Historical Analysis of National Subjective Well-being using Millions of Digitized Books*

Thomas Hills, Eugenio Proto and Daniel Sgroi†

November 2017

Abstract

We develop a new way to measure national subjective well-being across the very long run where traditional survey data on well-being is not available. Our method is based on quantitative analysis of digitized text from millions of books published over the past 200 years, long before the widespread availability of consistent survey data. The method uses psychological valence norms for thousands of words in different languages to compute the relative proportion of positive and negative language for four different nations (the USA, UK, Germany and Italy). We validate our measure against existing survey data from the 1970s onwards (when such data became available) showing that our measure is highly correlated with surveyed life satisfaction. We also validate our measure against historical trends in longevity and GDP (showing a positive relationship) and conflict (showing a negative relationship). Our measure allows a first look at changes in subjective well-being over the past two centuries, for instance highlighting the dramatic fall in well-being during the two World Wars and rise in relation to longevity.

JEL codes: N3, N4, O1, D6

Keywords: historical subjective well-being, language, big data, GDP, conflict

*The authors thank several colleagues for discussions on this and related research, especially Sean Allen, Sasha Becker, Steve Broadberry, Nick Crafts, Ray Duch, Paul Frijters, Richard Layard, Andrew Oswald, Luigi Pascali, Giovanni Ricco, David Ronayne, Jeremy Smith and Thijs Van Rens. We thank CAGE (The ESRC Centre for Competitive Advantage in the Global Economy at the University of Warwick) for funding and Tomas Engelthaler and Li Ying for research assistance.

†Hills: Department of Psychology, University of Warwick, t.t.hills@warwick.ac.uk. Proto: Department of Economics, University of Warwick, CAGE and IZA, e.proto@warwick.ac.uk. Sgroi: Department of Economics, University of Warwick, CAGE and Nuffield College, University of Oxford, daniel.sgroi@warwick.ac.uk.

1 Introduction

Subjective well-being is a time-honored preoccupation for economists. To give just a few examples, consider Di Tella, MacCulloch, and Oswald (2001); Deaton (2008); Stevenson and Wolfers (2008); Benjamin, Kimball, Heffetz, and Rees-Jones (2012) and Proto and Rustichini (2013). 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 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 well-being suffers 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 (Bolt and van Zanden (2014)) and for some nations we have centuries of additional data (Broadberry, Campbell, Klein, Overton, and Van Leeuwen (2012)). By comparison, consistent measures of subjective well-being have only been collected since the 1970s.

Here we present and validate a reliable historical measure of national subjective well-being, 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 drawn from the Google Books corpus, which is a collection of word frequency data for over 8 million books (see Lin, Michel, Aiden, Orwant, Brockman, and Petrov (2012)). Overall this data provides a digitized record of more than 6% of all books ever physically published (Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. (2011)). We use the words published in these books to compute the subjective well-being 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.

Our method relies on making inferences about public mood (i.e., sentiment) from large corpora of written text. This method has been used successfully in numerous contexts, including recovering large-scale opinions about political candidates (Connor, Balasubramanyan, Routledge, and Smith (2010)), predicting stock market trends (Bollen, Mao, and Zeng (2011)), understanding diurnal and seasonal mood variation (Golder and Macy (2011)), detecting the social spread of collective emotions (Chmiel, Sienkiewicz, Thelwall, Paltoglou, Buckley, Kappas, and Hołyst (2011)), and understanding the impact of events with the potential for large-scale societal effects such as celebrity deaths, earthquakes, and economic bailouts (Dodds, Harris, Kloumann, Bliss, and Danforth (2011); Thelwall, Buckley, and Paltoglou (2011)). Our approach is widely used (for instance, see (Dodds, Harris, Kloumann, Bliss, and Danforth, 2011, e.g.)) and directly supported by a study of 17 million blog posts, where Nguyen, Phung, Adams, Tran, and Venkatesh (2010) 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 well-being of the user, and those with negative valence are taken to have an equivalent negative connotation.[1]

Validation of our new measure is obtained through a series of comparisons with trusted historical measures. We show that our measure is entirely consistent with the Eurobarometer survey measure of aggregate life satisfaction: our measure of national subjective well-being is a statistically significant predictor of the Eurobarometer measure for the entire period during which both measures are available (from 1973 to 2009). We also examine the nature of the words in our measure, showing that the average valence of all words that correlate positively with the Eurobarometer life satisfaction measure is significantly higher than the average valence of words that correlate negatively. We examine the relationship between our measure and three long-run socio-economic series that also go back hundreds of years: GDP (provided by the Maddison Project), life expectancy and internal

[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.

conflict (both provided by the OECD). Our measure exhibits a positive relationship with GDP and life expectancy and a negative relationship with conflict. Finally we examine the data going back 200 years and pinpoint important historical events arguing that our measure picks up the effect of these events such as sharps falls during World Wars and other periods of intense conflict.

