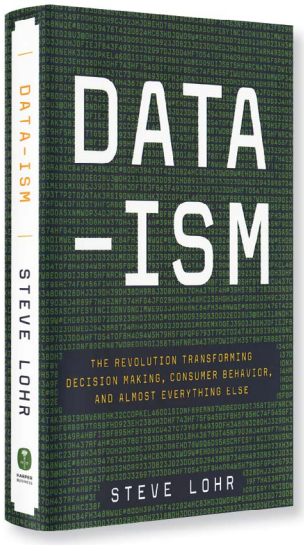
SOUNDVIEW Featured Book Review
www.summary.com

DATA-ISM
The Revolution Transforming Decision Making, Consumer Behavior and Almost Everything Else
by Steve Lohr

HarperCollins © 2015, 239 pages

In a January 2015 *New York Times Review of Books* essay, critic and magazine editor Leon Wieseltier warned against a post-humanist — after the human — culture in which technological devices and data replace human beings and thought. "Quantification is the most overwhelming influence upon the contemporary American understanding of, well, everything," he writes. "It is enabled by the idolatry of data, which has itself been enabled by the almost unimaginable data-generating capabilities of the new technology." In short, "Where wisdom once was, quantification will now be."



One might assume that Wieseltier does not have a copy of *Data-ism*, a new book from *New York Times* technology journalist Steve Lohr, on his bedside table. At first glance, *Data-ism* seems to be the embodiment of Wieseltier's fear that quantification has replaced wisdom. The "ism" title seems to promise an introduction (manifesto?) to the philosophy of quantification. The subtitle is not timid: "The revolution transforming decision making, consumer behavior and almost everything else." And within its pages, Lohr does a masterful job of describing all of the possibilities of "big data."

He starts, for example, with the operations of McKesson Corporation, which distributes one third of all pharmaceuticals in the U.S. to 26,000 customer locations, including roughly 240 million pills a day.

As Lohr explains, "Many companies live in a digital Tower of Babel, a hodgepodge of incompatible computer systems and data formats added over the years. Not McKesson." Using sensors and shipment-tracking software, the company accumulates massive amounts of data, although what is being measured (pills, prices and shipment miles) can be easily translated into numbers. As a result, Lohr writes, "There is not a log of randomness in the McKesson data. Its data was plentiful, stable and reliable, which provided sturdy building blocks for IBM to build its facsimile of McKesson's operations — a flight simulator for decision-making."

How exactly does this "flight simulator" work? In two ways, according to Lohr. One is to offer McKesson a granular analysis of operations, using data to provide profit and loss figures for every product, supplier and customer. In other words, the data describes what is happening today. But it can also help — and perhaps the key word there is "help" — McKesson executives figure out the future. Specifically, writes Lohr, "the technology is a tool for modeling what-if decisions, using the digital replica of the physical world to peer into the future to make more accurate predictions and better decisions."

One of those decisions was to centralize the distribution of very expensive drugs (such as cancer drugs that can cost thousands of dollars for one month's treatment). Centralization, which requires expensive air shipments to customers (nearly 10 times as expensive, as opposed to shipping the drugs by truck to distribution centers and then on to customers). However, IBM's modeling software predicted that savings in inventory levels for certain drugs would more than make up the higher air freight expenses. When McKesson tested the prediction with a pilot project, the software was vindicated: Inventory levels were cut in half, resulting in overall savings on distribution costs — and on-time delivery for the drugs went from 90 percent to 99

percent. "The software gave McKesson the clarity and the confidence to go ahead," Lohr writes. "McKesson's success illustrates where the big-data approach shines today. It shows data really being used to guide decisions and to make better decisions, ones that trump best guesses and gut feel, experience and intuition."

Those last words about how data will "trump best guesses and gut feel, experience and intuition" are no doubt the kind of attitude that will send chills into the hearts of social commentators such as Wieseltier. Is quantification indeed replacing wisdom? Will there be a time when humans — or human thinking — are considered superfluous? Will we let machines run the world?

