

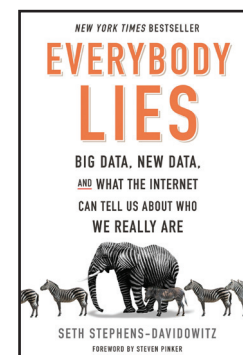
STEPHENS-DAVIDOWITZ, S. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are.*

New York: Dey Street Books, HarperCollins Publishers, 2017, 352 pages.
ISBN 978-0-06-239085-1.

BY MANUEL MARTÍNEZ NICOLÁS

Lecturer at Universidad Rey Juan Carlos

manuel.martinez.nicolas@urjc.es



Pantagruel in the Panopticon

It was the indescribable Dr House, the doctor-cum-detective star of the eponymous TV series, who popularised the scathing saw that “everybody lies”. An expert in diagnosing rare diseases based on strange or unusual symptoms, experience had convinced him that the last thing he could trust was what the patients themselves told him, as they were more concerned, even in those circumstances, about protecting their reputation than helping to clear up the source of their ailments. At all events, the social sciences have long assumed everybody lies since their primary source of data, people, exhibit an irrepressible tendency to say what they are, think, feel and do by passing it through the sympathetic sieve of what they would like to be, think, feel and do. Hence they lie and it is only with great difficulty that social scientists are able to control these well-known phenomena (social desirability, spiral of silence, etc.) when determining the attitudes, states of opinion or behaviour prevalent in human societies. And this is where we were when big data came along.

Stephens-Davidowitz has a PhD in Economics from Harvard and is a former data analyst for Google and currently a specialist columnist at *The New York Times*. He thus has a solid academic background, extensive experience in dealing with digital data and an undoubted vocation for communication, qualities which he pours into a book designed to show big data’s potential for social research by celebrating the advent of what he has no qualms about announcing as a “revolution”. The result is a vibrant and suggestive book, almost always carefully argued albeit sometimes a little over the top.

The author suggests at the start of this celebration that much of what we know about people is wrong. This is because what we think we know about them has no more empirical backing than what they tell us about themselves, and it is well known that everybody lies. However, henceforth we no longer need to trust what people tell us because the tracks left by Internet search engines (for example Google) and social media (for

example Facebook) are a sort of “digital truth serum” which finally discloses what people actually think, want and do. This is the potential of big data, honest indicators of what goes through people’s minds (almost a “cerebroscope” says Steven Pinker in his preface to the book) and which are especially useful for probing the kinds of behaviour most likely to be subjected to the nefarious psycho-sociological intermediations (reputation, desirability, protecting your own image, etc.) that lead people to lie: discriminatory attitudes (racism, homophobia, sexism), sexual behaviours, health problems, stigmatised conditions (for example mental illness) or socially frowned upon, banned or straightforwardly criminal practices (addictions, abortion, child abuse, harassment, etc.). A gargantuan data party, in short, that would turn these big data into a genuine panopticon from which little or nothing of what happens in the human mind could be removed.

But what exactly is big data’s potential to open up that “new path for social science in the 21st century” which Pinker envisions in the preface? At the end of chapter 2 (pages 53-54), Stephens-Davidowitz gives a summarised answer to this question by sketching the plan he will develop in the central part of the book (“The Powers of Big Data”), where, the burden of proof, he provides the results of the dozens of studies he has conducted with digital data, albeit with a deep-seated fondness for what can be done using Google algorithms.

To begin with, these digital data are “honest data”: “In the pre-digital age, people hid their embarrassing thoughts from other people. In the digital age, they still hide them from other people, but not from the internet and in particular sites such as Google or Pornhub, which protect their anonymity. These sites function as a sort of digital truth serum”. (page 54). And they are “honest”, says Stephens-Davidowitz, not only because anonymity is protected, but also because the user has an “incentive” not to distort the digital trail they leave in their wake (searches, visits to websites, etc.): getting their needs met. If I am depressed, I may not admit it in a survey, but it does not make much sense for me to search the internet for information

if I really am not or unless I want to help someone who is. There are no two ways about it.

More subtle and revealing is the argument that the potential of these new big data is not, against all expectation, that they are “big” but rather that they are “new”. They are “a new type of data”, “unconventional” data; at first glance trivial, even ridiculous, yet available with no restriction other than the analyst’s imagination to turn them into relevant social indicators. “The big data revolution is less about collecting more and more data. It is about collecting the right data” (page 62). The right data; that is to say, the significant data from among the inextricable digital tangle we have to hand. The potential of this revolution that the author foretells (and exemplifies with his books and papers) depends on setting off the imaginative spark in social scientists which examines searches for jokes about people of colour to map racial prejudice in the United States and relate it to the territorial distribution of electoral support for Donald Trump. Or which investigates pornography websites to estimate the high percentage of American homosexuals who have not yet dared to “come out of the closet” in the most intolerant states. Or which traces interest in ways of inducing an abortion (pills, herbs, vitamin C overdoses, etc.) to challenge the credibility of official figures in this field. Or which observes the frequency and distribution of “my mum/my dad hits me” searches to reconstruct the state of child abuse by linking it to the unemployment rate in particular places and at certain times.

This is the meaning of the revolution heralded here by Stephens-Davidowitz: shifting this “evidence” to the category of “data” which are significant for social research. A revolution whose advent will require, probably more than ever before, that intangible factor which hangs around his proposal: the “scientific imagination”, which in this arena can only be nourished by rigorous epistemological training that helps to pose the relevant questions and look for the best way to give them an empirically-based answer. It is not surprising, then, that it is these *apostles* of the algorithm (some excessive; others restrained) who are once more advocating (see also Finn, 2018) what had already been proclaimed half a century ago by Wright Mills (1961): turning the “sociological imagination” into the only valid attitude to properly address the world’s previously analogical and now digital complexity.

References

FINN, E. *La búsqueda del algoritmo. Imaginación en la era de la informática*. Barcelona: Alpha Decay, 2018 (first published: 2017).

WRIGHT MILLS, CH. *La imaginación sociológica*. Mexico: Fondo de Cultura Económica, 1961 (first published: 1959).