In the next section we describe the process of constructing the index. We then begin our series of validation exercises in section 3 with a comparison with the Eurobarometer life satisfaction measure. In section 4 we examine the relationship between our measure and important socio-economic data where historical data is available for similarly long periods of history: GDP, life expectancy and conflict (as measured by the OECD). We follow with some concluding remarks in section 5. The appendix includes a discussion of the role of the publishing industry, and analyses of lag structures and possible stochastic trends in the data.

2 Building the Measure

In order to construct our measure 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 "Affective Norms for English Words" (Bradley and Lang (1999)) 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.[2]

For English we used the affective rating norms which has become an important resource within psychology (Warriner, Kuperman, and Brysbaert (2013)). These norms are a database of nearly 14 thousand English words, all rated on a 1 to 9 valence scale. The 14 thousand words in the database contain a subset of the 1034 ANEW words. For German

[2]The Google Book corpora also includes books in Russian, Chinese and Hebrew but to the best of our knowledge valence has not been calculated for words in these languages. 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.

we used the affective norms for German sentiment terms (Schmidtke, Schröder, Jacobs, and Conrad (2014)). 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 (Montefinese, Ambrosini, Fairfield, and Mammarella (2014)), which contains 1121 Italian words. As with the English words, the ratings were collected on a 1 to 9 scale.

In the Appendix (in figure A.1) we present a sample of the words covered in all the languages we are considering from the start of the alphabet. We can also consider some examples at the high end such as happiness (8.53), enjoyment (8.37), vacation (8.53), joy (8.21), relaxing (8.19) and peaceful (8.00), lovemaking 7.95, celebrate 7.84. At the low end we have murder (1.48), abuse (1.53), die (1.67), disease (1.68), starvation (1.72), stress (1.79), unhappy (1.84) and hateful (1.9). In the middle we have words like neutral (5.5), converse (5.37), eight (5.37), century (5.36) and machinery (4.65).

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

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

In this expression $v_{j,i}$ 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 and represents the percentage of words covered by the weighted valence score.

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 where sufficient data is available (see Greenfield (2013) and Michel, Shen, Aiden, Veres, Gray, Pickett, Hoiberg, Clancy, Norvig, Orwant, et al. (2011)). 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, although this percentage is about 1%, which is consistent with the number of words covered in Italian and German being 10 times smaller compared to both US and British English. Despite this stability (see A.2), in all our analyses we control for words covered.

3 Validation against the Eurobaromter Life Satisfaction Measure

To validate the NVI, we compare it with existing survey-based measures of subjective well-being. 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 in the Eurobarmeter survey 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).[3]

Figure 1 shows the positive relationship between the NVI and aggregate life satisfaction derived from Eurobarometer data for the corresponding country. The first three panels show the NVI (in red with values on the left axis) and the corresponding Eurobarometer life satisfaction measure (in blue with values on the right axis) for Britain, Germany and Italy from 1973 to 2009 (the period when both measures exist). In order to get a feel for the overall strength of the correlation between the NVI and Eurobarometer data the final panel plots the NVI against life satisfaction for the same countries and time periods: the data are presented in the form of residuals after controlling for country fixed-effects and there is a positive correlation of 0.7079 (which provides the slope of the line-of-best-fit).

The analysis presented in Table 1 shows that the positive relationship is robust to the

[3]USA data is not included; to the best of our knowledge there are no comparable national-level data on life satisfaction in USA from the 1970s.

introduction of GDP, perhaps the most plausible omitted variable, and 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, or country-specific tends (column 2), to control for the possibility that the trends in NVI and life satisfaction generate a spurious correlation. In all cases, the coefficient on the NVI is positive and significant.

As an additional form of 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. To explore this idea further we first 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. The results are presented in figure 2 where we see that 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.

4 Further Validation and Historical Analysis

In figure 3 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.[4] The red vertical lines represent key political events in each country as indicated in the figure caption and, in combination with the NVI, strongly support a contemporary historical understanding of

[4] After 2000 there was a change in the book sampling method used by Google as described in Greenfield (2013). This can be observed in figure A.2 in the Appendix, 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.

these events' impacts on subjective well-being. 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 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. Moreover, we also analyze the effect of internal conflicts – which is reasonable to think should be detrimental to a country's welfare. Following convention, we use 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.[5] 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 (van Zanden, Baten, Mira dErcole, Rijpma, Smith, and Timmer (2014)). Other variables we use as controls are educational inequality (measured as a GINI index, which we use 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 (van Zanden, Baten, Mira dErcole, Rijpma, Smith, and Timmer (2014)). The data is further summarized in Table A.1 in the Appendix.

