prominent systems of the second half of the 20th century are, in order, politics, science, and economy, where the distance between politics and the other two systems is very significant.

A major difference between Roth et al. (2017) result and our result resides in the fact that, in the earlier case, the dominance of politics was not observed before the outbreak of World War I. As a result, the earlier study suggested a period from approximately 1870 to WW I in which law was the most important system. Moreover, unlike the earlier case, our research suggests that science and not economy is the second most important system in the second half of the 20th century. Major similarities between the two studies include the overall strong dominance of politics and the significant interaction between politics and WW I, with a concomitant absence of a similar interaction for WW II. Moreover, both studies confirm a relatively high (though declining) importance of the legal system as well as a comparably low significance of religion and a less pronounced process of secularisation.

5.3. German language area

Similar to the French case, the German language features the continuous dominance of politics throughout the entire sample period, which appears somewhat contested only in the first decades of the 19th century. The decline of religion is visible, though it is less pronounced

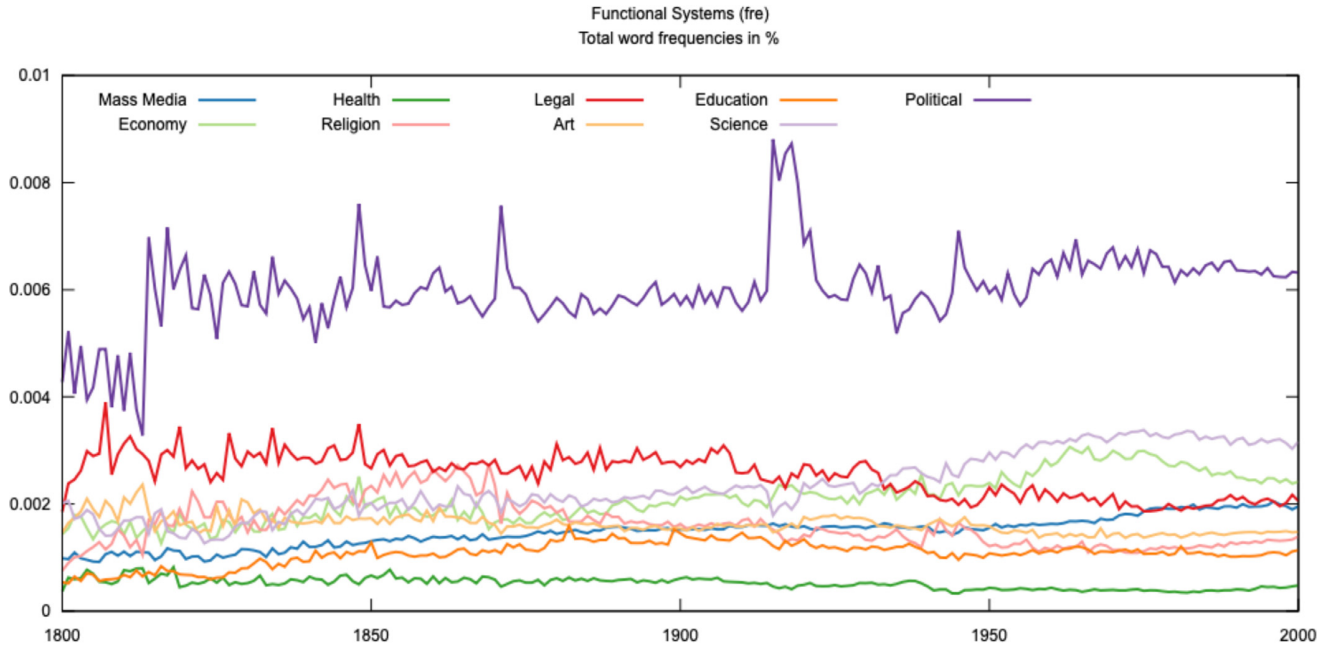

Fig. 3. Combined occurrence frequencies of all the "function system"-specific keywords as found in the French language Google Books corpus (1800–2000).

than in the English case, and there is a spike after WW II when religion is again the second most important system for a very short time before it is overtaken by science, which remains in the 2nd position until the end of the sample period. Economy appears as and remains the third most important system as of approximately 1950. Like with legal system, economy played a significant though intermittent role throughout the entire period.

