THE UNIVERSITY OF WARWICK

July 2015 (Revised Mar 2019) No.236

Historical Analysis of National Subjective Wellbeing using millions of Digitized Books

Thomas Hills, Eugenio Proto, Daniel Sgroi
and Chanuki Illushka Seresinhe

WORKING PAPER SERIES

This paper is a revised version of a paper originally circulated in 2015

REVISED March 2019

Centre for Competitive Advantage in the Global Economy

Department of Economics

CAGE
Competitive Advantage in the Global Economy

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Historical Analysis of National Subjective Wellbeing using Millions of Digitized Books

Thomas T. Hills, Eugenio Proto, Daniel Sgroi and Chanuki Illushka Seresinhe*

Thursday 28$^{\text{th}}$ March, 2019

Abstract

In addition to improving quality of life, higher subjective wellbeing leads to fewer health problems, higher productivity, and better incomes. For these reasons subjective wellbeing has become a key focal issue among scientific researchers and governments. Yet no scientific investigator knows how happy humans were in previous centuries. Here we show that a new method based on quantitative analysis of digitized text from millions of books published over the past 200 years captures reliable trends in historical subjective wellbeing across four nations. This method uses psychological valence norms for thousands of words to compute the relative proportion of positive and negative language, indicating relative happiness during national and international wars, financial crises, and in comparison to historical trends in longevity and GDP. We validate our method using Eurobarometer survey data from the 1970s onwards and in comparison with economic, medical, and political events since 1820 and also use a set of words with stable historical meanings to support our findings. Finally we show that our results are robust to the use of diverse corpora (including text derived from newspapers) and different word norms.

Introduction

Investigating subjective wellbeing is a time-honoured preoccupation across many social science disciplines (1–5). More recently, it has also become the focus of governments and international organizations who see it as an important target for government policy alongside the more traditional focus on national income. The United Nations released the the first *World Happiness Report* in 2011 alongside the OECD launch of the *Better Life Index*. Compared to national income—which has been collected since the 1930s in many nations—subjective wellbeing suffers

*Hills: Department of Psychology and CAGE, University of Warwick, t.t.hills@warwick.ac.uk. Proto: Department of Economics, University of Bristol, CAGE, CEPR, CESifo and IZA, e.proto@bristol.ac.uk. Sgroi: Department of Economics, University of Warwick, CAGE and IZA, daniel.sgroi@warwick.ac.uk. Seresinhe: Alan Turing Institute, British Library, cseresinhe@turing.ac.uk. The authors thank several colleagues for discussions on this and related research, especially Sean Allen, Sascha Becker, Steve Broadberry, Nick Crafts, Ray Duch, Andrew Oswald, Luigi Pascali, Giovanni Ricco, David Ronayne, Jeremy Smith and Thijs Van Rens. This research used cloud computing cloud computing resources kindly provided through a Microsoft Azure for Research Award. This work was supported by the ESRC Centre for Competitive Advantage in the Growing Economy (CAGE) at the University of Warwick, ESRC Grant RES-626-28-0001, and the Alan Turing Institute under the UK Engineering and Physical Sciences Research Council (EPSRC) Grant EP/N510129/1 (awards TU/D/000019). We thank Tomas Engelthaler and Li Ying for research assistance.

from a significant shortfall in the availability of long-run data. Historical approaches have computed national income statistics as far back as 1820 courtesy of the Maddison Project (6) and for some nations we have centuries of additional data (7). By comparison, consistent measures of subjective wellbeing have only been collected since the 1970s.

Here we present and validate a reliable historical measure of national subjective wellbeing, going back 200 years, which enables direct comparisons with Gross Domestic Product (GDP) and other long-run data. To do this, we derive an index from word-use.

Our main index is drawn from the Google Books corpus (8), which is a collection of word frequency data for over 8 million books. Overall, this data provides a digitized record of more than 6% of all books ever physically published (9). We use the words published in these books to compute subjective wellbeing at a given time by using affective word norms to derive sentiment from text. Affective norms are ratings provided by groups of individuals who examine a list of words and rate them on their valence, indicating how good or bad individual words make them feel. Using these ratings, we work through millions of books enumerating the complete published list of Google books by year and by language. Here we present this data for four countries which share a similar level of economic development: the USA, UK, Germany and Italy.

In part F of the Supplementary Material we show that the results presented here are further corroborated by alternative indices created from other independent corpora including “Find My Past" data from the British Library’s “British Newspaper Project" which covers 65 million newspaper and periodical articles from the UK across 200 periodicals going back to 1710, the US English COHA Corpora which includes 400 million words from 1810-2000 and two alternative indices (a “National Pleasantness Index" and “National Polarity Index") derived from SenticNet data.

Our method relies on making inferences about public mood (i.e., sentiment) from large corpora of written text. This method has widespread implications for predicting economic, political, and cultural trends, including recovering large-scale opinions about political candidates (10), predicting stock market trends (11), understanding diurnal and seasonal mood variation (12), detecting the social spread of collective emotions (13), predicting depression in medical patients (14), and understanding the impact of events with the potential for large-scale societal effects such as celebrity deaths, earthquakes, and economic bailouts (15, 16). Our specific approach is widely used (15, e.g.) and directly supported by a study of 17 million blog posts, where (17) found that a simple calculation based on the weighted affective ratings of words was highly accurate (70%) at predicting the mood of blogs as provided by the bloggers themselves. Thus, words with positive valence are taken to indicate positive connotations for the subjective wellbeing of the user, and those with negative valence are taken to have an equivalent negative connotation.[1]

We used the largest available sets of existing word valence rating norms for four languages: English (British), English (American), German and Italian. To allow for comparison across languages, all of our valence norms contain a subset of approximately 1000 words adapted from the "Affective Norms for English Words" (18) or ANEW, which are words chosen in part because they capture the range of emotional sentiment. The original ANEW list served as the basis for developing valence ratings for each of the other languages in our study. Here, we used exclusively the mean valence rating of words.[2] In the Supplementary Material (in Figure A.1) we present a sample of the words covered in all the languages we are considering.

In parts E and F of the Supplementary Material we also show that our results are supported by multiple alternative methods for computing historical sentiment including using only the most stable historical words (that are more resistant to changes in meaning over time), computing time-locked valences for each word, and using independent valence norms from the alternative AFINN word norms (see in particular Tables A.6 to A.10 and Figure A.9).

Using our historical record and word valences, for each language i we computed the National Valence Index (NVI), $NVI_{i,t}$, for each year, t, and language, i, as follows,

$$NVI_{i,t} = \sum_{j=1}^{n} v_{j,i} p_{j,i,t};$$

[1]As with any large dataset, this might not be true for any individual chosen at random or for any individual word, but the power of large data is that averaging over many readers and words, idiosyncratic noise is averaged out.

[2]The Google Book corpora also includes additional languages. For example, French and Spanish are included in the corpus and valence *is* available for these two languages, but our ability to draw sensible inferences for these countries is hampered by the market for books in French and Spanish outside of France and Spain.

where $v_{j,i}$ is the valence for word j in language i, and $p_{j,i,t}$ is the proportion of word j in year t for the language i. The proportion is computed over all words in the corpus for that year for which we have a valence rating. The Google Book database includes books from 1500 to 2009, but the number of books included for the first three centuries is fairly sparse, so we limit our analyses to the period from 1820 to to 2009, for which sufficient data is available (9, 19), and also considering that the complete series of data on national income collected in the Maddison project (6) – that we will use to validate our measure – typically starts in 1820. In addition, for US English and British English the percentage of words covered by our norms stabilizes between 10-12% at around 1800. For German and Italian, the percentage of covered words stabilizes after 1800, at approximately 1%. Despite this stability (see A.2), in all our analyses we control for words covered.

Comparison with Survey-Based Measures of Wellbeing

To validate the NVI, we compare it with existing survey-based measures of subjective wellbeing. The measure of life satisfaction we take as the ground truth is the average per year and per country data taken from the Eurobarometer survey conducted by the European Commission. The question answered was "On the whole, are you very satisfied, fairly satisfied, not very satisfied, or not at all satisfied with the life you lead?", coded on a 4 point scale from "Very satisfied" to "Not at all satisfied." This is the oldest survey available for the countries we used. The first wave covers each year dating back to 1973. It contains data from the UK (104,068 interviews), Germany (102,795 interviews with only West Germany covered before 1990) and Italy (103,789 interviews). Figure 1 shows the relationship between the NVI and aggregate life satisfaction derived from Eurobarometer data for the corresponding country. The data are presented in the form of residuals after controlling for country fixed-effects and have a positive correlation of 0.5508.[3]

The analysis presented in Table 1 shows that the positive relationship is robust to the introduction of GDP, perhaps the most plausible omitted variable. Moreover, the inclusion of year fixed-effects (column 1), to control for the possibility of biases generated by shocks common to all countries in the dataset, and country-specific tends

[3]While the US is not included in the Eurobarometer, there is some fragmented data available from the World Database of Happiness which is positively correlated with the NVI and is presented in the Supplementary Material in Figure A.5.

(column 2), show similar results. In all cases, the coefficient on the NVI is positive and significant. NVI is also a better predictor of life satisfaction than either GDP (correlation = 0.5014) or life expectancy, which we investigate more closely later (correlation = 0.1724), each computed after controlling for country fixed-effects. [4]

As a further validation, consider the following non-parametric exercise: if our measure is valid then the average valence of all the words (taken from the Google books corpora for each of the three countries) that have a frequency correlated significantly and positively with life satisfaction (taken from the Eurobarometer data) should be significantly higher than the average valence of the words that have a frequency correlated significantly and negatively. We find that this is indeed the case. The analysis is described in more detail in part A of the Supplementary Material and the results are presented in Figure A.3, also in the Supplementary Material.

Historical Analysis

In Figure A.8 we show the NVI for the UK, US, Germany and Italy from 1820 to 2009, which is the last year currently available from the Google corpora.[5] The red vertical lines represent key political events in each country as indicated in the figure caption. Their relationship with NVI strongly supports a contemporary historical understanding of these events' impacts on subjective wellbeing. Internal conflicts, such the American Civil War, and the 1848 "Year of Revolution" in Europe, and the two World Wars, coincide with falls in the NVI for the countries affected. In addition, the peak in the US data in the 1920s followed by a downward trend after the Wall Street crash in 1929, supports the view that the crash followed a period of over-optimism in response to sustained economic prosperity.

We compare the NVI to the two welfare indicators for which the longest series of data are available for our countries of interest; namely GDP and life expectancy at birth, with both showing a positive relationship with NVI. We also analyse the effect of internal conflicts, which shows a negative relationship. To do this, we use

[4] Rather than form any prior assumption about the relationship between the mood enshrined in text and the reported subjective wellbeing of individuals we let the data determine that relationship. Given the correlation in Figure 1 and the results in Table 1 we can see a link between individuals who on average report higher levels of life satisfaction and text that features higher levels of valence.

[5] After 2000 there was a change in the book sampling method used by Google (19). This can be observed in Figure A.2 in the Supplementary Material, where we observe a drop in the number of words used (especially for Italian). We checked robustness to excluding or including data from 2000-2009 and found that it made no substantive difference.

per capita GDP for the first analysis in Table 1 containing only observations after 1972 from the Penn dataset (version PWT 8.0) where data are in 2005 international dollars and are adjusted for purchasing power parity. For the historical analysis we use data from the Maddison Project (`http://www.ggdc.net/maddison/maddison-project/home.htm`, 2013 version.) where data are in 1990 international dollars.[6] The other main explanatory variables are the historical data on life expectancy at birth and on internal conflict from the OECD, which are available from 1820 onwards (20). Other variables we use as controls are educational inequality (measured as a GINI index, which we use as a proxy for the inclusivity of the demand for books within society) and the index of democracy (originally, from the Polity IV project) as an index of freedom from the OECD data available from 1820 onwards (20). The data is further summarized in Table A.1 in the Supplementary Material.