The NVI is likely to be affected by the market for literature and, more generally, by the evolution of literature and language (see section A in the Appendix 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

[5]The results of the analysis does not qualitatively change if we use the Maddison dataset instead of the Penn dataset also in the first analysis.

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. Model 2 assumes 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 4.5% of extra per capita annual GDP growth. 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 Appendix, section A 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.[6] The sign of 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.[7]

5 Concluding Remarks

Using conventional regression analysis and non-parametric methods we show that the NVI is highly consistent with existing well-being 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 also further validate our measure by showing a relationship with variables that are known to have a relationship with well-being, such as conflict, life expectancy and GDP.

The NVI makes visible a number of interesting patterns. For example, there is a rise in subjective well-being in Italy and Germany since the 1900s matched by a comparative decline in the UK and USA. Across all four nations, possibly excepting Germany, a steady rise in subjective well-being is visible since the 1970s. Internal and external conflicts represent dramatic shocks to subjective well-being, 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.

Both the market for books, and language itself, have evolved considerably over the period we consider e.g. Hills and Adelman (2015). We nevertheless 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 the arrival of urbanization, huge cultural

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

[7]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. Details are presented in part B of the Appendix.

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 (Jerven (2012)). 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 well-being and presents an initial foray into quantifying psychological history.

References

Alberge, J. (2015): "Publishers bypass literary agents to discover bestseller talent," Discussion paper, The Observer, February 22.

Benjamin, D. J., M. S. Kimball, O. Heffetz, and A. Rees-Jones (2012): "What do you think would make you happier? What do you think you would choose?," *The American economic review*, 102(5), 2083.

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

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

Bolt, J., and J. L. van Zanden (2014): "The Maddison Project: collaborative research on historical national accounts.," *The Economic History Review*, 67(3), 627–651.

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

Broadberry, S., B. Campbell, A. Klein, M. Overton, and B. Van Leeuwen (2012): "British Economic Growth, 1270-1870: an output-based approach," Discussion paper, School of Economics Discussion Papers.

Chmiel, A., J. Sienkiewicz, M. Thelwall, G. Paltoglou, K. Buckley, A. Kappas, and J. A. Hołyst (2011): "Collective emotions online and their influence on community life," *PloS one*, 6(7), e22207.

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

Deaton, A. (2008): "Income, Health, and Well-being Around the World: Evidence from the Gallup World Poll," *Journal of Economic perspectives*, 22(2), 53–72.

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

Dodds, P. S., K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth (2011): "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter," *PloS one*, 6(12), e26752.

Golder, S. A., and M. W. Macy (2011): "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures," *Science*, 333(6051), 1878–1881.

Greenfield, P. M. (2013): "The changing psychology of culture from 1800 through 2000," *Psychological science*, 24(9), 1722–1731.

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

Jerven, M. (2012): "An Unlevel Playing Field: National Income Estimates and Reciprocal Comparison in Global Economic History," *Journal of Global History*, 7, 107–128.

Lin, Y., J.-B. Michel, E. L. Aiden, J. Orwant, W. Brockman, and S. Petrov (2012): "Syntactic annotations for the google books ngram corpus," in *Proceedings of the ACL 2012 system demonstrations*, pp. 169–174. Association for Computational Linguistics.

Michel, J.-B., Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, et al. (2011): "Quantitative analysis of culture using millions of digitized books," *Science*, 331(6014), 176–182.

Montefinese, M., E. Ambrosini, B. Fairfield, and N. Mammarella (2014): "The adaptation of the Affective Norms for English Words (ANEW) for Italian," *Behavior research methods*, 46(3), 887–903.

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

Proto, E., and A. Rustichini (2013): "A reassessment of the relationship between GDP and life satisfaction," *PloS one*, 8(11), e79358.

Schmidtke, D. S., T. Schröder, A. M. Jacobs, and M. Conrad (2014): "ANGST: Affective norms for German sentiment terms, derived from the affective norms for English words," *Behavior research methods*, 46(4), 1108–1118.

Stevenson, B., and J. Wolfers (2008): "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*, pp. 1–87.

Thelwall, M., K. Buckley, and G. Paltoglou (2011): "Sentiment in Twitter events," *Journal of the American Society for Information Science and Technology*, 62(2), 406–418.

van Zanden, J. L., J. Baten, M. Mira dErcole, A. Rijpma, C. Smith, and M. Timmer (2014): "How was life? Global well-being since 1820," .

Warriner, A. B., V. Kuperman, and M. Brysbaert (2013): "Norms of valence, arousal, and dominance for 13,915 English lemmas," *Behavior research methods*, 45(4), 1191–1207.