The continuous dominance of politics in our results is a major difference from the results of the previous study, where the declining religion remained the most dominant system until the turn of the 20th century. There is also no "interregnum of the legal system" (Roth et al., 2017, p. 315), with a dominant position of law in the last quarter of the 19th century (though the legal system is confirmed to be the second most important system in precisely this period). Further differences are less significant, whereas the similarity between both studies comprises the political growth pattern of the 20th century, the religious spike after WW II, and the picture in the second half of the 20th century, where politics is first, science second, and economy third.

6. Replicating functional diversity. A contrastive reading of the results

Whereas Roth et al. (2017) analysis was limited to only the five most frequent keywords per function system, our research included all the function system-specific keywords within the sample of the 2000 most frequent ngrams per language area in the sample period. As a result of our enhanced approach, the dominance of politics appears to be even more pronounced than it had in the previous study, evidently because politics has not only the *most frequent* but also the *most* keywords in the sample. We have also seen fairly significant shifts in the importance of science and economy in the case of the English and French language areas. Whereas Roth et al. (2017) research suggested that economy is only the fourth most important system in the English language area by the end of the sample period, economy ranks third in our case. Conversely, Roth et al.(2017) study suggested that economy was second in the French language area by the end of the sample period, whereas in our study, it is also ranks third.

Setting aside these and a number of smaller differences, however, our results mostly confirm those of Roth et al. (2017). This confirmation is particularly noteworthy in view of our four hypotheses:

1) We hypothesized that we would confirm the general observation of trends in functional differentiation, i.e., the circumstance that the importance of the different function systems changes over time. As the importance of all function systems sometimes features more or less pronounced changes over time in all three language areas, this first hypothesis is corroborated. There are trends in functional differentiation that can be made visible using big data research methods.
2) We also hypothesized that our results would confirm the idea of a secularisation of the investigated language areas in the 19th century. In line with the previous results, this expected trend is confirmed, particularly for the English language area, whereas the trend now appears less pronounced in the German language area and much less pronounced in the French language area. In particular, the latter result is in line with the secular post-revolution history of that language area. Our second hypothesis is therefore (partly) confirmed for those language areas in which secularisation could still occur.
3) In line with our third hypothesis, our research showed that a focus on not only the strong but rather on all function system-specific keywords in our samples further increased the dominance of political system, which has by far been the most dominant system throughout the entire sample period in all the investigated language areas. This result corroborates and further radicalises those of the previous study and supports the idea that we need to conceive of the modern era as a heavily politicised era.
4) Although our additional focus on not only the prominence but also the number of function system-specific keywords points at a higher importance of economy, particularly in the English language area, than was suggested by Roth et al. (2017) previous study, this effect has been even stronger for science, which is why our fourth hypothesis, regarding a significant increase in the relative importance of economy compared to the other systems, is not corroborated. By the end of the 20th century, economy had been no better than the third most important system in all the language areas for several decades. The only likely not completely negligible modification is that the period in which economy was the second most important system in the English language area has been longer than was suggested by the original research.

As counter-intuitive as they still may seem, the results of our replication study have corroborated those presented in the previous study by Roth et al. (2017). The empirical basis of the authors' capitalism-sceptical arguments (Roth et al., 2018; 2019) therefore remains valid, or at least plausible, insofar as we have overcome a serious limitation of their research design and can now exclude a systematic keyword-selection bias. In this sense, we hope to have provided slightly more than the "50% of all available evidence" (Miller and Bamberger, 2016, p. 314) that a first replication study necessarily adds to a counter-intuitive and therefore often poorly understood phenomenon.