In his *New York Times* essay, Wieseltier wrote, "All revolutions exaggerate, and the digital revolution is no exception." Anyone who has been reading business books for the past few years will know what he is talking about — and why *Data-ism* is an exceptional book. Lohr is a data enthusiast — there is no doubt of that — but in his book he demonstrates that he is first and foremost a conscientious journalist with the perspective to see all sides of an issue. He extolls the virtues of big data but also compellingly describes its limitations.

The Girl and the Butterfly

A fundamental challenge for data-driven artificial intelligence is described through the example offered to Lohr by Tom Mitchell, chairman of the machine-learning department at Carnegie Mellon. The example involves two sentences:

The girl caught the butterfly with the spots.

The girl caught the butterfly with the nets.

Only one word in the two sentences is different. Humans reading those two sentences understand that girls don't have spots and butterflies don't have nets. Thus, the spots in the first sentence refer to the butterfly, while the nets in the second sentence are associated with the girl. The advantage of humans in this example can be summarized in one word: context. The greatest challenge for artificial intelligence is to be able to put the data it accumulates in context.

To explain the challenge, Lohr starts with a simple example: the number 39. By itself, it is a data point but tells us very little. Add the word degrees, and the information becomes a little clearer, although we could still be talking about a temperature or an angle. Add the word Celsius, and we know it is a hot temperature (about 102 degrees Fahrenheit). Tell us that this temperature comes from a person's mouth, and we now know that the number 39 signifies illness.

Such is the power of context. Piles of numbers can mean nothing without context. Context can be achieved in two ways, Lohr explains: correlation, which connects data to action or behavior in the real world, and association, the more sophisticated ability of computers to draw context from putting words together (as in the exercise above with the number 39).

Correlation, which is what most people might think of when they refer to "data mining," is not new. Lohr gives the example of Walmart, which, more than 10 years ago, noticed from mining its data that when hurricanes were predicted, Walmart consumers stocked up on strawberry Pop-Tarts (buying seven times more than usual) and beer (the best-selling pre-hurricane item).

The startup Zest Finance uses correlation to determine the risk of payday borrowers. Its sophisticated algorithms allow Zest Finance to reduce the risk of default by its borrowers by 40 percent. The secret is to find the unconventional numbers, or data points, to enter in the equations. In the case of Zest Finance, for example, one of the data points they include is how long a potential borrower has had a cellphone or, even more surprising, how they type their names into websites: all uppercase means they are less likely to repay the loan, while proper-case users are the best risks.

Zest Finance may or may not know why there is a correlation between how you type your name into a website and how likely you are to repay a loan, but does it matter? The answer to that question — does "why" matter? — is a topic of heated debate in the big-data community. Lohr notes that the authors of *Big Data* insist that "identifying causal mechanisms is a self-congratulatory illusion; big data overturns this."

Other data enthusiasts believe that understanding the "why" is vital, and this is achieved only by pairing measurement with theoretical models. Richard Berner, the first director of the U.S. Treasury Department's Office of Financial Research, tells Lohr that the measurement-is-everything mentality "is what led us into trouble in the financial crisis." Lohr explains Berner's position: The naive assumption that housing prices would only go up was largely based on a blinkered view that analyzed data in more recent years, when data was plentiful and consistent. It ignored earlier financial crises, when data sets were sparse and messy.

For Berner, the promise of big data is achieved when data is combined with "theories or models of how the economy behaves," Lohr explains. Both are necessary.

Even David Ferrucci, the man whose IBM team created the Watson, agrees with Berner. "People are so enamored with the data-driven approach that they believe correla-

tion is enough." More sophisticated decision-making in an area such as business strategy requires understanding "the why of things," he tells Lohr. "You're going to have to be able to address the question, 'What's really going on?'"

Intuition Has Its Own Problems

Lohr does not gloss over the limitations of machines, as some data enthusiasts might. However, he also notes that human intuition, despite the ability to place spots on butterflies and nets in the girl's hands, has its own limitations.

