

## Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books

Thomas T. Hills, Eugenio Proto and Daniel Sgroi\*

### Highlights:

- Governments and international bodies (UN, OECD) routinely collect national-level happiness data to use alongside national income to help measure national-level wellbeing.
- Policy-makers routinely make use of historical national income data to better understand the long-run drivers of wealth, but to-date have no equivalent long-run data for happiness. The typical way to measure happiness is through surveys, but survey measures only started in the 1970s and for some nations only from 2011.
- We use the “big data” made available by the Google book digitization project to give us unprecedented access to billions of words used across several centuries and for several nations. Performing semantic text analysis on these words reveals an implied happiness measure that we can use to build a national happiness index (we call the HPS index) going back more than two centuries.
- The HPS index is highly correlated with survey measures where both coexist (from the 1970s). Using this index, we chart the major drivers of happiness (wars, economic crises, natural disasters, epidemics) and the long-run changes that have affected happiness.
- Aside from big shocks like wars and economic disasters which have a massive short-run effect, the biggest long-run drivers are health measures such as life expectancy and child mortality. Income has no perceivable effect on long-run happiness (confirming the “Easterlin paradox”).

### Full press release:

Happiness may be our ultimate goal as humans, but it has proved difficult to measure and understand despite the best attempts of some of our greatest philosophers and social scientists from Aristotle onwards. Unlike other concepts such as wealth and income, we lack long-run data on how happiness has evolved over time and how the happiness of the population has reacted to long-run changes in society or major shocks (like wars, epidemics, natural disasters and economic crises). Our work seeks to fill this gap by building a long-run measure of historical happiness going back more than two hundred years and to use this new measure (which we call the HPS index) to begin to address some of these gaps in our understanding.

Building such a series is not easy. The main way we measure happiness now is through surveys, but how can we survey our ancestors who lived in the distant past? Our answer is to derive sentiment from existing sources, in particular from the Google Books corpus of several million books stretching back several centuries. Text analysis of this sort has been used by everyone from market researchers, the security services and psychiatrists to understand what is going through the mind of the authors of text. Using the latest semantic analysis, the “big data” afforded to us by the Google digitization project and psychological techniques for extracting mood from text, we form an index that goes back to 1776. This index is highly correlated with normal survey-based measures from the 1970s onwards, and allows us to track changes in happiness for about two hundred years further back than previous studies.

While this is a very new area in which we expect to make significant progress in the years to come, we can already identify particularly gloomy times in the past associated with wars and economic

crises, and can answer some of the most pressing questions about the history of happiness and main drivers of unhappiness through time. For instance, for the 6 advanced nations that we analyse (the USA, UK, Germany, France, Italy and Spain) happiness does not appear to correlate with income over time, confirming the well-known “Easterlin paradox” in a long-run historical context for the first time. We can also see that happiness is very much related to two key measures of health: life expectancy and infant mortality (which operate differently on happiness suggesting both are independently important). Just as historical measures of income have been well-used over the past decades to help our understanding of what matters for wealth, we would hope that our measure of historical happiness will provide the tools to further our understanding of wellbeing. Moreover, the innovative methods we use need not be restricted to understanding happiness and we are already looking at how to understand changes in other important socio-economic concepts over the long-run.