The NVI is likely to be affected by the market for literature and, more generally, by the evolution of literature and language (see section B in the Supplementary Material for a discussion). Over the long run, as the target for a typical published book moved from the wealthy elite to the general public, the content of these books changes. Moreover patterns in literary style changed considerably in the early part of the nineteenth century with the advent of literary realism (and social commentary) within literature. To help deal with problems of this sort we include control variables specifically chosen to correct for year-on-year trends. We also use the two alternative econometric specifications presented in Table 2 corresponding to two different hypotheses on the evolution of literature and language: Model 1 assumes that the market for books and language itself evolved in a similar way across the different countries we are considering, hence the introduction of year fixed-effects should correct any source of bias; in Model 2 we assume that the evolution of the market for books and of language itself affects written texts of different languages differently, hence by including country-specific trends we correct any source of bias to the extent that it generates roughly linear trends. Our results show that these two models generate similar findings.

Column 1 of Table 2 shows the NVI's response to a number of explanatory variables. Alongside per capita GDP and life expectancy, we introduced controls for year fixed-effects, words covered, democracy, and education inequality. To give a feel for the relative impact on the NVI, one extra year of life expectancy is worth as much as

[6]The results of the first analysis do not change qualitatively if we use post-1972 data from the Maddison Project instead of the Penn dataset.

4.5% growth in per capita annual GDP. To account for potential lags between changes in the key variables and the appearance of their influence in published text, we empirically determined the lags for each variable based on their influence on the NVI (details are provided in the Supplementary Material, in section B and Tables A.3, A.4, and A.5). Since year fixed effects are heavily correlated with the years in which internal conflicts took place and there are likely to be spillover effects from such conflicts in one country to another, we cannot include both year fixed-effects and a measure of major conflict in the same regression. We overcome this in column 2 of Table 2 where we estimate an alternative specification which replaces the year-specific fixed-effects in column 1 with country-specific trends: this allows us to include internal conflict, which is composed of major conflicts that directly affect the domestic population such as internal unrest or invasions. The use of country-specific trends also helps us to deal with spurious correlation across countries. In column 2, the effects of per capita GDP and life expectancy remain positive, though their magnitudes become smaller, and less statistically significant. This is likely due to the strong trend components in both GDP and life expectancy.[7] The internal conflict variable is negative and significant. In terms of the effect on the NVI, one fewer year of internal conflict is worth as much as 40% more per capita annual GDP growth. In both estimations presented in Table 2, we cluster errors at the country level to calculate standard errors. We also carry out an analysis of possible stochastic trends (including appropriate Augmented Dicky Fuller tests), that might affect the regressions presented in Tables 1 and 2 in part D of the Supplementary Material.

Concluding Remarks

Using conventional regression analysis and non-parametric methods we show that the NVI is highly consistent with existing wellbeing measures going back to 1973 and indicates that on average the valence enshrined in literature matches the mood of the population as represented in published books. We further validate our measure by showing a relationship with variables that are known to have a relationship with wellbeing, such as conflict, life expectancy and GDP. In part F of the Supplementary Material we show that our findings are robust to the use of alternative corpora including newspapers and periodicals and to alternative word norms.

[7]See column 2 of Table A.2 which shows that life expectancy loses significance when we add country-specific trends, even without internal conflict.

The NVI makes visible a number of interesting patterns. For example, there is a rise in subjective wellbeing in Italy and Germany since the 1900s matched by a comparative decline in the UK and USA. However, since the 1970 all four nations, possibly excepting Germany, have seen a steady rise in subjective wellbeing. Internal and external conflicts represent dramatic shocks to subjective wellbeing, but people tend to bounce back following these shocks even if they do not always bounce back to pre-war levels. These observations currently stand as hypotheses, but the NVI makes them possible by presenting psychological history in a form available for explanation. A longer overview of how the NVI has changed in response to major historical events is presented in part G of the Supplementary Material.

It is worth commenting on the relationship between the NVI and historical GDP in the light of the controversy surrounding the link between national income and national happiness, normally referred to as the Easterlin Paradox (3, 21, 22). We note here that in our analysis of the relationship with GDP we find only a very mild relationship across the full duration of the NVI and only very modest increases as a percentage since 1990. It is also worth noting that Easterlin's key work usually considers "happiness" rather than "life satisfaction" and is typically also based on short-duration time-series and cross-sectional data, whereas our findings are based on a first attempt to provide a long-run proxy for life satisfaction. We also know that life satisfaction reacts less to income than measures of more emotional wellbeing such as happiness (23). We therefore argue that our findings lie somewhere in between those who adhere to an extreme form of the Easterlin Paradox and those who believe there is in fact a significant relationship.

Both the market for books, and language itself, have evolved considerably over the period we consider e.g. (24). To help control for this we show in part E of the Supplementary Material that our findings remain in place when we restrict our sample of words to a pool of words that have stable meanings over time. We also argue that this is a similar issue in spirit to the problem of comparing economic growth and income levels across many centuries when lifestyles have changed beyond recognition. Caution is needed when considering *any* long-run socio-economic data, but the utility of having long-run data is hard to overstate. Consider for instance urbanization, huge cultural and political shifts, increased technological advances (mechanization, computerization, mobile telephony, the internet and so on) and countless other important changes that make inter-temporal comparisons of national income challenging but have not prevented the development and widespread use of historical measures of GDP (25). In all cases, there is a need for what historians call a "close read" of the historical literature. Our approach offers the additional contribution of establishing an economic indicator of historical wellbeing and presents an initial foray into quantifying psychological history.

References

1. Di Tella, R., MacCulloch, R. J. & Oswald, A. J. Preferences over inflation and unemployment: Evidence from surveys of happiness. *American Economic Review* **91**, 335–341 (2001).

2. Deaton, A. Income, health, and well-being around the world: Evidence from the gallup world poll. *Journal of Economic perspectives* **22**, 53–72 (2008).

3. Stevenson, B. & Wolfers, J. Economic growth and subjective well-being: Reassessing the easterlin paradox. *Brookings Papers on Economic Activity* 1–87 (2008).

4. Benjamin, D. J., Kimball, M. S., Heffetz, O. & Rees-Jones, A. What do you think would make you happier? what do you think you would choose? *American Economic Review* **102**, 2083–2110 (2012).

5. Proto, E. & Rustichini, A. A reassessment of the relationship between gdp and life satisfaction. *PloS one* **8**, e79358 (2013).

6. Bolt, J. & van Zanden, J. L. The Maddison Project: collaborative research on historical national accounts. *The Economic History Review* **67**, 627–651 (2014).

7. Broadberry, S., Campbell, B., Klein, A., Overton, M. & Van Leeuwen, B. British economic growth, 1270-1870: an output-based approach. Tech. Rep., School of Economics Discussion Papers (2012).

8. Lin, Y. *et al.* Syntactic annotations for the google books ngram corpus. In *Proceedings of the ACL 2012 system demonstrations*, 169–174 (Association for Computational Linguistics, 2012).

9. Michel, J.-B. *et al.* Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).

10. Connor, B., Balasubramanyan, R., Routledge, B. R. & Smith, N. A. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM* **11**, 122–129 (2010).

11. Bollen, J., Mao, H. & Zeng, X. Twitter mood predicts the stock market. *Journal of Computational Science* **2**, 1–8 (2011).

12. Golder, S. A. & Macy, M. W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* **333**, 1878–1881 (2011).

13. Chmiel, A. *et al.* Collective emotions online and their influence on community life. *PloS one* **6**, e22207 (2011).

14. Eichstaedt, J. *et al.* Facebook language predicts depression in medical records. *Proceedings of the National Academy of Sciences* **115**, 11203–11208 (2018).

15. Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A. & Danforth, C. M. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one* **6**, e26752 (2011).

16. Thelwall, M., Buckley, K. & Paltoglou, G. Sentiment in twitter events. *Journal of the American Society for Information Science and Technology* **62**, 406–418 (2011).

17. Nguyen, T., Phung, D., Adams, B., Tran, T. & Venkatesh, S. Classification and pattern discovery of mood in weblogs. In *Advances in Knowledge Discovery and Data Mining*, 283–290 (Springer, 2010).

18. Bradley, M. M. & Lang, P. J. Affective norms for English words (ANEW): Instruction manual and affective ratings. Tech. Rep., Technical Report C-1, The Center for Research in Psychophysiology, University of Florida (1999).

19. Greenfield, P. M. The changing psychology of culture from 1800 through 2000. *Psychological Science* **24**, 1722–1731 (2013).

20. van Zanden, J. L. *et al. How was life? Global well-being since 1820* (OECD Publishing, Paris, France, 2014).

21. Easterlin, R. A. Does economic growth improve the human lot? some empirical evidence. *Nations and households in economic growth* **89**, 89–125 (1974).

22. Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O. & Zweig, J. S. The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences* **107**, 22463–22468 (2010).

23. Kahneman, D. & Deaton, A. High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences* **107**, 16489–16493 (2010).

24. Hills, T. T. & Adelman, J. S. Recent evolution of learnability in American English from 1800 to 2000. *Cognition* **143**, 87–92 (2015).

25. Jerven, M. An unlevel playing field: National income estimates and reciprocal comparison in global economic history. *Journal of Global History* **7**, 107–128 (2012).

26. Alberge, J. Publishers bypass literary agents to discover bestseller talent. Tech. Rep., The Observer, February 22 (2015).

27. Bernard, A. *Rotten Rejections* (Robson Books, 1990).

28. Warriner, A. B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* **45**, 1191–1207 (2013).

29. Schmidtke, D. S., Schröder, T., Jacobs, A. M. & Conrad, M. ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words. *Behavior Research Methods* **46**, 1108–1118 (2014).

30. Montefinese, M., Ambrosini, E., Fairfield, B. & Mammarella, N. The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods* **46**, 887–903 (2014).

31. Recchia, G. & Louwerse, M. Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly journal of experimental psychology* **107**, 1–41 (2014).

32. Bullinaria, J. P., John A.and Levy. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods* **39**, 510–526 (2007).

33. Davies, M. The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International journal of corpus linguistics* **14**, 159–190 (2009).

34. Cambria, E., Poria, S., Bajpai, R. & Schuller, B. Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, 2666–2677 (2016).

35. Veenhoven, R. World Database of Happiness, collection Happiness in Nations, Overview of happiness surveys using Measure type: 121A / 2-Step verbal Lifesatisfaction. *viewed on 2019-01-16 at rlddatabaseofhappiness.eur.nl* (2019).