Figures and Tables

Figure 1: **The National Valence Index and Aggregate Life Satisfaction.** In the first 3 panels which present time-series data, the National Valence Index (our measure of subjective well-being derived from digitized text) is represented in red (values on the left axis) and life satisfaction (taken from the Eurobarometer survey-based measure) is represented in blue (values on the right axis). In the last panel, we plotted the National Valence Index against life satisfaction for the same countries and periods: 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. The correlation between the National Valence Index text-based measure of subjective well-being and the survey-based Eurobarometer measure of life satisfaction is positive with a value of 0.7079 which is indicated in the final panel as the slope of the line-of-best-fit.

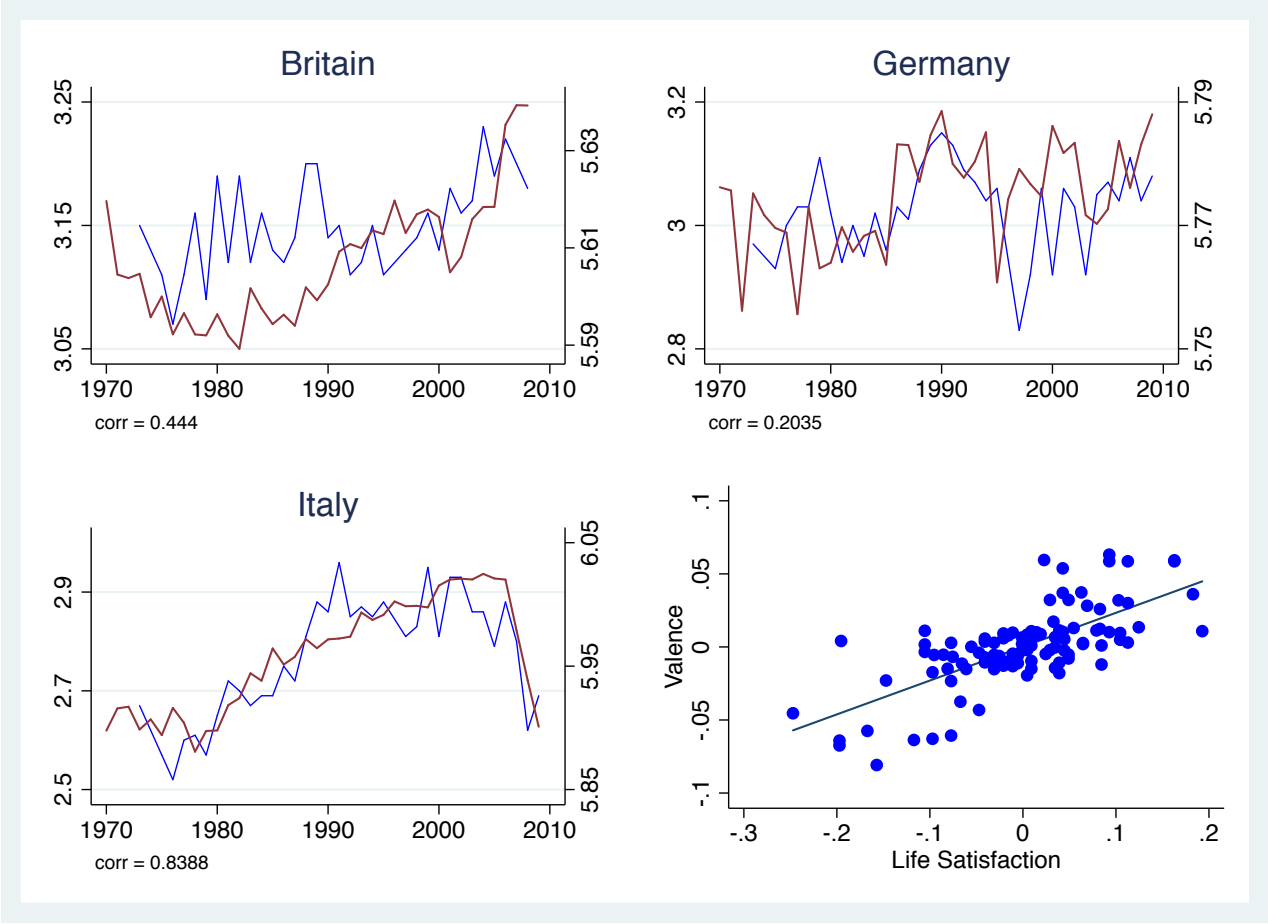

Figure 2: **Average Valence and Correlations with Life Satisfaction: All Countries Available.** We selected the words in our dataset for which the level of correlation between valence and life satisfaction (from the Eurobaromter survey-based measure) is significant at the 0.05% level and then calculated 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.

