
Historical analysis of national subjective wellbeing using millions of digitised books: Introducing the HPS Index

Thomas Hills, Eugenio Proto, Daniel Sgroi

With records of subjective wellbeing going back less than half a century, this column asks if we can know the impact of key past events on the happiness of our ancestors. It presents a new historical index that draws on millions of digitised books in the Google Books corpus of words using sentiment analysis. The index – which goes back to the 1776 US Declaration of Independence, 200 years earlier than any other index of happiness – makes it possible to analyse the historical drivers of happiness in France, Germany, Italy, Spain, the UK and the US.

Calls from UK Prime Minister David Cameron, the United Nations' World Happiness Report, the OECD's Better Life Index, along with psychologists and economists, all reflect on the need to develop a better understanding of subjective wellbeing ('happiness'). Though many contemporary economies have tracked crime, education and economic production for the best part of a century, subjective wellbeing only began to become a staple of world economic indicators in the 1970s.

Unlike national income accounting, which initiated the collection of GDP in the 1930s, subjective wellbeing is a rather young indicator. Though there have been successful projects to roll back GDP (e.g. Bolt and van Zanden 2014, Broadberry 2015), attempts to construct historical series for wellbeing have been notably lacking. Without such a series, we are left wondering how wellbeing responds to key historical events, such as expansionary monetary policies, education and longevity.

But if constructing historical series for wellbeing makes sense, how can we extend existing measures when direct survey evidence was only initiated in the 1970s? The key insight in our new research paper (Hills *et al.* 2015) is that language conveys sentiment, and that the growing availability of digitised text provides unprecedented resources to construct a quantitative history of wellbeing based on historical language use.

In particular, the foundation of our work involves combining multiple large collections of texts of natural language going back two centuries with state-of-the-art methods for deriving public mood (i.e. sentiment) from language. The recent digitisation of books, newspapers and other sources of natural language – such as the Google Books Ngram database – represent historically unprecedented amounts of data ('big data') on what people thought and wrote over the past few centuries (see Michel 2011). These databases have already proved fruitful in detecting large-scale changes in language, which in turn correlate with social and demographic change, for instance in Hills and Adelman (2015).

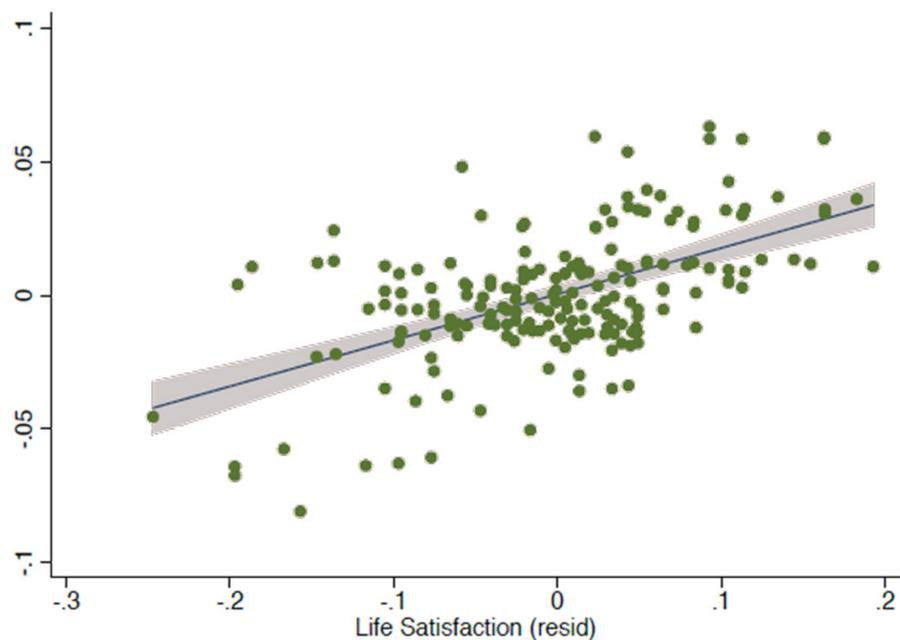
These data offer the capacity to infer public mood using sentiment analysis. Deriving sentiment from large collections of written text represents a growing scientific endeavour. Examples include recovering large-scale opinions about

political candidates, predicting stock market trends, understanding diurnal and seasonal mood variation, detecting the social spread of collective emotions, and understanding the impact of events with the potential for large-scale social impact such as celebrity deaths, earthquakes and economic bailouts (e.g. Pang and Lee 2008). Applying the same methods to historical text we can begin to produce more quantitative accounts of national happiness.