Figures and Tables

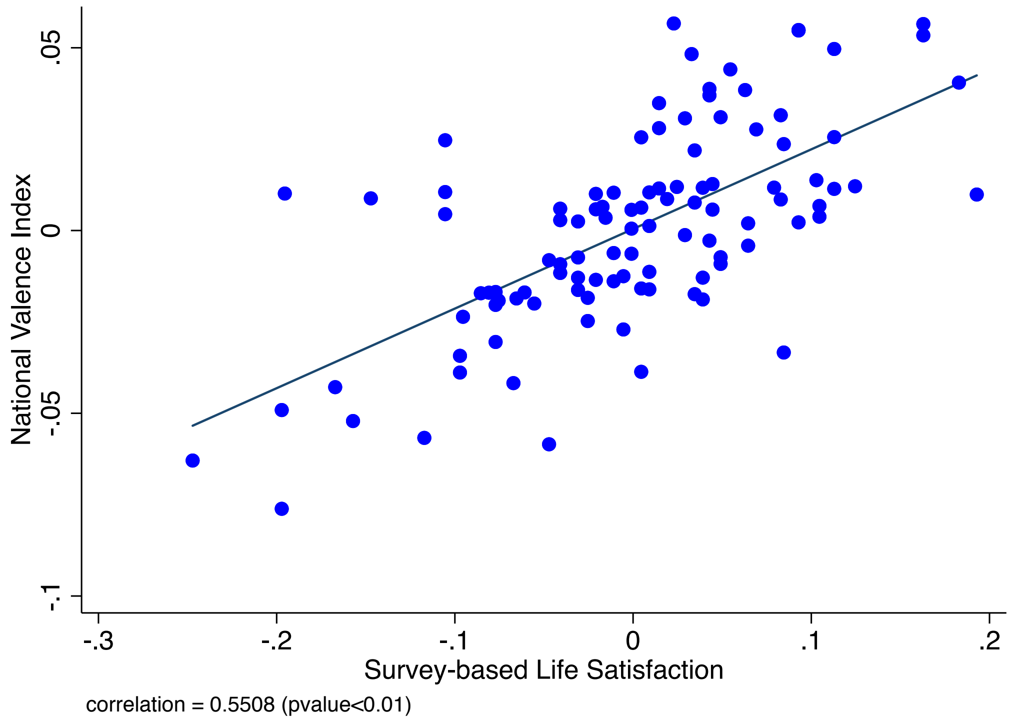


Figure 1: **A Correlation Plot of the National Valence Index against Eurobarometer Aggregate Life Satisfaction.** The scatter plot represents the National Valence Index (our measure of subjective wellbeing derived from digitized text) plotted against aggregate life satisfaction (taken from the Eurobarometer survey-based measure) for the UK, Germany and Italy (the three countries for which both measures exit) from 1973 to 2009 (the period over which both measures are available). Both variables (the National Valence Index and Eurobarometer Life Satisfaction measures) are expressed in the form of residuals after controlling for country fixed-effects, so that values represent variations around the averages for each of the three countries.

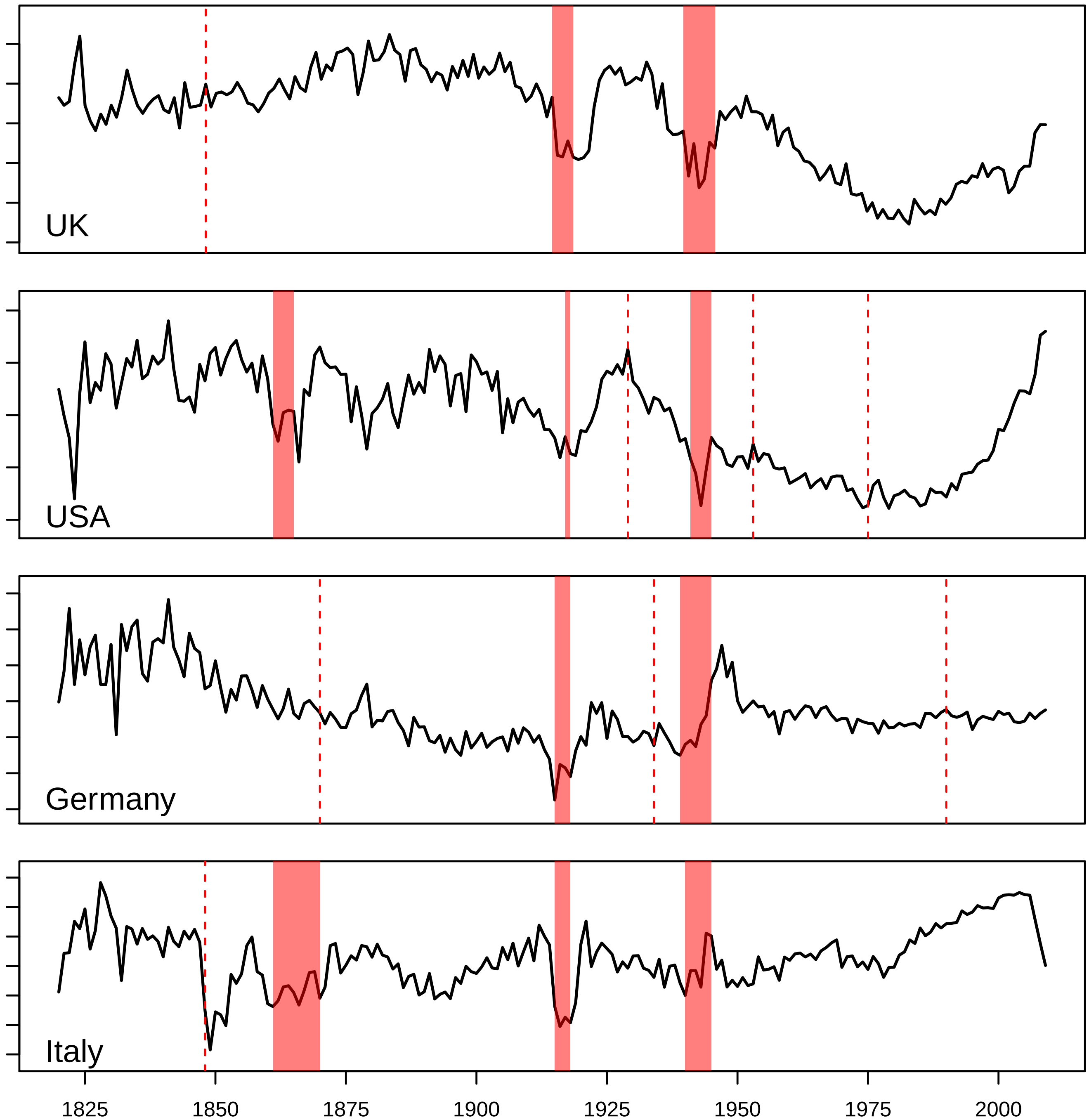


Figure 2: **A Time-Series Plot of the National Valence Index Over the Period 1820-2009.** The figure shows the National Valence Index (our measure of subjective wellbeing derived from digitized text) plotted from 1820 to 2009. Various important events have been highlighted in shaded red (for periods of time) or with a vertical (dashed) red line for events corresponding to a single year. For all countries the red shaded lines include World War I (approximately 1914-18) and World War II (approximately 1938-45). In the 3 European countries a line is drawn in 1848, the "Year of Revolution". In the USA, there is an additional shaded area representing the Civil War (1861-65) and the vertical red lines representing the Wall Street Crash (1929), the end of the Korean War (1953) and the fall of Saigon (1975). For Germany, the vertical red lines represent the end of Franco-Prussian War and reunification (1870), Hitler's ascendency to power (1934) and the reunification (1990). In Italy, there is an additional shaded area representing the unification (1861-70).

	1	2
	Year FE	CS trends
	b/se	b/se
National Valence Index	2.8551***	1.6596**
	(0.2867)	(0.2246)
GDP	Yes	Yes
Country Specific Trend	No	Yes
Year FE	Yes	No
r2	0.730	0.588
N	104	104

Table 1: **The National Valence Index Predicts Aggregate Life Satisfaction.** The table indicates that the National Valence Index (our measure of subjective wellbeing derived from digitized text) is a statistically significant predictor in an OLS estimate of aggregate life satisfaction. The dependent variable is average life satisfaction per country and year taken from the Eurobarometer survey-based measure. The period covered is 1973 to 2009, the period over which both measures exist. The countries considered are Germany, Italy and the UK, the three countries for which both data exist. Per Capita GDP (expressed in terms of purchasing power parity) is from the PWT 8.0 dataset. Column 1 includes year fixed-effects (to help deal with spurious correlations over time) and column 2 includes country-specific trends (to help deal with spurious correlations across countries). Robust standard errors clustered at country levels are given in brackets. Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0826***		0.0698***	0.0550**
	(0.0090)		(0.0106)	(0.0130)
Life Expectancy(t-1)		0.0048**	0.0030	0.0016
		(0.0013)	(0.0014)	(0.0013)
Internal Conflict(t-1)				–0.0184**
				(0.0040)
Words Covered(t)	Yes	Yes	Yes	Yes
Democracy(t)	Yes	Yes	Yes	Yes
Education Inequality(t)	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	No
Country-Specific Trends	No	No	No	Yes
r2	0.752	0.705	0.774	0.571
N	412	412	412	412

Table 2: **Historical Determinants of the National Valence Index from 1820 to 2009.** The table displays an OLS regression of the National Valence Index (our text-based measure of wellbeing). The countries included are Germany, Italy, the UK and the United States. The explanatory variables include per capita GDP (in 1990 international dollars taken from the Maddison Project), words covered (the percentage of all words that are included in the text-derived valence measure) and a variety of measures provided by the OECD, including life expectancy from birth, internal conflict (such as civil wars, revolutions and internal unrest), democracy and education inequality (which offers a control for literacy). The estimation controls for year fixed-effects in column 1 (to help deal with spurious correlations over time) and country-specific trends in column 2 (to help deal with spurious correlations across countries). Internal conflict is not included when year fixed-effects are controlled since year fixed-effects are heavily correlated with the years in which internal conflicts took place and there are likely to be spillover effects from such conflicts in one country to another. The lags of the regressors are empirically determined and details are provided in part B.2 of the Supplementary Material. Statistical significance is indicated as follows: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

Historical Analysis of National Subjective Wellbeing using Millions of Digitized Books
Supplementary Material

Thomas Hills, Eugenio Proto, Daniel Sgroi and Chanuki Illushka Seresinhe

A Correlations between Words and Average Life satisfaction

In this section, we conduct a non-parametric analysis that complements the conventional regression analysis in the main text. First, we calculated the relative frequency of all words for which there is a valence measure for every year. The relative frequency is simply the number of times the word appears in each year t and country j in the Google book corpus data, divided by the average frequency of every word in the same language j and year t; then we select the words for which the level of correlation between the valence and life satisfaction is significant at the usual threshold of the 0.05% level and calculate the averages of the valence across the words correlating positively and negatively.

If the valences of the words carry information about life satisfaction then the average valence of all words that correlate positively with life satisfaction should be significantly higher than the average valence of the words that correlate negatively. This is exactly what the bars of Figure A.3 suggest. Words that correlate positively (negatively) with life satisfaction also correlate positively (negatively) with valence. This indicates that valence is aligned with reported life satisfaction over the period for which both are available.

B The Publishing Industry: Market Forces and Lags

In this section we analyse the possible channels of transmission from events like wars or reflected in GDP and life expectancy through to literature and then to the NVI. We also empirically determine plausible time lags.