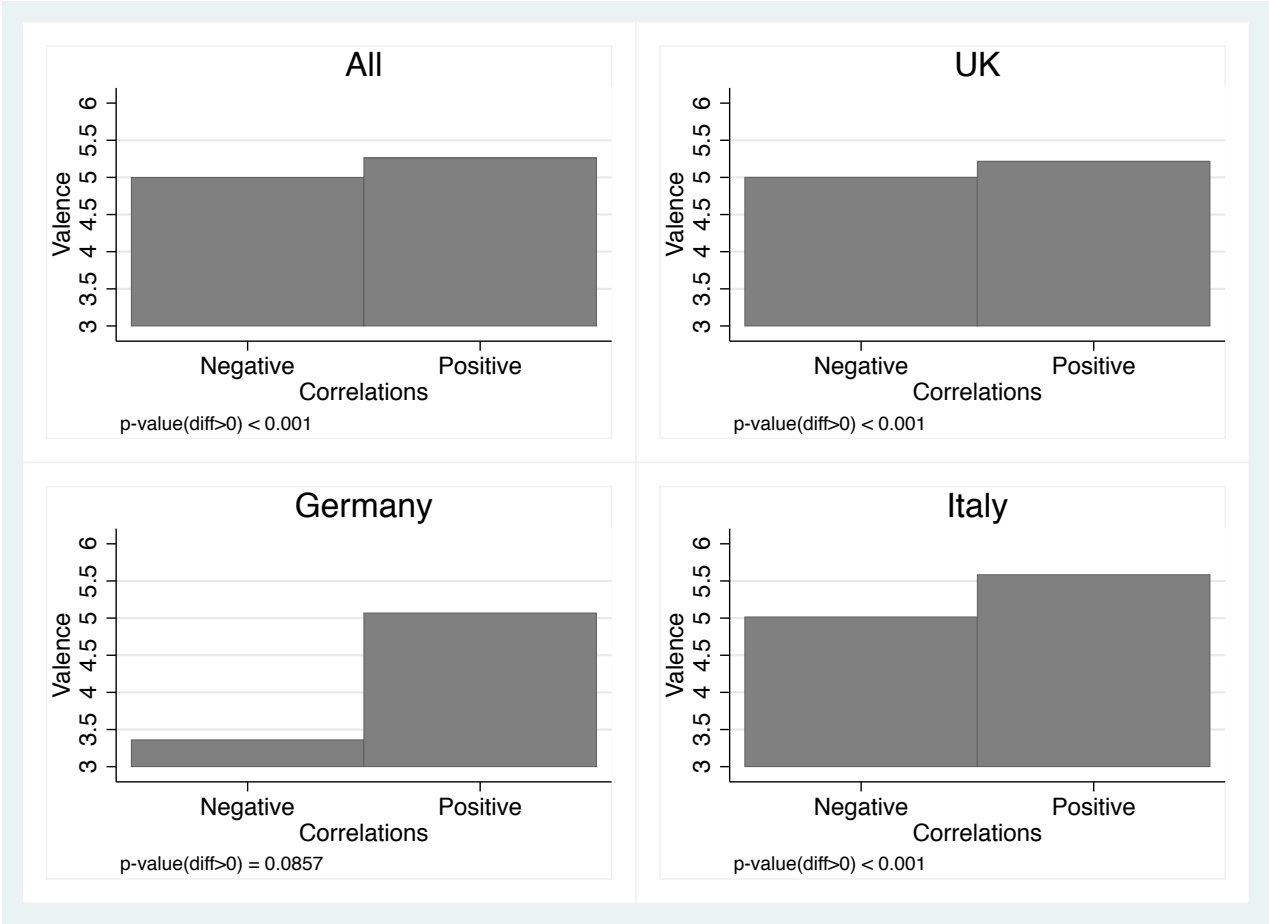

Figure 3: **A Time-Series Plot of the National Valence Index Over the Period 1820-2009.** The figure shows the National Valence Index (our measure of subjective well-being derived from digitized text) plotted from 1820 to 2009. Various important events have been highlighted in shaded red (for periods 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).