Lohr takes an example from Daniel Kahneman's seminal book *Thinking, Fast and Slow*. Fast thinking is the intuitive, gut-feel conclusions we make and decisions we take without taking the time to investigate and reflect on the evidence. Slow thinking takes more time and requires more effort but is based on research and reflection on that research.

It would, of course, be nearly impossible to "navigate daily life in real time," in Lohr's words, "if every decision we made involved slow thinking." People quickly assess any situation based on the past, which allows them to interpret, he writes, "a likely future, one that is emotionally and socially coherent with that person's experience."

However, fast thinking can sometimes take us off base. Lohr uses an illustration from Kahneman's book, in which research participants are told about a man who is named "Steve" and who is described as "meek." The participants are then asked if this man is more likely to be a farmer or a librarian. The vast majority reply "librarian" even though the data shows that there are 20 times more male farmers than librarians.

> "Lohr draws a picture of a world in which both people and computers can help each other overcome the others' weaknesses. The core question of big data... is the 'division of labor' between man and machine."

Intuitive, fast thinking based on the clues given — the name "Steve" and the description as "meek" — leads to the wrong answer. Just a little data, however, would have made a difference.

In short, Lohr draws a picture of a world in which both people and computers can help each other overcome the other's weaknesses. The core question of big data, in other words, is the "division of labor" between man and machine. For some, machines are destined to be a "clever helper," Lohr writes, "a tireless digital assistant to see what its overworked human boss missed." As noted above, the best decisions are based on measurements that can inform theoretical models.

Others, however, see a toppling of the traditional order of humans as the senior partner. Put forth more by scientists, researchers and academics than businesspeople, Lohr writes, this school of thought predicts that computers will some day be able to engage in "experiential learning at scale" — in essence becoming smarter than humans. Humans might still make the rules, but instead of computers assisting humans, it will be the reverse: humans assisting the computers.

Will We Trust Machines?

One of the case studies in the book involves a function — maintaining the temperature in the home — that could be moving away from humans to machines. The future is perhaps already visible in the sophisticated algorithms of the Nest "learning" thermostats, which use sensors to collect data on the energy use and activity in a house and then, after about a week, have enough data to take over the regulation of the temperature in the house, choosing the most energy-efficient settings when appropriate.

The trouble, as Lohr recounts in his case study on the company, is that people didn't like the thermostat controlling the temperature settings in their house. So, instead of the thermostat automatically changing the settings to the optimal number, the thermostats would show a green leaf when the user chose an energy-saving setting. Users who "chased the leaf" would be saving the most energy. "The green-leaf solution tells us something about the uneasy alliance between people and computers," Lohr notes. "The people who tried out the prototype thermostats — mechanical autocrats who set temperatures on their own — felt the machines had taken over."

The idea of machines taking over and pushing humans into an inferior position clearly rankles, even when the issue is setting energy-saving thermostat settings. However, Lohr writes, people can learn to relinquish control if they learn to trust the computers. Today, for example, most owners of Nest thermostats let the machine decide when to change the temperature settings.

However, it's one thing to let machines set the thermostat. It's another to let machines drive the car (as in the self-driving cars currently being developed) or, even more frightening, make patient-treatment decisions. Yet even in medicine, big data can play a major role, according to Lohr.

From Wall Street to Medicine

The career of 32-year-old Jeffrey Hammerbacher is one of the threads that runs through Lohr's journey into the world of dataism. A Harvard graduate from the Midwest,

Hammerbacher started his career on Wall Street, built the first team of data scientists at Facebook, and walked away from potential millions of dollars at Facebook to launch a San Francisco startup that specializes in software for data scientists, before, most recently, moving to Mount Sinai Medical Center in New York City.

"The best minds of my generation are thinking about how to make people click ads. That sucks." Lohr disagrees, noting that "online advertising is an economic virtue" because the resulting marketing efficiency reduces costs, which frees up money for investment elsewhere in the economy. To which Hammerbacher responds, "So why not work on the elsewhere?"