B.1 The Publishing Industry

Unless we have reason to suspect some behavioural forces or market failure, economists would normally assume that firms aim to profit maximize. To put this in context, we can think of publishers as fulfilling two roles. First, they attend to the physical (and costly) production of books, which for the period in question almost entirely concerns the manufacture and distribution of printed texts: crucially they cannot publish every book they receive. This leads to their second role, filtering from the mass of submitted books those they wish to publish in order to maximize sales. In this way they act as an intermediary, taking the supply of (largely) unsolicited written books and selecting from them books they feel will match the demand of the reading public.[8] The end result is that only a small minority of authors end up with a publishing contract: some estimates suggest that publishers (and more recently, agents) can receive hundreds or thousands of unsolicited manuscripts a year and might select only a handful. (26) gives two specific examples of publishers' acceptances from unsolicited fiction submissions: 3/5,000 at Jonathan Cape, and 1/400 at HarperCollins. On that basis the text of published books represents a tiny proportion of the words written by all (published and unpublished) authors. The insight from economic theory is that in order to maximize profits publishers filter in a non-random way to match their choice of which books to publish with the demand from potential readers. The positive correlation we find in Figure 1 also indicates that publishers match books typified by predominantly high valence words ("happy books") to "happy people" and books typified by predominantly low valence words ("sad books") to "sad people." Later in this section we will list some quotes from publishers and authors concerning their rationale for rejecting books submitted for publication. The aim is to provide some supporting evidence for the importance of the potential demand-side to publishers.

We first need to note that there is a strong "survivor bias" when examining rejection letters: the vast majority of books that are rejected by publishers will not see print and it is highly unlikely that rejection letters for these books will come to light. The rejection letters that survive tend to be for books which become successful. What is helpful for us is that the bias works in favour of our hypothesis: if publishers are rejecting books that later do become a success on market-based grounds, it seems likely that they are rejecting many more books that never

[8] Recently this role has been partly carried out by "agents" who receive unsolicited manuscripts and select from those they wish to bring to the attention of publishers.

come to print on the same grounds. What follows are a few notable examples for quite famous books which hint at the importance that publishers place on the marketable nature of books and whether books are a good match for readers: note that these authors and books were eventually printed at some later date which might mean that a book was not a good match at one point but later became a better match for the market, or of course that different publishers had different ideas about what might be a good match.

The examples included here are derived from a very much longer list that can be found in (27) and directly relate the decision to reject to demand from the reading population:

- John Gallsworthy's book "A Man of Property" from "The Forsyte Saga" was rejected on the grounds that "The author writes to please himself rather than to please the novel reading public and accordingly his novel lacks popular qualities" and that the book "would have no real sale in this country".
- Simon Brett recalled the following rejection: "I'm afraid the current state of the fiction market is too depressing for me to offer you any hope for this": this could mean that literally the market demanded depressing books but more likely it is a statement that the publisher felt that demand in the market offered no hope to Brett whose work was not a good match. Either way it supports our argument.
- Harlan Ellison recalls having a piece rejected by Playboy magazine because, while the story was "a knockout piece of writing" it did not match the philosophy of action of the "young urban male readership".
- Laurence J. Peter's book "The Peter Principle: Why Things Always Go Wrong" was rejected by McGraw-Hill in 1964 with the following words: "I can foresee no commercial possibilities for such a book and consequently can offer no encouragement".
- Stephen King remarks that he sent three chapters of a book to a publisher before he had published anything else and the rejection informed him that "We are not interested in science fiction which deals with negative utopias, they do not sell".
- Thomas Hardy's book "Tess of the D'Urbervilles" was rejected on the grounds that the readership might be concerned by "improper explicitness".

- Sherwood Anderson's book "Winesburg, Ohio" was rejected on the grounds that readers might find it "far too gloomy".
- George Moore was told about his book "Esther Waters" that it would "hardly go down here" because of certain scenes (such as childbirth) that might upset the potential readers.
- Herman Melville was told that "Moby Dick" would be "unsuitable for the Juvenille Market in [England]".
- Laurence Wylie's chronicle of French country life "A Village in the Vaucluse" was rejected on the grounds that "It is so far from being a book for the general reader that nothing can be done about it".
- Barbara Pym was told after submitting her novel "An Unsuitable Attachment": "Novels like (this), despite their qualities, are getting increasingly difficult to sell." Barbara Pym was also told of her novel "The Sweet Dove Died" that is was "Not the kind of thing to which people are turning."

Finally, note that in part F of this Supplementary Material we also compare the NVI derived from the Google Books corpus with alternative indices derived from other corpora including text taken from newspapers and find that they are positively correlated. We would argue that this is not surprising as newspaper publishers are also driven by the desire to sell newspapers and so match the mood of their readers.

B.2 Different Lags of the Regressors

From the discussion above, we can argue that events happening in one year could feasibly be featured in literature in the same year (if publishers correctly predict the evolution of public mood) or with a lag of several years if publication is time-consuming or delayed. The choice of appropriate lags for the different variables we are considering then becomes an empirical question.

In what follows, we compare different models determining the channels through which a country's subjective wellbeing is factored into the different written languages based on a lag of $t - \tau$ years, with $\tau = 0, 1, 3, 5, 10$.

In the Tables A.3, A.4 and A.5, we present the estimation corresponding to the above models for life expectancy, GDP, and internal conflicts using lags 1, 3, 5 and 10. In GDP the maximal magnitude is at 5 years lag, in conflict the maximal magnitude is at a 1 year lag. For Life Expectancy, it is a bit more complicated, since it goes down after t-1, but then goes up in t-10. We preferred to use t-1 because in t-10 we lose several datapoints. From this specification the resulting lags that best explain changes in the NVI are zero lag for life expectancy, a three year lag for GDP, and zero lag for internal conflicts.

C Valence Norms

For English we used the affective rating norms (28). These norms are a database of nearly 14 thousand English words, all rated on a 1 to 9 valence scale. Each word was rated by 20 participants and the mean valence rating was used for the purpose of our study.[9] The 14 thousand words in the database contain a subset of the 1034 ANEW words. For German we used the affective norms for German sentiment terms (29). This is a list of 1003 words, and German translations of the ANEW list. The valence ratings were collected on a -3 to +3 scale. The mean values were adjusted to reflect a 1 to 9 scale in our analysis. For Italian we used an adaptation of the ANEW norms (30), which contains 1121 Italian words. As with the English words, the ratings were collected on a 1 to 9 scale.

D Stochastic Trends

In this section we discuss the possibility of stochastic trends in both the comparison of the NVI and aggregate life satisfaction data from the Eurobarometer (in Table 1 in the main text) and in the historical analysis of the NVI (in Table 2 in the main text).

D.1 Aggregate Life Satisfaction

In column 2 of Table 1 of the main text, we introduced a control for deterministic trends. However, stochastic trends may also bias our results. To address this issue we used the Augmented Dickey-Fuller unit-root test for stationarity of the NVI from 1970 onwards for all countries separately: the approach we use is typical and involves a null hypothesis defined as the presence of a unit root (a stochastic trend) and the alternative hypothesis of stationarity.

[9] 20 participants may seem to induce concerns about measurement error, but since we are considering a thousand words per participant with unbiased measurement error, this should cancel out in aggregate. Moreover, these ratings have high reliability, which also explains why this dataset has become an important resource within psychology.

The test for a unit root can be rejected in all but Italy (MacKinnon approximate $p - value$ for $Z(t) = 0.6898$), which was integrated of order 1 (so is stationary in differences: see below). For the UK, the unit root can be rejected at 10% confidence levels (MacKinnon approximate $p - value$ for $Z(t) = 0.0696$). For these 3 countries we performed the same test on the life satisfaction variable. For life satisfaction in the UK, the test for a unit root can be strongly rejected (MacKinnon approximate $p - value$ for $Z(t) = 0.0000$). This implies that for the UK a stochastic trend is not a confounding variable in the relationship between the NVI and life satisfaction.

For life satisfaction in Italy the unit root test cannot be rejected (Italy: MacKinnon approximate $p - value$ for $Z(t) = 0.2743$), but can be rejected on the first differences; the two series are then integrated of order 1. Accordingly, there are stochastic trends in both life satisfaction and the NVI for Italy. We therefore tested for cointegration between the NVI and life satisfaction in Italy. The test for cointegration between valence and life satisfaction cannot be rejected: in the residuals of the regression of valence on life satisfaction in Italy the test allows us to reject the existence of a unit root (MacKinnon approximate $p - value$ for $Z(t) = 0.0011$).[10] The existence of cointegration between two variables provides a further test of the existence of a link between these variables, establishing a correlation between long-term shocks in both variables. Hence a permanent shock in life satisfaction is featured in the valence as well.

D.2 Historical Data

In the analysis in Table 2, we addressed the possibility that trends generated by languages, culture or other omitted factors might have biased our initial results. Here we explicitly address the possibility that omitted variables might have generated stochastic trends and biased the correlations presented above. If our estimated life satisfaction and the other regressors are integrated of order bigger than 0, this could potentially be a source of spurious correlation.

We tested the order of integration of our estimated life satisfaction for all languages and years we are considering with the Augmented Dickey-Fuller unit-root test, and we find that for all the presence of a unit root hypothesis can largely be rejected (while, as it is expected, for both GDP and life expectancy the same hypothesis cannot be

[10]The details of all tests can be provided upon request.

rejected).[11]

E Word Stability

In this section we recalculate our main index using a set of words that have stable meanings over time. In order to identify the most stable words over time we use the following process. We use our list of ANEW words for all languages (US English, British English, Italian, German) and compute the positive pointwise mutual information (PMI) vectors using the method employed by Recchia and Louwerse (31) and initially introduced by Bullinaria and Levy (32). For each ANEW word for every year from 1800, the PMI vector is computed as

$$PMI(x,y) = log_2(\frac{P(x,y)}{P(x)P(y)}) \tag{A-1}$$

If we wanted to calculate the PMI for the word “blossom”, then x would be “blossom” and y would be every other word in the ANEW list. $P(x,y)$ would the number of times "blossom" co-occurs with all the different ANEW words divided by the total number of words in the corpus. When calculating co-occurrences we check for ANEW words which co-occur in any 2 word window either before or after word x:

worda wordb blossom wordc worde

$P(x)$ and $P(y)$ is calculated as the frequency of x and y (respectively) divided by the total number of words in the corpus. We then take the log and set any elements containing negative values to zero.[12]

We then see how each word changes over time and calculate the decadal changes over time using the PMI vectors we have computed for each word for every year. We take the cosine distance of word x of $year_t$ and $year_{t+10}$, where t is every year from 1800 to 2009. The cosine distance between any two elements (u,v) is defined as $\frac{1-uv}{u_2 v_2}$.

For each word, we then take all the cosine distance values and calculate the maximum difference. As an extra robustness check, we also checked that our results held when computing the average difference of the cosine

[11] The details of all tests can be provided upon request.

[12] Negative values, i.e. when $P(x,y) < P(x)P(y)$, indicate less than the expected number of co-occurrences, which can arise for many reasons, including a poor coverage of the represented words in the corpus. A potentially useful variation, therefore, is to set all the negative components to zero, and use only the Positive PMI.

distance values for each word.

Finally, in order to identify the most stable words, we take three different methods. We order all our words in terms of average difference or maximum difference and take the top 25% or top 50% where the top words are the most stable. Table A.6 shows the most stable and least stable words identified for each language.

We then recompute our new valence indices by using only the stable set of words identified and the corresponding valence scores from ANEW.

Additionally, we also test how our valence indices by computing a time-locked yearly valence score for each word based on which of the top words word x has co-occurred with. Therefore, for word x, we find the top 15 words that word has co-occurred with every year. We then calculate the valence of word x in $year_t$ as the average valence of its top 15 co-occurring words.