As economy does not appear to be a particularly dominant system in our research either, Roth et al. (2018) may therefore still speculate that the modern society of the 19th and 20th centuries might not have been a capitalist society after all. This result remains particularly surprising, as our research has also found particularly low values for the importance of economy in the 19th century, part of which is commonly assumed to coincide with a first golden age of capitalism.

Despite our use of the same big data source, however, it is worthwhile to highlight that neither the research by Roth et al. (2017) nor our own research should be confused with a deductive research programme. Rather, what both articles have presented are the results of an explorative approach to big data and is therefore most suitable as a basis for abductive reasoning. This is what (Roth et al., 2018, p. 44) are actually—though only implicitly—suggesting when they concede that it has not been their "ambition to test alternative hypotheses such as the increasing insignificancy of religion or the growing importance of economy in a strictly deductive sense" and that they "just venture the expectation that the frequency of the combined keywords as per function system be both unevenly distributed and subject to change in time, and that the commonly implied loss of significance of religion as well as the turn to a capitalist society be somehow reflected in the English language subset of the largest online text corpus worldwide".

7. Insistence on capitalism? Outlook on routine checks of prominent social theories

Whereas the results of our replication study largely confirm those of Roth et al. (2017), and thus the basis of their capitalism-skeptical arguments (Roth et al., 2018; 2019), it is noteworthy to point at the fact that according to both their and our data there has been a limited period in the English language area where politics has been the most and economy the second-most important function system. If capitalism is understood as a specific focus on or even bias to political-economic issues, then the period between the eve of the first and the shorter or longer aftermath of the second World War may with some right be considered as capitalist (especially if we consider that both capitalist and anti-capitalist ideologies secretly concur in their strong emphasis on political and economic issues). Marxist and other variants of political- and economy-focused theories—including Parsons (1937, 1951) structural functionalism and its comparably over-strong emphasis on politics and economy, e.g., in the context of Parson's AGIL

scheme)—may therefore be observed to correspond well with the strong importance of politics and economy during this period.

The situation is different though for the French and German language area. Whereas the French language area has been defined mainly defined by politics and law throughout the long 19th century, the German language area witnesses a comparably early rise of science to the second rank behind politics.

The most striking finding therefore remains the great convergence of the language areas during the second half of the 20th century, when all three areas are defined by dominant position of politics with science as second- and economy as third-most important function system; and it is interesting to note that this development is congruent with the contemporaneous transition from Marxism to critical theory as guiding paradigm of (heterodox) social theorizing.

In this sense, we might argue that Marxism has been more adequate to the description of societies in the first half of the 20th century, whereas critical theory with its focus on how scientific-technological rationality shapes and moderates the interplay of politics and economy has been more adequate for the description of societies in the second half of the 20th century. The latter assessment, however, is also true for theoretical frameworks such as the triple helix model of industry, government, and university cooperation (Etzkowitz and Leydesdorff, 1995; 2000; Ivanova and Leydesdorf, 2014), which also highlight the critical role of the interplay of politics, science, and economy. In more applied contexts, the same constellation also appears in the form of popular strategic management tools such PEST, which are and produce results that are characterized by the same (over-) emphasis of *p*olitical, *e*conomic, and scientific-*t*echnological issues that moves everything else into the residual *s*-quadrant. Recent extensions of PEST into PESTLE or STEEPLED, however, might be indicators that the narrow focus on politics, science, and economy is out of touch with the key issues of 21st century societies, and there are good reasons why the triple helix model is basically open to extensions or shifts of focus (Leydesdorff, 2012), too. In the same way, social theories of the 21st century will need to be similarly open to big data-driven routine checks. This skeptical attitude to even the most foundational assumptions about the nature of past, present, and future societies is essential for both the quality management and the survival of many a social theory as there clearly is little point in expending our scarce resources for projecting a narrow local mid-20th century vision of a war society into an unknown future which is bound to be different.

Acknowledgements

Two authors gratefully acknowledge financial support from the Research Council of Lithuania and the European Regional Development Fund-ERDF/FEDER (National R&D Project 01.2.2-LMT- K-718-02-0019).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.techfore.2019.119759.