Currently, the elsewhere for Hammerbacher is in medicine. At Mount Sinai Medical Center, he is leading the charge on transforming medicine into what one Mount Sinai scientist calls "an information game": doctors no longer reign supreme, as data-driven software takes the lead in monitoring patients … and recommending treatments. Hammerbacher is focused on projects related to the treatment of cancer, diabetes, Alzheimer's and Crohn's disease, but his long-term interest is mental health, sparked by a close friend's downward spiral into depression, schizophrenia and eventual suicide, and his own battles with anxiety and depression and a self-destructive lifestyle that landed him in the hospital at 27 — the same age, he noted at the time, that Jimi Hendrix, Janis Joplin, Jim Morrison and Kurt Cobain died. Hammerbacher is now healthy and fit and ready to focus on using all the power of big data to combat mental illness, which he believes is a wide-open opportunity.

"The causes of most mental disorders are little understood," Lohr writes, "but vast, new data sets are emerging from genomics and brain-scanning imagery." Or, as Hammerbacher tells a Silicon Valley audience, "Think about mental disorders as one of the most critical and most challenging data problems for us out there."

Remember Taylorism

In *Data-ism*, Lohr leads his readers on a wide-ranging tour of industries and domains, from medicine to viticulture, to hotels and people, from ambitious young math wizards, to groundbreaking academics, to progressive CEOs. He draws a picture of a world in which big data has a growing importance in every facet of our lives. However, he also warns about data enthusiasts who believe that dataism is the answer to every question and problem. These enthusiasts, he writes, may be guilty of the "overreaching hubris" of the early 20th-century Frederick Taylor enthusiasts, who believed that Taylor's "scientific management" method of measurement and analysis held the answer to all issues of work and home.

"What struck me while reporting these stories, and what came up repeatedly in conversations with artificial intelligence experts, is what an awesome thing the human brain and what we call general human intelligence really are," Lohr writes in the final paragraphs of his book.

General intelligence involves the effortless capacity to tap life experience and make intuitive connections and quick decisions… Then there is the human brain as a processor, cramming incredible computing power into a tiny space and using only 20 watts of energy. By contrast, the Watson computer that won its Jeopardy! contest with human champions burned 85,000 watts.

Data-ism is perhaps one of the most balanced, level-headed examinations of the potential of big data. Lohr never hesitates to give voice to the critics or skeptics of a data-driven world, nor fails to point out the limitations of artificial intelligence. It is this balance and restraint, however, that makes Lohr and his book the most persuasive champions of the massive and generally positive changes that "the virtuous cycle of more and more varied data and smarter and smarter algorithms, written by human programmers" will make in our lives. In short, quantification will not replace wisdom, as Wieseltier fears; but, Lohr shows, it will augment our wisdom — working with our amazing human brains — to help us make better decisions, free our time and energy to focus on the tasks where we can make the most difference, and, ultimately, make the world a much better place.

The author: Steve Lohr reports on technology, business and economics for the *New York Times*. He was a foreign correspondent for a decade and served brief stints as an editor before covering technology starting in the early 1990s. In 2013, he was a member of the team that was awarded the Pulitzer Prize for explanatory reporting. He has written for magazines such as the *New York Times Magazine*, the *Atlantic* and the *Washington Monthly*. He is the author of a history of computer programming, *Go To: The Story of the Math Majors, Bridge Players, Engineers, Chess Wizards, Maverick Scientists and Iconoclasts — the Programmers Who Created the Software Revolution.*

1-800-SUMMARY
service@summary.com

Published by Soundview Executive Book Summaries® (ISSN 0747-2196), 511 School House Rd, Suite 300, Kennett Square, PA 19348 USA. To subscribe, call 1-800-SUMMARY (240-912-7513 outside the United States), or order online at www.summary.com. Multiple-subscription discounts and corporate site licenses are also available.

Rebecca S. Clement, Publisher; Sarah T. Dayton, Editor-in Chief; Ashleigh Imus, Senior Editor; Amanda Langen, Senior Graphic Designer
Review by Chris Murray, Editor-at Large. #FBR-0615