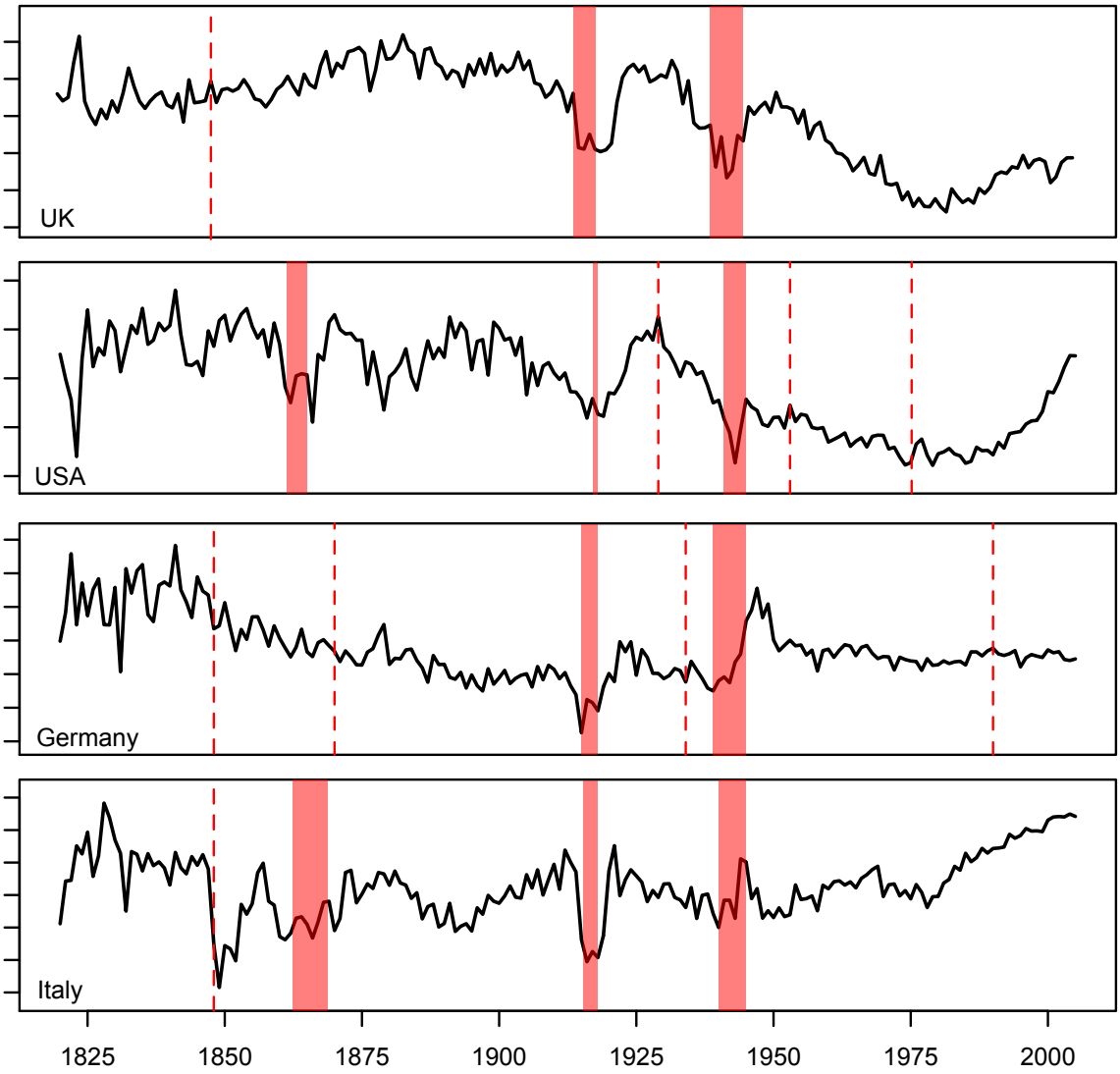

Table 1: **The National Valence Index Predicts Aggregate Life Satisfaction.** The table indicates that the National Valence Index (our measure of subjective well-being derived from digitized text) is a statistically significant predictor in an OLS estimate of aggregate life satisfaction. The dependent 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
	Year FE	CS trends
	b/se	b/se
National Valence Index	2.2536***	1.3349**
	(0.1959)	(0.1608)
GDP	Yes	Yes
Country Specific Trend	No	Yes
Year FE	Yes	No
r2	0.765	0.580
N	104	104

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 well-being). 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 A.2 of the Appendix. Statistical significance is indicated as follows: * $p-value < 0.1$, ** $p-value < 0.05$, *** $p-value < 0.01$.

	1	2
	Year FE	CS Trends
	b/se	b/se
(log) GDP(t-3)	0.0821**	0.0517*
	(0.0174)	(0.0213)
Life Expectancy(t)	0.0036**	0.0016
	(0.0008)	(0.0014)
Internal Conflict(t)		–0.0190**
		(0.0049)
World Covered(t)	Yes	Yes
Democracy(t)	Yes	Yes
Education Inequality(t)	Yes	Yes
Year FE	Yes	No
Country-Specific Trends	No	Yes
r2	0.736	0.494
N	412	412

Appendix

A The Publishing Industry: Market Forces and Lags

In this appendix 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.

A.1 The Publishing Industry

Unless we have reason to suspect some behavioral forces or market failure, we 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 a handful, for instance Alberge (2015) 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 (of 0.7079) which is evident in the final panel of Figure 1 also indicates that publishers match books typified by predominantly high

[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.

valence words ("happy books") to "happy people" and books typified by predominantly low valence words ("sad books") to "sad people." In this appendix we note some quotes from publishers and authors concerning their rationale for rejecting books submitted for publication. The aim is to provide some support for the importance of the demand-side of the book market to publishers deciding which books to select for publication.