Annex

A1. Script used to plot Figs. 2-4 based on the Google Books ngram dataset available at http://storage.googleapis.com/books/ngrams/books/datasetsv2.html

```
#!/bin/bash set -euo pipefail set -x
HERE="$(readlink -f "$(dirname "$0″)")"
#tmp=$(mktemp -d) tmp="$HERE/debug" mkdir -p "$tmp" function cleanup {
#head $tmp/*
#rm -rf $tmp echo >> /dev/null
} trap cleanup EXIT export LC_ALL=C for lang in "ger" "eng" "fre"; do
```

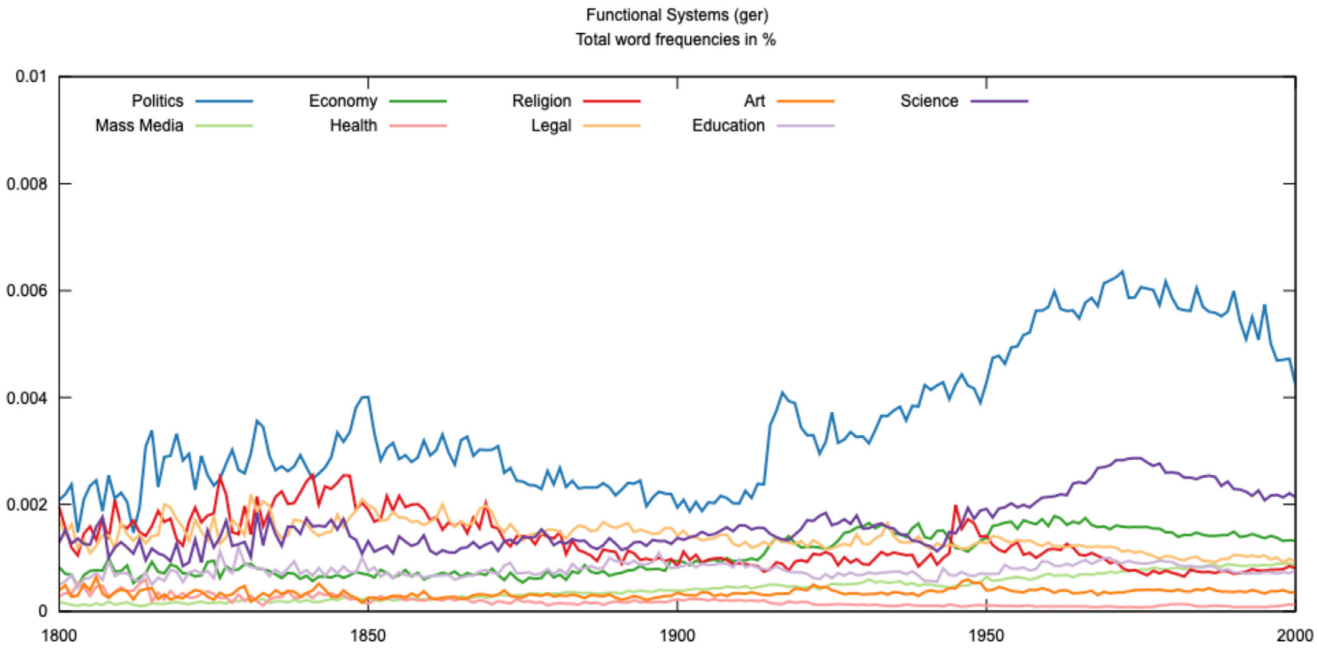

Fig. 4. Combined occurrence frequencies of all the "function system"-specific keywords as found in the German language Google Books corpus (1800–2000).

```
# download ngrams: mkdir -p google/$lang for k in $(echo {a..z}) "other"; do wget –limit-rate "2.5M" -nc -O "google/$lang/$k.gz" "http://storage.googleapis.com/books/ngrams/books/googlebooks-$lang-all-1gram-20,120,701-$k.gz" || ["$?" -le "1″] done
# also download totals and convert to tsv on the fly wget –limit-rate "2.5M" -nc -O "google/$lang/total" "http://storage.googleapis.com/books/ngrams/books/googlebooks-$lang-all-totalcounts-20,120,701.txt" \
&& tr '\t, ' '\n\t\t' < "google/$lang/total" \
| sponge "google/$lang/total" || ["$?" -le "1″]
# create a script that for each word returns the category if existant
# roth/cat/$lang.tsv will be the associative table used in this script generation mawk -F' '
BEGIN{print "#!/usr/bin/awk -f\nBEGIN{FS=\"\t\";OFS=FS;"}
{print "k[\""$1″\"] = \""$2″\";"}
END{print "}{$1 == k[$1];if($1)print $0}"}
' roth/cat/$lang.tsv > $tmp/tocat.$lang.awk chmod +x $tmp/tocat.$lang.awk
# debugging: following lines validate containment of few awaited sample lines
#cat -n "$tmp/tocat.$lang.awk"
#echo "Waffen 2001 2 3″ | $tmp/tocat.$lang.awk
#echo "AAA 2001 2 3″ | $tmp/tocat.$lang.awk
#exit
# create a script that divides the total of occurances by the total
# occurances of all words (year-wise) mawk -F$'\t' '
BEGIN{print "#!/usr/bin/awk -f\nBEGIN{FS=\"\t\";OFS=FS;"}
{print "total[\""$1″\"] = \""$2″\";"}
END{print "}{if(total[$2]!=0){$3=$3/total[$2]}else{$3=\"inf\"}}1″}