So, taking our word "blossom", the valence for "blossom" in 1800 will be calculating using the ANEW valence from the words 'freshness', 'flourish', 'firewood', 'canvas', 'foliage', 'ripe', 'blooming', 'glossy', 'bosom', 'awning', 'badger', 'girdle', 'pristine', 'mantle', 'gallop' whereas the valence for "blossom" in 2009 will be calculated using the ANEW valence of the words 'foliage', 'blooming', 'lavender', 'magnolia', 'leaf', 'vine', 'wreath', 'fade', 'lily', 'flourish', 'spring', 'tree', 'spray', 'rot' and 'lemon.'

The results of these analyses are shown in Tables A.7 to A.10, with related plots of the NVI using only the most stable 50% or 25% of words in Figure A.8.

F Alternative Corpora and Word Norms

In this section we highlight the similarity between our reported results on the NVI based on text derived from the Google Books corpora using the ANEW word norms and variations based on alternative corpora or word norms.

Firstly in Figure A.6 we recalculate the NVI using the COHA Corpora. The Corpus of Historical American English (COHA), collected independently of the Google Books corpus, represents a balanced and representative corpus of American English containing more than 400 million words of text from 1810 to 1990, by decade, and composed of newspaper and magazine articles (33). Also plotted in the same figure is the NVI based on the Google

Book corpus. The two display a positive correlation of 0.6144 (with a p-value of 0.0051).

In Figure A.7 we once again compare our own NVI based on the Google Books corpus but this time to an alternative derived from the “Find My Past" data from the British Library’s “British Newspaper Project" which covers 65 million newspaper and periodical articles from the UK across 200 periodicals from 1710-1953. There is a positive correlation between the two of 0.4554 (with a p-value of under 0.000). Table A.10 provides a direct comparison of the historical determinants of the two indices for the period 1820-1950.

Tables A.12 and A.13 present a regression analysis of two alternative indices derived from SenticNet data. SenticNet is a well-known resource for sentiment analysis and offers the values for 30,000 concepts in either single word or multi-word expressions (34). The regression analysis mirrors the analysis of the NVI in Tables 2 and A.2.

Finally, Figure A.9 presents a recalculation of the NVI using the alternative AFINN word norm rather than the ANEW word norm used in the main text. The comparison is made for British English and American English and display a positive correlation of 0.9040 and 0.7850 respectively (with p-values under 0.01).

G Overview of the NVI over time

The NVI provides a first attempt to measure changes in national mood over the long-run. It also provides a way to assess how significant historical events affected national mood.

Looking at the UK some interesting patterns emerge. The NVI in the 19th century in the UK is high compared to the 20th century. The index falls with the American War of Independence (1775-83), and the loss of the American colonies, the two World Wars, and the stock market crash of 1929 and the subsequent Great Depression. In the post-World War II period the NVI reached a notable high point in 1957, the year of Harold Macmillan’s speech that most Briton’s had “never had it so good". After that the NVI falls through the 1960s and on into the 1978-79 "Winter of Discontent", with the trend rising back in the late 20th century.

Across all of the countries we consider we can see major historical events being picked up by changes in the NVI. To give a few examples: the Year of Revolutions (1848 for the European countries), the outbreak of World

War I (1914 for Germany and the UK), the Wall Street Crash (1929 for the USA), Hitler takes power (1933 for Germany), the outbreak of World War II (1939 for Germany and the UK), the end of Korean War (1953 for the USA), the end of Vietnam War (1975 for the USA) and German reunification and the end of Cold War (in 1990 for all countries).

H Additional Figures and Tables

This section includes additional figures and tables that are referenced in the main text and in the Supplementary Material

Variable	Mean	Std. Dev.	Min.	Max.	N
National Valence Index	5.798	0.164	5.589	6.128	760
FindMyPast National Valence Index	5.884	0.007	5.859	5.9	131
COHA National Valence Index	5.685	0.029	5.639	5.722	19
Life Satisfaction	2.98	0.181	2.52	3.23	104
Life Satisfaction (US)	1.835	0.033	1.77	1.88	28
per capita GDP (Maddison)	11980.032	11270.36	400	50902	728
per capita GDP (Penn)	25233.999	7193.752	13069.197	43511.594	170
Life Expectancy	61.457	14.088	25.81	82.400	493
Internal Conflict	0.097	0.296	0	1	762
Democracy	5.649	5.894	-9	10	624
Education Inequality	31.526	22.722	6.111	98.935	504
Words Covered Google	0.079	0.068	0.01	0.218	759
Words Covered FindMyPast	0.016	0.001	0.015	0.018	131

Table A.1: **Main Variables.** These are the mean, standard deviation, minimum value and maximum value of the key variables described in the main text.

ENGLISH	VALENCE	GERMAN	VALENCE	ITALIAN	VALENCE
aardvark	6.26	Aas	-2.6	abbaglio	3.94
abalone	5.3	Abenddämmerung	-2.35	abbandonato	2
abandon	2.84	Abendessen	2.1	abbondanza	6.82
abandonment	2.63	Abenteuer	0.81	abbraccio	7.7
abbey	5.85	Abfall	1.44	abete	6.17
abdomen	5.43	abkochen	0.4	abitante	5.67
abdominal	4.48	Abschaum	1.9	abitazione	6.46
abduct	2.42	Abscheu	-1.38	abito	7.27
abduction	2.05	Absturz	-1.6	abitudini	4.91
abide	5.52	absurd	-2.7	aborto	2.06
abiding	5.57	Abtreibung	-2.55	abuso	1.74
ability	7	aggressiv	-1.8	accettazione	5.79
abject	4	aktivieren	-0.6	accogliente	8.03
ablaze	5.15	Alarm	1.5	accomodante	6.4
able	6.64	Alimente	-0.79	accordo	6.71
abnormal	3.53	Alkoholiker	2.15	acqua	7.78
abnormality	3.05	Allee	-1.9	adorabile	7.33
abode	5.28	allein	-1.27	adulto	5.78
abolish	3.84	Allergie	-1.56	aereo	6.56
abominable	4.05	Alptraum	-1.56	affamato	4.74
abomination	2.5	anbetungswürdig	-1.22	affascinare	7.97
abort	3.1	angeekelt	0.73	affaticato	3.73
abortion	2.58	angespannt	1.53	affetto	7.48
abracadabra	5.11	Angriff	-2.1	afflizione	1.94
abrasive	4.26	ängstlich	1	affogare	1.79
abreast	4.62	Anreiz	-1.93	aggressione	2.53
abrupt	3.28	Anstellung	-2.21	aggressivo	3.48

Figure A.1: **A Sample of Word Valence in Different Languages.** For English and Italian the words are scaled from 1 to 9. For Germany the valence ratings were collected on a -3 to +3 scale. The German mean values were adjusted to reflect a 1 to 9 scale in our analysis.

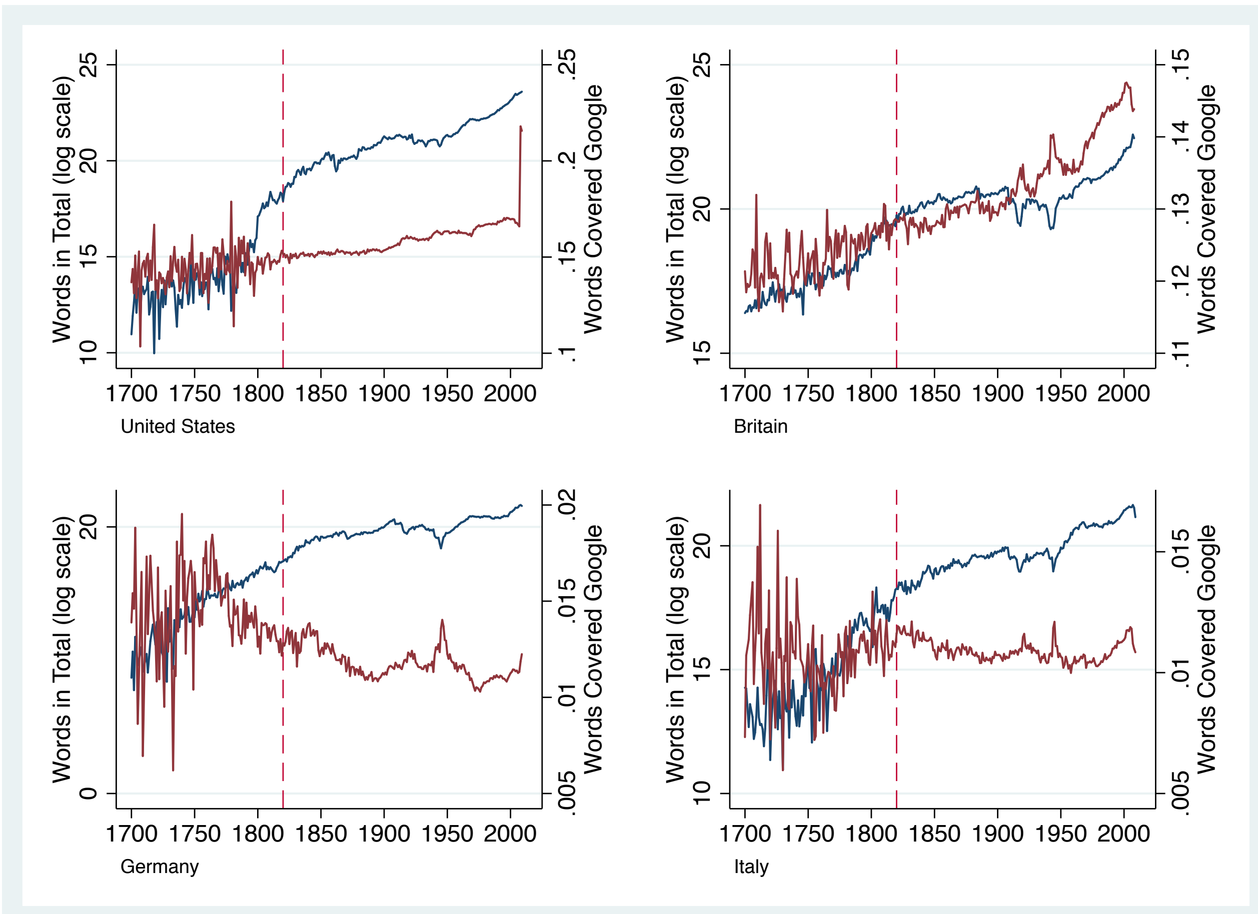


Figure A.2: **The Number of Words and Share of Words Covered.** The red line represents the proportion of words in the corpus covered in the text analysis by the valence norms and the blue line represents the total number of words—in logarithmic scale—for all countries considered in the analysis.

