

Book Review: Big Data: A Revolution That Will Transform How We Live, Work, and Think

by blog admin

May 2, 2013

In **Big Data: A Revolution That Will Transform How We Live, Work and Think**, two of the world's most-respected data experts reveal the reality of a big data world and outline clear and actionable steps that will equip the reader with the tools needed for this next phase of human evolution. **Niccolo Tempini** finds that rather than showing how the impact of data-driven innovations will advance the march of humankind, the authors merely present a thin collection of happy-ending business stories.



Big Data: A Revolution That Will Transform How We Live, Work, and Think. Kenneth Cukier and Viktor Mayer-Schonberger. Hodder. March 2013.

Find this book

My issue with Big Data is that it does not take big data seriously enough. Although the authors have pedigree (Editor at the Economist; Professor at Oxford) this is not an academic text: it belongs to that category of popular essays that attempt to stimulate debate. Anyone who works with data (e.g. technologists, scientists, politicians, consultants) or questions what will be borne from our age of data affluence may have expectations for this book - unfortunately it falls short on providing any real answer.

The book paints an impending revolution in mighty strokes. The authors claim the impact of data-driven innovations will advance the march of humankind. What they end up presenting is a thin collection of happy-ending business stories — flight fare prediction, book recommendation, spell-checkers and improved vehicle maintenance. It's too bad the book's scientific champion [Google Flu Trends](#), a tool which predicts flu rates through search queries, has proven so fallible. Last February it forecast almost twice the number of cases reported by the official count of the [Center for Disease Control](#).



amazon

Big data will certainly affect many processes in a range of industries and environments, however, this book gestures at an inevitable social revolution in knowledge making ('god is dead'), for which I do not find coherent evidence.

The book correctly points out that data is rapidly becoming the "raw material of business". Many organisations will tap into the new data affluence, the outcome of a long historical process that includes 'datafication' (I'll define later) and the diffusion of technologies that have tremendously reduced the costs involved in data production, storage and processing.

So, where's the revolution? The book argues for three rather simplistic shifts.

The first shift – the new world is characterised by "far more data". The authors say that just as a movie emerges from a series of photographs, increasing amounts of data are as important because quantitative changes bring about qualitative changes. The technical equivalent in big data is the ability to survey a whole population instead of just sampling random portions of it.

The second shift is that “looking at vastly more data also permits us to loosen up our desire for exactitude”. Apparently, in big data, “with less error from sampling we can accept more measurement error”. According to the authors, science is obsessed with sampling and measurement error as a consequence of coping in a ‘small data’ world.

It would be amazing if the problems of sampling and measurement error really disappeared when you’re “stuffed silly with data”. But context matters, as Microsoft researcher [Kate Crawford cogently argues](#) in her blog. It is easy to treat samples as n=all as data get closer to full coverage, yet researchers still need to account for the representativeness of their sample. Consider how the digital divide – some people are on the Internet, others are not — affects the data available to researchers.

While a missed prediction does not cause much damage if it is about book recommendations on Amazon, a similar error when doing policy making through big data is potentially more serious. Crawford reminds us that *Google Flu Trends* failed because of measurement error. In big data, data are proxies of events, not the events themselves. *Google Flu Trends* cannot distinguish with certainty people who have the flu from people who are just searching about it. Google may tune “its predictions on hundreds of millions of mathematical modelling exercises using billion of data points”, but volume is not enough. What matters is the nature of the data points and Google has apples mixed with oranges.

The third and most radical shift implies “we won’t have to be fixated on causality [...] the idea of understanding the reasons behind all that happens.” This is a straw man argument. The traditional image of science the authors discuss (fixated with causality, paranoid about exactitude) conflates principles with practices. Correlational thinking has been driving a lot of processes and institutional behaviours in the real world. Nevertheless, “*Felix, qui potuit rerum cognoscere causas*” (Fortunate who was able to know the causes of things) – which happens to be the motto of the LSE – is still bedrock in Western political life and philosophy. The authors cannot dismiss causation so cavalierly.

However, it appears that they do. Big data, they say, means that the social sciences “have lost their monopoly on making sense of empirical data, as big-data analysis replaces the highly skilled survey specialists of the past. The new algorithmists will be experts in the areas of computer science, mathematics, and statistics; and they would act as reviewers of big data analyses and predictions.” This is an odd claim given that the social sciences are thriving precisely because expert narratives are a necessary component of how data becomes operational. This book is a shining example that big data speaks the narrative experts give it. What close observers know is that even at the most granular level of practice, analytic understanding is necessary when managers attempt to implement these systems in the world.

The book is blinded by its strongest assumption: that quantitative analysis is devoid of qualitative assessment. For the authors, to *datafy* is merely to put a phenomenon “in a quantified format so it can be tabulated and analysed.” Their argument, that “mathematics gave new meaning to data – it could now be analysed, not just recorded and retrieved”, implies that analysis begins only after phenomena get reduced to quantifiable formats. Human judgement is just an inconvenience of a ‘small data’ world that has no role in the process of making data. This is why they warn that in the impending world of big data, “there will be a special need to carve out a place for the human”.

It is hard to see how imagination and practical context will suddenly cease to play a fundamental role in innovation. But innovation could definitely be jeopardised if big data systems are not recognized for what they are – tools for optimising resource management. Big data may not be an instrument of discovery; while certainly it is a way of managing entities that are already known. Big data promises to be financially valuable – because it is primarily a managerial resource (e.g. pricing fares, finding books, moving spare parts, etc.).

In the world according to Cukier and Mayer-Schönberger, all the challenges of knowledge-making are about to evaporate. With big data affluence – sampling, exactitude, and the pursuit of causality will no longer be issues. The most pressing question is the problem of data valuation. Now there is a problem the authors are willing to discuss seriously: how can data be transformed into a stable financial asset when most of its utility as a predictive resource is not predictable?

So eager are the authors to mark the potential value of big data for organisations (data can only be an asset to a corporation) that they overlook the impact of these systems on other social actors. So what if big data environments reconfigure social inequalities? While the citizen will earn new responsibilities (like privacy management), only corporate entities will be able to systematically generate, own and exploit big data sets.

Big data is serious. There will be winners and there will be losers. What the public need is a book that explains the stakes so that they can be active participants in this revolution, rather than be passive recipients of corporate competition.

Niccolò Tempini is a [PhD Candidate in Information Systems](#) at the London School of Economics and Political Science. You can follow Niccolò on Twitter [@tmpncl](#). Read more reviews by Niccolò.