' google/$lang/total > $tmp/normalize.$lang.awk
# debugging: following lines validate containment of few awaited sample lines
#cat -n "$tmp/normalize.awk"
#echo "Himmel 2001 2 3″ | $tmp/normalize.$lang.awk
#exit
# go through the list of word-counts and assign a category to each if available
# 1. data rate measurement
# 2. decompress
# 3. categorization (generated above)
# 4. remove categorizations concerning Sport
# 5. divide by total count of all words of that year
# 6. sort by category and year pv "google/$lang/"*.gz \
| zcat \
| $tmp/tocat.$lang.awk \
| grep -v Sport \
| mawk -F$'\t' 'BEGIN{
OFS=FS;
}
{ count[$1 FS $2]+=$3
}
END{ for(k in count) print k,count[k]
}' \
| mawk -f $tmp/normalize.$lang.awk \
| awk -F$'\t' 'BEGIN{
OFS=FS;
}
{ table[$2][$1]=$3; kats[$1]=1;
}
END{ printf "Year"
i==0 for(year in table){ for(kat in kats) { done[kat]++; if(done[kat]<2) { pos[kat]=i++ printf "\t%s",kat
}
}
} printf "\n" for(year in table){ printf "%d",year for(kat in kats) if(table[year][kat]) printf "\t%e",table[year][kat] else printf "\t%e",0 printf "\n"
}
}' \
| sort -r \
> cats-$lang.tsv
# plot relative total frequencies of category counts ncols=$(($(head -n1 cats-$lang.tsv | tr ' ' '\n' | wc -l))) gnuplot -p -e ' load "'$HERE'/../gnuplot-colorbrewer/qualitative/Paired10.plt"; set datafile separator "\t"; set terminal svg size 1000,500; set key left top horizontal; set key autotitle columnhead; set yrange [0:0.01]; set xrange [1800:2000]; set output "cats-'$lang'.svg"; set title "Functional Systems ('$lang')\nTotal word frequencies in%"; plot for [i==2:'$ncols'] "'cats-$lang.tsv'" using 1:i ls i lw 2 with lines;' done
```

References

Beck, U., Lau, C., 2005. Second modernity as a research agenda: theoretical and empirical explorations in the "meta-change" of modern society. Br. J. Sociol. 56 (4), 525–557.

Bergthaller, H., Schinko, C., 2011. Introduction: from national cultures to the semantics of modern society. In: Bergthaller, H., Schinko, C. (Eds.), Addressing Modernity. Social Systems Theory and U.S. Cultures. Edition Rodopi, Amsterdam, pp. 5–34.