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 later become successful. What is helpful for us is that the bias works in our favour: 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 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 Bernard (1990) 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."

A.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 three different models determining the channels through which a country's subjective well-being is factored into the different written languages based on a lag of $t - \tau$ years:

- the books published in the market reflect current subjective well-being (publishers selecting books that match the current mood of the population). In this case $\tau = 0$.
- as before, but the publisher's decision to publish is taken on the basis of subjective well-being one year before, i.e there exists a publishing lag of one year. In this case $\tau = 1$.
- a book published at time t reflects subjective well-being of the population three years prior to publication, i.e. there exists a publishing lag of three years. In this case we assume $\tau = 3$.

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. 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.

B Stochastic Trends

In this appendix 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).

B.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 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. 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 cannot be a source of confounding 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 $Z(t) = 0.0011$).[9] 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.

B.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

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

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 rejected).[10]

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

C Additional Figures and Tables

This section includes additional figures and tables that are referenced in the main text and in the Appendix.

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.

Variable	Mean	Std. Dev.	Min.	Max.	N
National Valence Index	5.746	0.12	5.589	6.042	759
Life Satisfaction	2.98	0.181	2.52	3.23	104
per capita GDP (Maddison)	7562.123	6641.594	1076.852	31357	635
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	0.099	0.299	0	1	759
Democracy	5.654	5.89	-9	10	625
Education Inequality	31.526	22.722	6.111	98.935	504
Words Covered	0.069	0.06	0.009	0.191	759

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 share of words covered in the text analysis over the total, the blue line represents the total number of words, for all countries considered in the analysis.