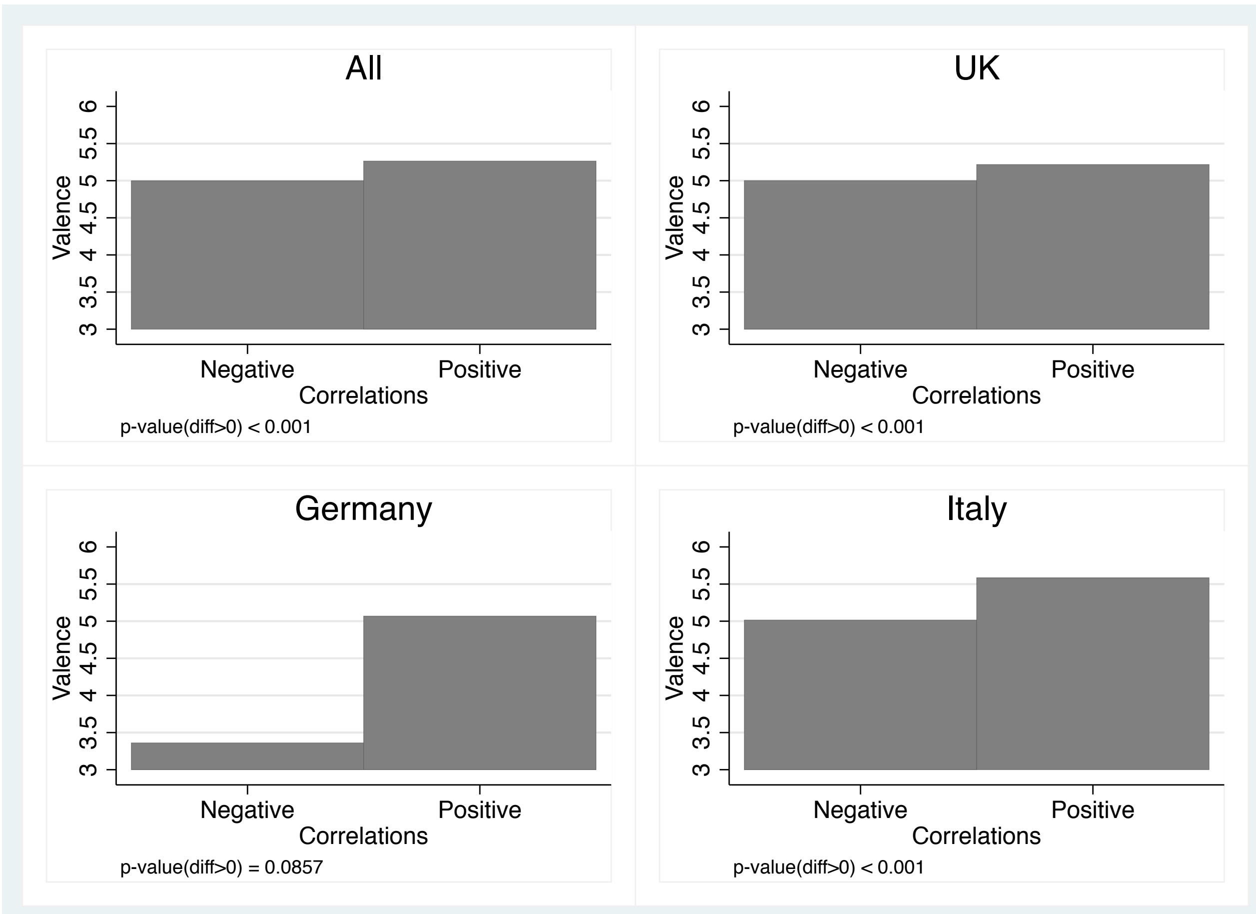


Figure A.3: **Average Valence and Correlations with Life Satisfaction: All Countries Available.** We selected the words in our dataset for which the level of correlation between valence and life satisfaction (from the Eurobaromter survey-based measure) is significant at the 0.05% level and then calculated the averages of the valence across the words correlating positively and negatively for the UK, Germany and Italy. The bars in the figure represent the average valence of words that correlate positively and negatively. By looking at the bars it is possible to see that the average valence among words that correlate positively with life satisfaction is higher than the average valence among words that correlate negatively with life satisfaction.

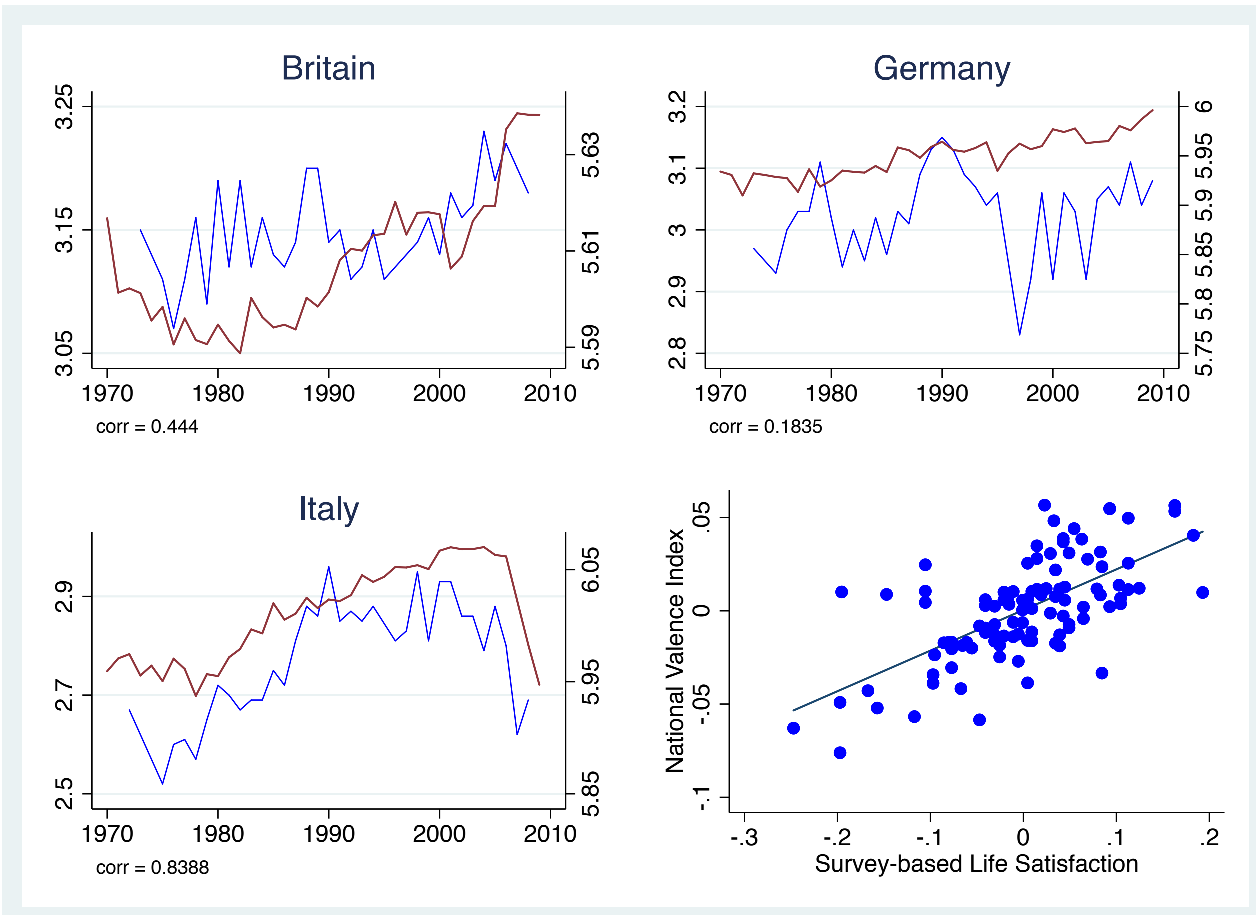


Figure A.4: **The National Valence Index and Aggregate Life Satisfaction.** In the first 3 panels which present time-series data, the National Valence Index is represented in red (values in the left axis) and life satisfaction is represented in blue (values in the right axis). In the last panel, we plotted the National Valence Index against life satisfaction for the same countries and periods; both variables are expressed in the form of residuals after controlling for country fixed-effects.

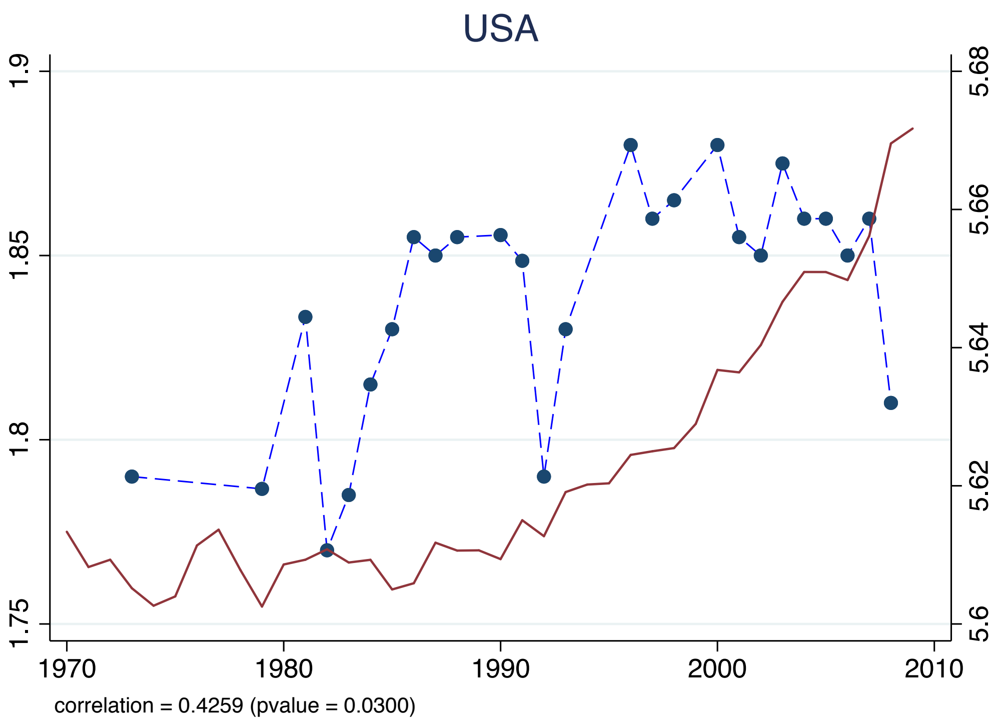


Figure A.5: **The National Valence Index and Aggregate Life Satisfaction in the US.** The National Valence Index is represented in red (values on the left axis) and life satisfaction is represented in blue (values on the right axis). Life Satisfaction data are from the World Database of Happiness (35) and are coded as 1 (= "disatisfied") and 2 (= "satisfied"). They are available only for the years 1973, 1979, 1981-1993, 1996-1998 and 2000-2008.

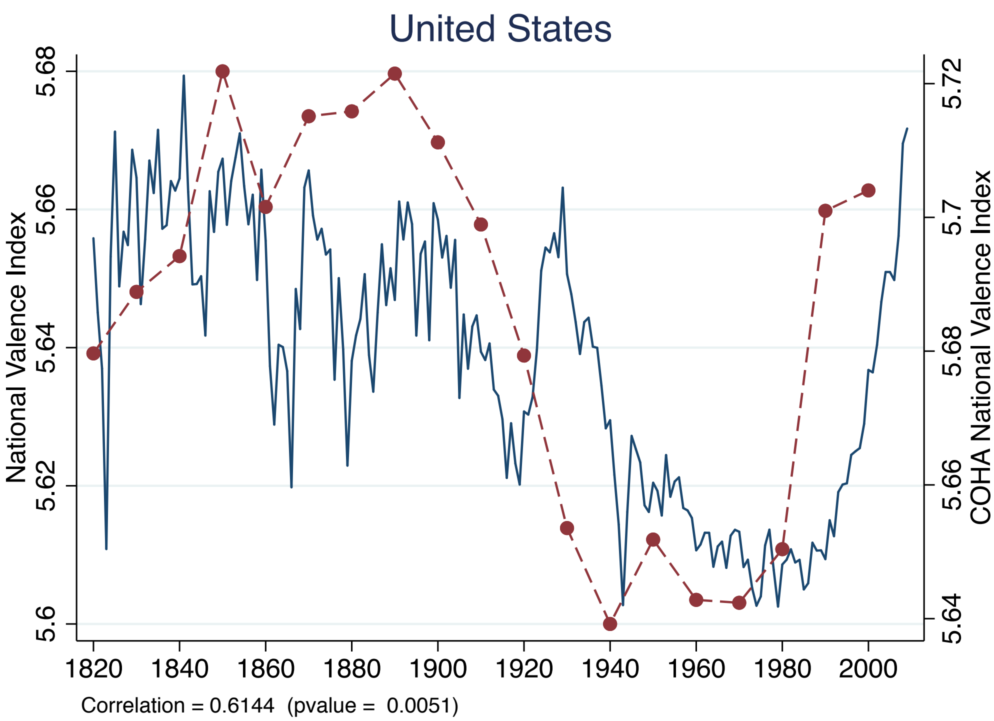


Figure A.6: **The National Valence Index Derived from Two Different Corpora of US Data.** The red line represents the National Valence Index calculated using the COHA Corpora - based on 400 million words of text from 1810 to 1990, by decade, and composed of newspaper and magazine articles. The blue line represents the US National Valence Index derived from the Google Books corpus.