In the approach we take, sentiment measures are based on valence norms for thousands of words. These already exist in the literature and are collected from a large group of individuals who are asked to rate a list of words on how those words make them feel (e.g. Gilbert 2007). In the present case, valence norms based on the affective norms for English words have already been collected for five languages: English, French, Spanish, Italian, and German.

We applied these norms to the Google Books corpus for each of these languages, allowing us to derive a new index for subjective wellbeing going back to 1776, which we tentatively call the HPS index. An initial comparison with subjective wellbeing collected with survey data is shown in Figure 1. The data reflect the residuals after controlling for country fixed effects and clearly show a strong and significant correlation with our measure based on historical language.

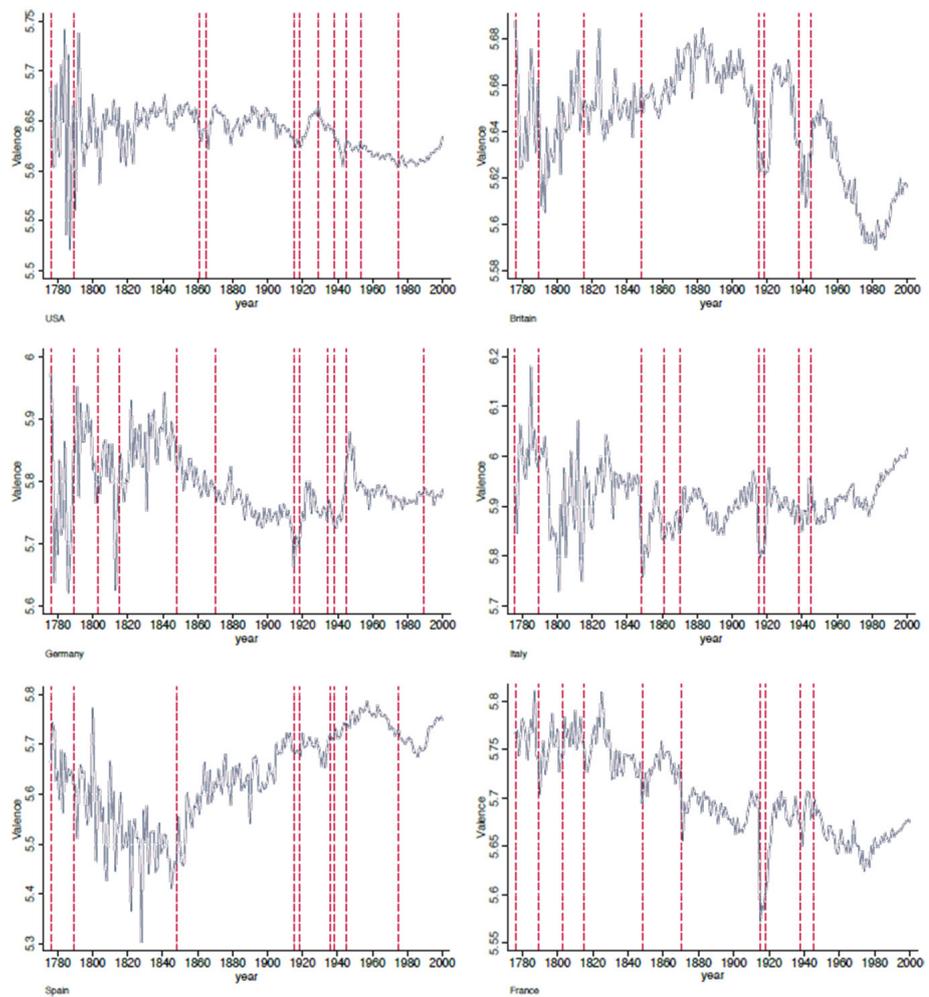
Figure 1. Comparison between survey measures of life satisfaction and residuals (after controlling for country fixed effects) for our measure based on sentiment from historical text.



Note: The grey area represents the 95% confidence interval.

Rolling the text-derived measures of subjective wellbeing back to 1776 reveals a quantitative picture of how public sentiment has changed across the six countries we considered: France, Germany, Italy, Spain, the UK and the US. Though we make clear in our research that we need to exercise caution when examining very long-run trends (as language itself has evolved so much), it is nonetheless clear in Figure 2 that short-term events, such as the exuberance of the 1920s, the Depression era, and World Wars I and II show clear and distinguishable influences on subjective wellbeing.