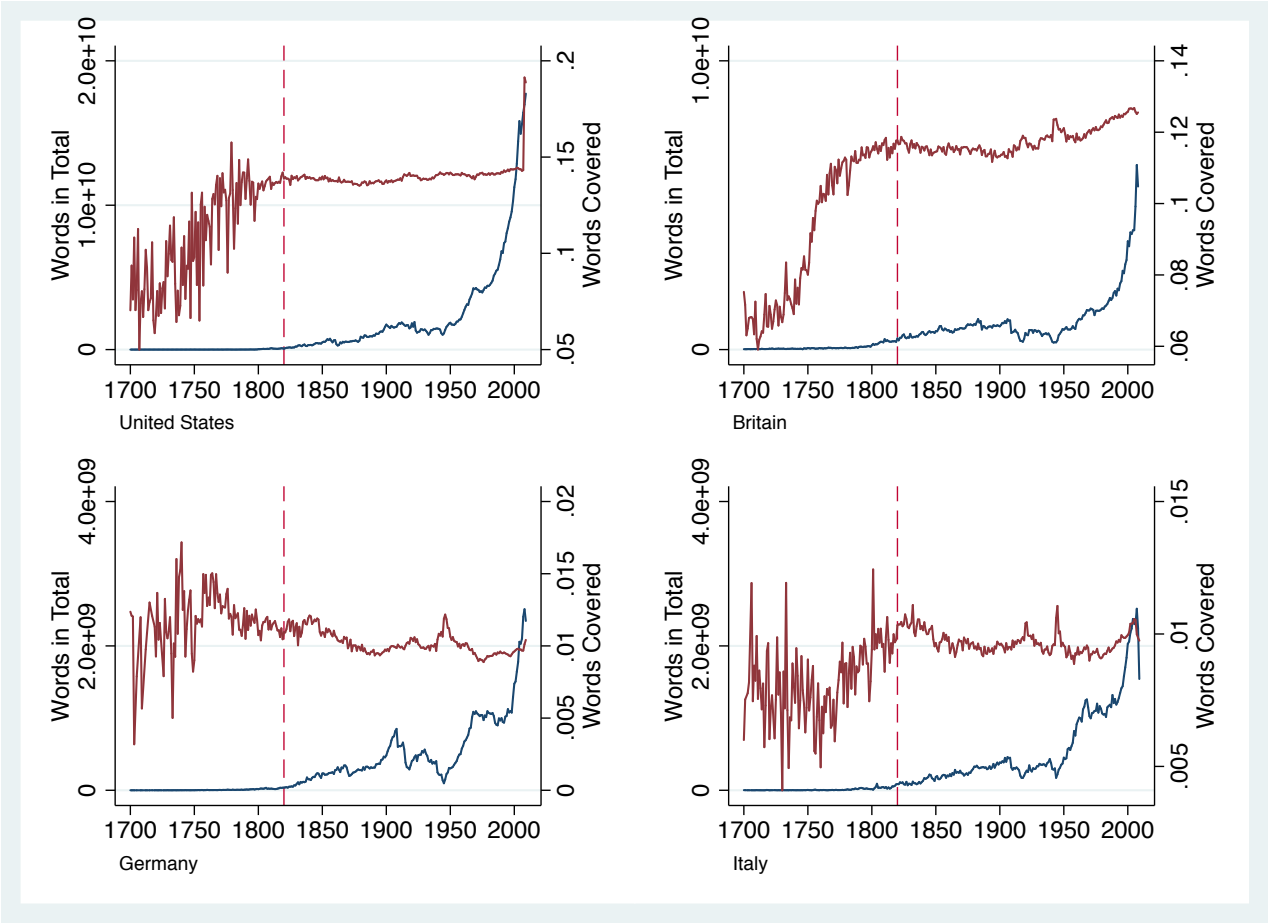

Table A.2: **Historical Determinants of the National Valence Index.** The countries are Germany, Italy, UK and the United States and the period considered is 1820-2009. The regressions are carried out with OLS and either a year fixed-effect (to help deal with spurious correlations over time) or country fixed-effect (to help deal with spurious correlations across countries). This table repeats the analysis in Table 2 but highlights the fact that it is not the inclusion of internal conflict that eliminates the significance of life expectancy but rather the addition of country-specific trends. 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
	Year FE	CS Trend	CS Trend + Conflicts
	b/se	b/se	b/se
GDP (log)	0.0821**	0.0544*	0.0517*
	(0.0174)	(0.0212)	(0.0213)
Life Expectancy	0.0036**	0.0015	0.0016
	(0.0008)	(0.0014)	(0.0014)
Internal Conflict			–0.0190**
			(0.0049)
Words Covered	0.4058	1.1290*	1.0030
	(0.7868)	(0.3780)	(0.4497)
Democracy	0.0018**	–0.0002	–0.0003
	(0.0005)	(0.0009)	(0.0008)
Education Inequality	0.0004**	0.0000	0.0001
	(0.0001)	(0.0004)	(0.0003)
Italy Trend		–0.0011	–0.0010
		(0.0008)	(0.0008)
Germany Trend		–0.0013	–0.0013
		(0.0008)	(0.0008)
UK Trend		–0.0019*	–0.0017*
		(0.0006)	(0.0006)
USA Trend		–0.0018	–0.0017
		(0.0008)	(0.0008)
Year FE	Yes	No	No
r2	0.736	0.472	0.494
N	412	412	412

Table A.3: **Effect of Life Expectancy on the National Valence Index, using Different Time Lags in the Regressors.** 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
	no lag	1 year lag	3 years lag
	b/se	b/se	b/se
Life Expectancy(t)	0.0060***		
	(0.0008)		
Life Expectancy(t- 1)		0.0059***	
		(0.0007)	
Life Expectancy(t- 3)			0.0044***
			(0.0003)
Democracy(t)	0.0026**	0.0026**	0.0035***
	(0.0007)	(0.0005)	(0.0004)
Education Inequality(t)	0.0009***	0.0008***	0.0009**
	(0.0001)	(0.0001)	(0.0002)
Words Covered(t)	–0.3207	–0.3485	–0.3312
	(1.5429)	(1.6123)	(1.5827)
Year FE	Yes	Yes	Yes
r2	0.645	0.646	0.605
N	412	412	408

Table A.4: **Effect of the GDP on the National Valence Index, using Different Time Lags in the Regressors.** 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
	no lag	1 year lag	3 years lag
	b/se	b/se	b/se
(log) GDP(t)	0.0779***		
	(0.0122)		
(log) GDP(t-1)		0.0843**	
		(0.0153)	
(log) GDP(t-3)			0.0952***
			(0.0160)
Democracy(t)	0.0023	0.0026*	0.0028*
	(0.0011)	(0.0011)	(0.0010)
Education Inequality(t)	0.0002	0.0003	0.0003
	(0.0002)	(0.0002)	(0.0002)
Words Covered(t)	–1.1749	–0.2375	0.0031
	(1.5057)	(1.0520)	(0.9174)
Year FE	Yes	Yes	Yes
r2	0.624	0.621	0.660
N	456	459	459

Table A.5: **The Effect of Internal Conflicts on the National Valence Index, using Different Time Lags in the Regressors.** 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: * $p - value < 0.1$, ** $p - value < 0.05$, *** $p - value < 0.01$.

	1	2	3
	no lag	1 year lag	3 years lag
	b/se	b/se	b/se
Internal Conflict(t)	–0.0285**		
	(0.0068)		
Internal Conflict(t-1)		–0.0255**	
		(0.0066)	
Internal Conflict(t-3)			–0.0208*
			(0.0074)
Words Covered(t)	–0.6451**	–0.6763**	–0.6798**
	(0.1772)	(0.1643)	(0.1636)
r2	0.025	0.025	0.022
N	1227	1223	1215