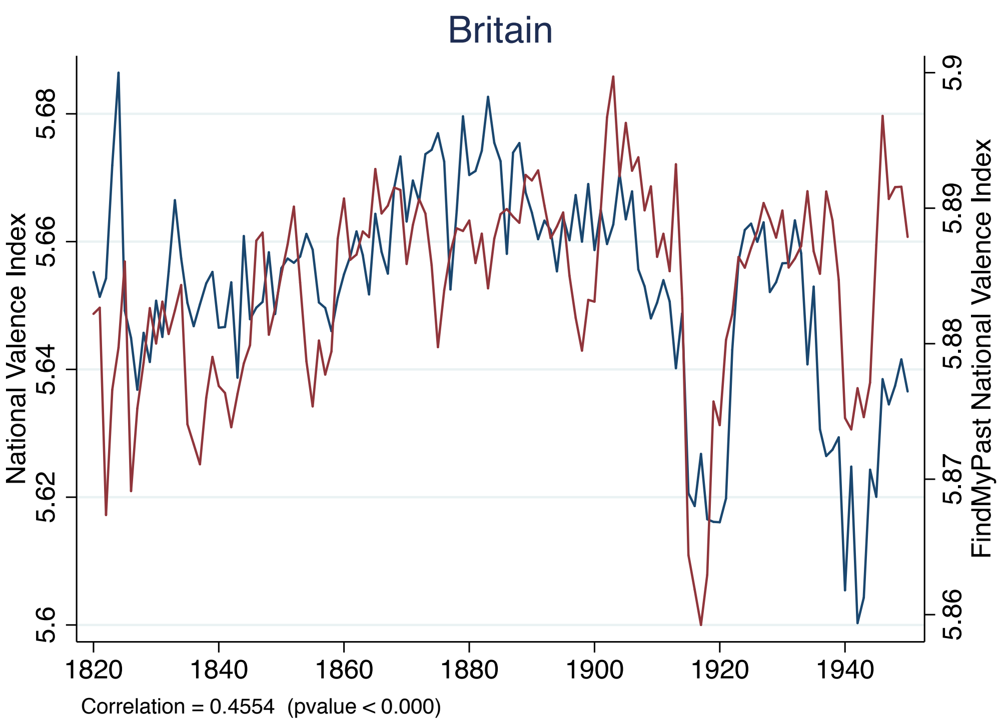
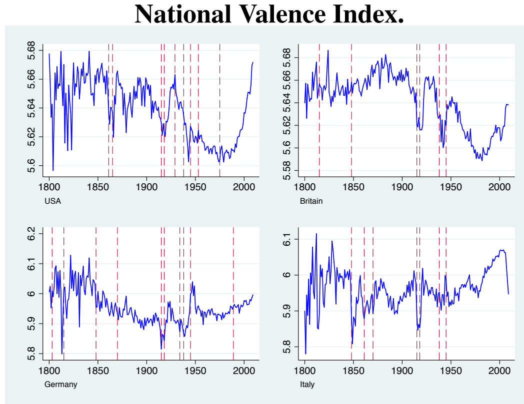
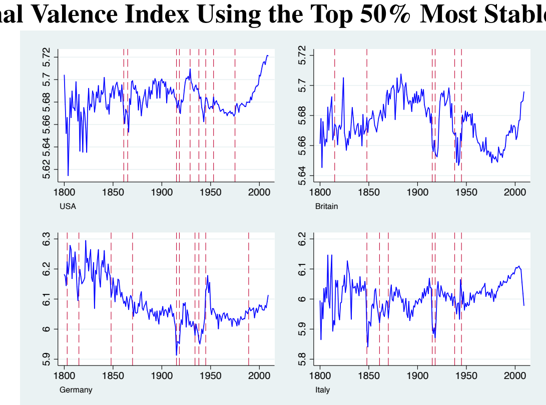


Figure A.7: **The National Valence Index Derived from Two Different Corpora of British Data.** The red line represents the National Valence Index calculated using FindMyPast data - based on 200 British periodicals from 1710-1953. The blue line represents the British National Valence Index derived from the Google Books corpus.

National Valence Index.



National Valence Index Using the Top 50% Most Stable Words.



National Valence Index Using the Top 25% Most Stable Words.

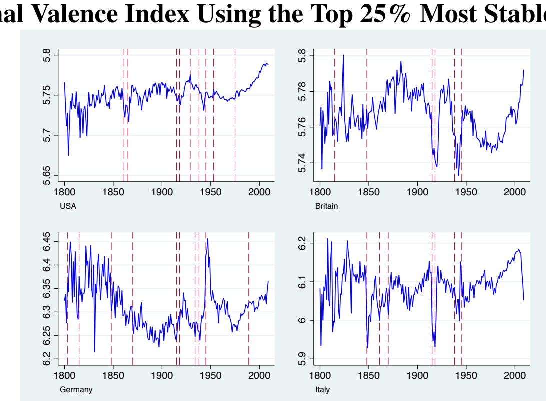


Figure A.8: **A Time-Series Plot Over the Period 1800-2009.**

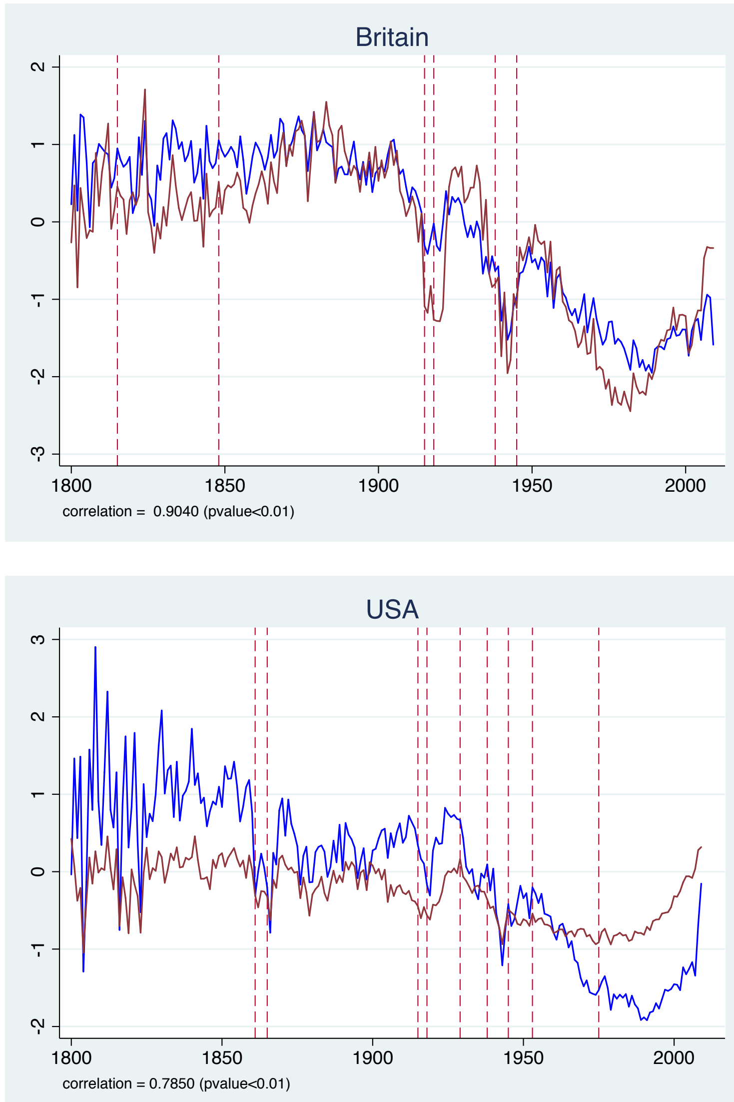


Figure A.9: **The NVI Derived from the AFINN Word Norm vs the ANEW Word Norm over the Period 1800-2009.** The blue line represents the National Valence Index derived from the AFINN word norm and the red line the National Valence Index derived from the ANEW word norm. The National Valence Indices are transformed in standard deviations to ease comparability.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0826***		0.0698***	0.0550**
	(0.0090)		(0.0106)	(0.0130)
Life Expectancy(t-1)		0.0048**	0.0030	0.0016
		(0.0013)	(0.0014)	(0.0013)
Internal Conflict(t-1)				–0.0184**
				(0.0040)
Words Covered	–1.5813	–2.0859	–1.2282	0.4901
	(1.3370)	(2.2393)	(1.3712)	(0.7027)
Democracy	0.0030*	0.0024*	0.0021**	–0.0006
	(0.0010)	(0.0008)	(0.0005)	(0.0006)
Education Inequality	0.0003	0.0008**	0.0004**	0.0001
	(0.0003)	(0.0002)	(0.0001)	(0.0002)
Italy Trend				–0.0009
				(0.0007)
Germany Trend				–0.0007
				(0.0006)
UK Trend				–0.0016**
				(0.0005)
USA Trend				–0.0018*
				(0.0006)
Year FE	Yes	Yes	Yes	No
r2	0.752	0.705	0.774	0.571
N	412	412	412	412

Table A.2: **Historical Determinants of the National Valence Index – all coefficients are visible.** The countries are Germany, Italy, UK and the United States and the period considered is 1820-2009. The regressions are carried out with OLS and either a year fixed-effect (to help deal with spurious correlations over time) or country fixed-effect (to help deal with spurious correlations across countries). Robust standard errors clustered at country levels are given in brackets. Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4	5
	M1	M2	M3	M4	M5
	b/se	b/se	b/se	b/se	b/se
Life Expectancy(t)	0.0046**				
	(0.0013)				
Life Expectancy(t- 1)		0.0048**			
		(0.0013)			
Life Expectancy(t- 3)			0.0044**		
			(0.0008)		
Life Expectancy(t- 5)				0.0027*	
				(0.0010)	
Life Expectancy(t- 10)					0.0049***
					(0.0007)
Democracy(t)	0.0026*	0.0024*	0.0029*	0.0035**	0.0026*
	(0.0011)	(0.0008)	(0.0009)	(0.0010)	(0.0010)
Education Inequality(t)	0.0009**	0.0008**	0.0008**	0.0008**	0.0007*
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Words Covered(t)	–2.0159	–2.0859	–1.9190	–2.2976	–1.8185
	(2.2155)	(2.2393)	(2.2140)	(2.4087)	(2.1879)
Year FE	Yes	Yes	Yes	No	No
r2	0.696	0.705	0.699	0.672	0.698
N	412	412	408	404	394

Table A.3: **Effect of Life Expectancy on the National Valence Index, using Different Time Lags in the Regressors.** The dependent variable is the NVI at time t. The countries included are Germany, Italy, UK and the United States and the period considered is 1820-2009. This table highlights the significance level of different possible lags of Life Expectancy. Robust standard errors are clustered at country levels are given in brackets. Statistical significance is indicated as follows: $^{*}\ p-value < 0.1$, $^{**}\ p-value < 0.05$, $^{***}\ p-value < 0.01$.