Figure 2. The average valences over the period 1776-2000.



Note: For all countries the vertical red lines correspond to 1789 (the year of the French Revolution), World War I (1915-18) and World War II (1938-45). In the five European countries, a line is drawn for 1848 (the year of the revolutions). In the US, the vertical lines represent: the Civil War (1861-65), the Wall Street Crash (1929), the end of Korean War (1953) and the fall of Saigon (1975); in the UK, the Napoleonic Wars (1803-15). In Spain, the starting of Civil War (1936); in France, the Napoleonic Wars (1803-15), the end of the Franco-Prussian War (1870); for Germany, the vertical lines represent the Napoleonic Wars (1803-15), the Franco-Prussian War and unification (1870), Hitler's ascendancy to power (1934), the reunification (1990); for Italy, the unification (1861-70).

Why is a quantitative history of wellbeing important?

The fledgling state of wellbeing data has limited our collective ability to understand how wellbeing responds to different historical events. This has in turn limited the use of wellbeing in public policy, health initiatives and financial decision-making. In practice, if subjective wellbeing is to become a key factor in guiding our collective behaviour, then we need accounts of wellbeing on a par with those of GDP.

Using wellbeing as a measure to guide behaviour, however, takes more than the desire to simply improve wellbeing. As noted by Gilbert (2007), people have problems understanding so-called 'affective forecasting' – the ability to understand how one will feel in the future – and with this also comes a limited

capacity to understand how prior events and decisions influenced our past happiness.

To overcome this, especially at the government level, we must develop our capacity to predict how wellbeing responds to both deliberate and unexpected events. Better predicting economic fortunes was the motivation of the national income accounting, which later became GDP, following the Depression in the 1930s. Of course, now numerous decisions are based on GDP, despite a near global acceptance that, in the words of John F Kennedy, "it measures everything in short, except that which makes life worthwhile" (Presidential Library and Museum, North Dakota).

Thus, as with GDP, governments and other agencies recognise the importance of this additional 'emotional accounting' and, by all accounts, they want to understand how better to use it to improve future wellbeing. But to do that we need historical informed accounts of what this means, and our index represents a first attempt.

References

Bolt, J and van Zanden, J L (2014), "The Maddison Project: Collaborative Research on Historical National Accounts", *The Economic History Review* 67(3): 627–651.

Broadberry, S, Campbell, B, Klein, A, Overton, M and van Leeuwen, B (2015), *British Economic Growth, 1270-1870*. Cambridge University Press.

Hills, T T, Proto, E and Sgroi, D (2015), "Historical Analysis of National Subjective Wellbeing using Millions of Digitized Books", IZA Discussion Paper No. 9195.

Michel, J-B, Shen, Y K, Aiden, A P, Veres, A, Gray, M K, Pickett, J P, et al. (2011), "Quantitative Analysis of Culture using Millions of Digitized Books", *Science* 331 (6014): 176–182.

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

Pang, B and Lee, L (2008), "Opinion Mining and Sentiment Analysis", *Foundations and Trends in Information Retrieval* 2(1-2): 1–135.

Warriner, A B, Kuperman, V and Brysbaert, M (2013), "Norms of Valence, Arousal, and 21 Dominance for 13,915 English Lemmas", *Behavior Research Methods* 45 (4), 1191–1207.

Gilbert, D (2007), *Stumbling on Happiness*. Random House.

About CAGE

Established in January 2010, CAGE is a research centre in the Department of Economics at the University of Warwick. Funded by the Economic and Social Research Council (ESRC), CAGE is carrying out a five year programme of innovative research.

The Centre's research programme is focused on how countries succeed in achieving key economic objectives, such as improving living standards, raising productivity and maintaining international competitiveness, which are central to the economic well-being of their citizens.

CAGE's research analyses the reasons for economic outcomes both in developed economies such as the UK and emerging economies such as China and India. The Centre aims to develop a better understanding of how to promote institutions and policies that are conducive to successful economic performance and endeavours to draw lessons for policy-makers from economic history as well as the contemporary world.



This piece first appeared on Voxeu on 18 September 2015
<http://voxeu.org/article/building-historical-index-happiness-using-google-books>

VOX

Research-based policy analysis and commentary from leading economists

© 2016 The University of Warwick

Published by the Centre for Competitive Advantage in the Global Economy
Department of Economics, University of Warwick, Coventry CV4 7AL
www.warwick.ac.uk/cage

Artwork by Mustard, www.mustardhot.com