Bohannon, J., 2011. The science hall of fame. Science 331 (6014), 143.

Brier, S., 2007. Applying Luhmann's system theory as part of a transdisciplinary frame for communication science. Cybern. Hum. Knowing 14 (2–3), 29–65.

Etzkowitz, H., Leydesdorff, L., 1995. The Triple HelixUniversity-Industry-Government relations: a laboratory for knowledge based economic development. EASST Rev. 14 (1), 14–19.

Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. Res. Policy 29 (2), 109–123.

Guy, J.-S., 2013. Why does society describe itself as global? Re-examining the relation between globalization and the states from a second-order perspective. Curr. Sociol. 61 (7), 1040–1057.

Guy, J.S., 2019. Digital technology, digital culture and the metric/nonmetric distinction. Technol. Forecast. Soc. Change 145, 55–61.

Ivanova, I.A., Leydesdorff, L., 2014. Rotational symmetry and the transformation of innovation systems in a Triple Helix of university–industry–government relations. Technol. Forecast. Soc. Change 86, 143–156.

Karakilic, E., 2019. Rethinking intellectual property rights in the cognitive and digital age of capitalism: an autonomist Marxist reading. Technol. Forecast. Soc. Change 147, 1–9.

Kjaer, P.F., 2010. The metamorphosis of the functional synthesis: a continental European perspective on governance, law, and the political in the transnational space. Wis. Law Rev 2010 (2), 489–533.

Kloumann, I.M., Danforth, C.M., Harris, K.D., Bliss, C.A., Dodds, P.S., 2012. Positivity of the English language. PLoS One 7 (1), e29484.

La Cour, A, 2006. The concept of environment in systems theory. Cybern. Hum. Knowing 13 (2), 41–55.

Laermans, R., 2007. Theorizing culture, or reading Luhmann against Luhmann. Cybern. Hum. Knowing 14 (2–3), 67–83.

Leydesdorff, L., 2012. The triple helix, quadruple helix, …, and an N-tuple of helices: explanatory models for analyzing the knowledge-based economy. J. Knowl. Econ. 3 (1), 25–35.

Luhmann, N., 1977. Differentiation of society. Can. J. Sociol. 2 (1), 29–53.

Luhmann, N., 1982. The Differentiation of Society. Columbia University Press, New York.

Miller, C.C., Bamberger, P., 2016. Exploring emergent and poorly understoodphenomena in the strangest of places: The footprint of discovery in replications, meta-analyses, and null findings. Academy of Management Discoveries 2 (4), 313–319.

Ophir, S., 2010. A new type of historical knowledge. Inf. Soc. 26 (2), 144–150.

Ossewaarde, M., 2019. Digital transformation and the renewal of social theory: unpacking the new fraudulent myths and misplaced metaphors. Technol. Forecast. Soc. Change 146, 24–30.

Palmås, K., 2019. From hacking to simulation: periodizing digitally-inspired social theory. Technol. Forecast. Soc. Change 145, 105–112.

Parsons, T., 1937. The Structure of Social Action 1 McGraw-Hill Book Company, New York.

Parsons, T., 1951. The Social System. Free Press, New York.

Rivero, J.J.B., 2019. The fractal geometry of Luhmann's sociological theory or debugging systems theory. Technol. Forecast. Soc. Change 146, 31–40.

Roth, S., 2019. Digital transformation of social theory. A research update. Technol. Forecast. Soc. Change 146, 88–93.

Roth, S., Clark, C., Trofimov, N., Mkrtichyan, A., Heidingsfelder, M., Appignanesi, L., Mkrtichyan, A., Heidingsfelder, M., Appignanesi, L., Pérez-Valls, M., Berkel, J., Kaivo-Oja, J., 2017. Futures of a distributed memory. A global brain wave measurement (1800–2000). Technol. Forecast. Soc. Change 118, 307–323.