	1	2	3	4	5
	M1	M2	M3	M4	M5
	b/se	b/se	b/se	b/se	b/se
(log) GDP(t)	0.0614***				
	(0.0072)				
(log) GDP(t-1)		0.0611***			
		(0.0079)			
(log) GDP(t-3)			0.0659***		
			(0.0081)		
(log) GDP(t-5)				0.0735***	
				(0.0111)	
(log) GDP(t-10)					0.0728***
					(0.0079)
Democracy(t)	0.0025*	0.0026*	0.0028*	0.0029*	0.0027*
	(0.0010)	(0.0010)	(0.0010)	(0.0010)	(0.0010)
Education Inequality(t)	0.0004	0.0004	0.0004	0.0003	0.0002
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)
Words Covered(t)	–2.5082	–2.4601	–2.2927	–2.1053	–2.1659
	(1.4543)	(1.4147)	(1.2709)	(1.0832)	(1.0778)
Year FE	Yes	Yes	Yes	Yes	Yes
r2	0.707	0.707	0.718	0.735	0.728
N	459	459	459	459	459

Table A.4: **Effect of the GDP on the National Valence Index, using Different Time Lags in the Regressors.** The dependent variable is the NVI at time t. The countries included are Germany, Italy, UK and the United States and the period considered is 1820-2009. This table highlights the significance level of different possible lags of GDP. Robust standard errors clustered at country levels are given in brackets. Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4	5
	M1	M2	M3	M4	M5
	b/se	b/se	b/se	b/se	b/se
Internal Conflict(t)	–0.0372				
	(0.0161)				
Internal Conflict(t-1)		–0.0393*			
		(0.0133)			
Internal Conflict(t-3)			–0.0316**		
			(0.0090)		
Internal Conflict(t-5)				–0.0278**	
				(0.0064)	
Internal Conflict(t-10)					–0.0224*
					(0.0072)
Words Covered(t)	0.0380	–0.0161	0.0231	–0.0876	–0.4527
	(1.5854)	(1.5378)	(1.5244)	(1.4500)	(1.3528)
r2	0.008	0.010	0.006	0.005	0.006
N	1227	1223	1215	1207	1187

Table A.5: **The Effect of Internal Conflicts on the National Valence Index, using Different Time Lags in the Regressors.** The dependent variable is the NVI at time t. The countries are Germany, Italy, UK and the United States. This table highlights the significance level of different possible lags of Internal Conflict. Robust standard errors clustered at country levels are given in brackets. Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

Language	Most stable words	Least stable words
UK English	hugger, can, would, will, may	daybreak, daresay, daisy, banter, irrigate
USA English	can, will, would, shall, hundred	stairs, staircase, stainless, sportsman, holly
German	frühling, räuber, liebe, gesundheit, gott	schlüssel, schnee, vogel, sauer, heu
Italian	regina, santo, colore, lago, ferro	saggio, salice, salutare, ratto, gelosia

Table A.6: **The Most Stable and Least Stable Words for each Language, for Words that Existed in 1800.**

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0669**		0.0507**	0.0488**
	(0.0138)		(0.0152)	(0.0087)
Life Expectancy(t-1)		0.0048***	0.0032*	0.0024
		(0.0007)	(0.0010)	(0.0016)
Internal Conflict(t-1)				–0.0134***
				(0.0011)
Words Covered	0.2436	0.3088	0.2814	0.9849
	(0.6590)	(0.6382)	(0.6851)	(0.4898)
Democracy	0.0024**	0.0017	0.0013**	–0.0008
	(0.0004)	(0.0008)	(0.0002)	(0.0006)
Education Inequality	0.0001	0.0005**	0.0002	0.0001
	(0.0002)	(0.0001)	(0.0001)	(0.0002)
Italy Trend				–0.0011
				(0.0007)
Germany Trend				–0.0009
				(0.0006)
UK Trend				–0.0015*
				(0.0005)
USA Trend				–0.0016*
				(0.0006)
Year FE	Yes	Yes	Yes	No
r2	0.691	0.673	0.725	0.464
N	412	412	412	412

Table A.7: **Historical Determinants of the National Valence Index (valence computed using the 50% most stable words identified using the maximum difference in cosine distances), from 1820 to 2009**. Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0514***		0.0375*	0.0492***
	(0.0084)		(0.0119)	(0.0078)
Life Expectancy(t-1)		0.0041**	0.0030*	0.0026
		(0.0010)	(0.0011)	(0.0019)
Internal Conflict(t-1)				–0.0102**
				(0.0021)
Words Covered	0.9801	1.0423	0.6331	1.2139
	(0.7372)	(0.9230)	(0.5019)	(0.6098)
Democracy	0.0015*	0.0008	0.0005	–0.0009
	(0.0005)	(0.0008)	(0.0004)	(0.0006)
Education Inequality	0.0002	0.0005**	0.0003	0.0003
	(0.0001)	(0.0001)	(0.0001)	(0.0002)
Italy Trend				–0.0012
				(0.0007)
Germany Trend				–0.0011
				(0.0007)
UK Trend				–0.0013*
				(0.0005)
USA Trend				–0.0015*
				(0.0006)
Year FE	Yes	Yes	Yes	No
r2	0.671	0.673	0.703	0.408
N	412	412	412	412

Table A.8: **Historical Determinants of the National Valence Index (valence computed using the 25% most stable words identified using the maximum difference in cosine distances), from 1820 to 2009.** Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0921**		0.0708**	0.0543**
	(0.0201)		(0.0191)	(0.0135)
Life Expectancy(t-1)		0.0064***	0.0043**	0.0023
		(0.0008)	(0.0009)	(0.0019)
Internal Conflict(t-1)				–0.0145**
				(0.0037)
Words Covered	0.5820	0.6227	0.4637	0.8523
	(0.5977)	(0.6138)	(0.6590)	(0.6286)
Democracy	0.0042***	0.0034**	0.0027***	–0.0003
	(0.0005)	(0.0009)	(0.0002)	(0.0006)
Education Inequality	0.0003	0.0009*	0.0004*	0.0002
	(0.0004)	(0.0003)	(0.0002)	(0.0002)
Italy Trend				–0.0009
				(0.0009)
Germany Trend				–0.0011
				(0.0008)
UK Trend				–0.0016
				(0.0007)
USA Trend				–0.0018
				(0.0008)
Year FE	Yes	Yes	Yes	No
r2	0.739	0.711	0.780	0.605
N	412	412	412	412

Table A.9: **Historical Determinants of the National Valence Index (valence computed using the 25% most stable words identified using the average difference in cosine distances), from 1820 to 2009.** Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0470**		0.0394**	0.0362*
	(0.0101)		(0.0101)	(0.0146)
Life Expectancy(t-1)		0.0027**	0.0015	0.0029*
		(0.0007)	(0.0010)	(0.0010)
Internal Conflict(t-1)				–0.0069
				(0.0042)
Words Covered	1.1891	1.2274	1.2068	0.2085
	(0.6269)	(0.6296)	(0.6328)	(0.3967)
Democracy	0.0018	0.0016	0.0012	0.0007
	(0.0008)	(0.0007)	(0.0010)	(0.0006)
Education Inequality	–0.0006**	–0.0002	–0.0005**	–0.0004
	(0.0001)	(0.0001)	(0.0002)	(0.0003)
Italy Trend				–0.0022*
				(0.0007)
Germany Trend				–0.0020**
				(0.0006)
UK Trend				–0.0020**
				(0.0006)
USA Trend				–0.0022**
				(0.0006)
Year FE	Yes	Yes	Yes	No
r2	0.547	0.526	0.554	0.299
N	412	412	412	412

Table A.10: **Historical Determinants of the National Valence Index (time-locked valences computed using the valence of the 50% most stable words, identified using the maximum difference in cosine distances, based on their co-occurence with the observed word), from 1820 to 2009.** Statistical significance is indicated as follows: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	1820-1950 FindMyPast b/se	1820-1950 Google b/se
GDP (log) t	0.9149***	
	(0.1512)	
GDP (log) t-5		0.6519**
		(0.2654)
WW1	–2.1139***	–1.0180***
	(0.2163)	(0.2439)
WW2	–1.4433***	–1.0039***
	(0.2171)	(0.2570)
Words Covered(t)	146.1456	–139.5449***
	(101.3410)	(34.2593)
r2	0.529	0.486
N	130	130

Table A.11: **Comparing Historical Determinants of the National Valence Indices from 1820 to 2009 in Britain, using Find My Past Data and Google**. The NVI are transformed in standard deviations to ease comparability. Statistical significance is indicated as follows: * $p{-}value < 0.1$, ** $p{-}value < 0.05$, *** $p{-}value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0224**		0.0166*	0.0002
	(0.0062)		(0.0056)	(0.0020)
Life Expectancy(t-1)		0.0017***	0.0012**	0.0002
		(0.0001)	(0.0003)	(0.0003)
Internal Conflict(t-1)				0.0020
				(0.0036)
Words Covered	0.0756	0.0128	0.0574	0.1126***
	(0.1002)	(0.1643)	(0.1109)	(0.0177)
Democracy	0.0011**	0.0009**	0.0007**	–0.0001***
	(0.0002)	(0.0002)	(0.0001)	(0.0000)
Education Inequality	0.0000	0.0002	0.0001	–0.0001
	(0.0002)	(0.0001)	(0.0001)	(0.0000)
Italy Trend				0.0001
				(0.0001)
Germany Trend				0.0001
				(0.0001)
UK Trend				–0.0003**
				(0.0001)
USA Trend				–0.0003***
				(0.0000)
Year FE	Yes	Yes	Yes	No
r2	0.668	0.653	0.724	0.872
N	412	412	412	412

Table A.12: **Historical Determinants of the SenticNet National Pleasantness Index from 1820 to 2009.** Statistical significance is indicated as follows: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3	4
	Year FE	Year FE	Year FE	CS Trends
	b/se	b/se	b/se	b/se
(log) GDP(t-5)	0.0122*		0.0089	0.0034*
	(0.0048)		(0.0048)	(0.0013)
Life Expectancy(t-1)		0.0009**	0.0007*	0.0002
		(0.0002)	(0.0003)	(0.0002)
Internal Conflict(t-1)				0.0022
				(0.0032)
Words Covered	0.1181	0.0855	0.1089	0.0927***
	(0.0995)	(0.1090)	(0.0992)	(0.0105)
Democracy	0.0007**	0.0005**	0.0005**	–0.0000
	(0.0002)	(0.0001)	(0.0001)	(0.0000)
Education Inequality	–0.0000	0.0000	–0.0000	–0.0000
	(0.0001)	(0.0001)	(0.0001)	(0.0000)
Italy Trend				–0.0000
				(0.0001)
Germany Trend				0.0001
				(0.0000)
UK Trend				–0.0002*
				(0.0001)
USA Trend				–0.0003***
				(0.0000)
Year FE	Yes	Yes	Yes	No
r2	0.537	0.533	0.577	0.762
N	412	412	412	412

Table A.13: **Historical Determinants of the SenticNet National Polarity Index from 1820 to 2009.** Statistical significance is indicated as follows: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.