Roth, S., Valentinov, V., Augustinaitis, A., Mkrtichyan, A., Kaivo-oja, J., 2018. Was that capitalism? a future-oriented big data analysis of the English language area in the 19th and 20th century. Futures 98, 41–48.

Roth, S., Schwede, P., Valentinov, V., Pérez-Valls, M., Kaivo-oja, J., 2019. Harnessing big data for a multifunctional theory of the firm. Eur. Manag. J. https://doi.org/10.1016/j.emj.2019.07.004.

Wenzel, M., Will, M.G., 2019. The communicative constitution of academic fields in the digital age: the case of CSR. Technol. Forecast. Soc. Change. https://doi.org/10.1016/j.techfore.2019.05.006. online first.

Wilson-Lee, E. (2017) The man who tried to read all the books in the world. University of Cambridge Research Features, 26. October 2017. Available at: https://www.cam.ac.uk/research/features/the-man-who-tried-to-read-all-the-books-in-the-world/(accessed 18 February 2019).

Steffen Roth is Full Professor of Management at the La Rochelle Business School, France, and Research Professor of Digital Sociology at the Kazimieras Simonavičius University in Vilnius, Lithuania. He holds a Habilitation in Economic and Environmental Sociology awarded by the Italian Ministry of Education, University, and Research; a PhD in Sociology from the University of Geneva; and a PhD in Management from the Chemnitz University of Technology. He is an associate editor of *Kybernetes* and the field editor for social systems theory of *Systems Research and Behavioral Science*. The journals his research has been published in include *Journal of Business Ethics, Journal of Cleaner Production, Administration and Society, Technological Forecasting and Social Change, Journal of Organizational Change Management, Journal of Economic Issues*, and *Futures*. His ORCID profile is available at orcid.org/0000-0002-8502-601X.

Peter Schwede works as an independent coder as well as a systems administrator at the Saxonian State Office of Taxes and Finances in Dresden, Germany. He holds a master's degree in computer sciences from the Dresden University of Technology, Germany.

Vladislav Valentinov is Senior Researcher at the Leibniz Institute of Agricultural Development in Transition Economies in Halle, Germany. He has served on the Editorial Boards of several journals, such as the Journal of Economic Issues, Voluntas, and Kybernetes. He was awarded the title of Adjunct Full Professor by the Martin-Luther-University Halle-Wittenberg, Germany, in 2017. His research has been published in journals such as *Cambridge Journal of Economics, Journal of Institutional Economics, Ecological Economics, Foundations of Science, Administration and Society, Journal of Business Ethics*, and *Nonprofit and Voluntary Sector Quarterly*.

Krešimir Žažar is Assistant Professor at the University of Zagreb, Faculty of Humanities and Social Sciences, Department of Sociology, Croatia. His main research interests are in the areas of general sociology as well as the epistemology of social sciences and (classical) sociological. He is particularly interested in the examination of social dynamics and structural transformations of modern societies. He has co-published 3 books and over 20 scientific and research papers.

Jari Kaivo-oja is one of the leading foresight experts and innovation management researchers in the Nordic countries. He has worked in numerous EU Commission projects in the 6th and 7th Framework programs and recently in the Horizon 2020 programme. He is Research Director at the Finland Futures Research Centre of the University of Turku School of Economics as well as Adjunct Professor at the University of Helsinki and at the University of Lapland. He has worked for the European Commission, for the European Foundation, for the Nordic Innovation Center (NIC), for the Finnish Funding Agency for Technology and Innovation (TEKES), for the EUROSTAT, for the RAND Europe, and for the European Parliament. He has worked in cooperation with the Chinese Academy of Social Sciences (CASS) in the field of global energy and climate policy. His research activities are focused on foresight tools and methodologies, on innovation research and management, on global sustainability analyses, knowledge and ubiquitous society developments, global energy and climate change policies, modern management research and governance theories. He is author of articles published in journals such as *Futures, Journal of Cleaner Production, Foresight, Technological Forecasting and Social Change, Energy, European Integration Studies*, or *Journal of Organizational Change Management*.