**BOOK REVIEWS**

**Information Shift**

**BY JEFFREY L. DEAL**

**BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK**

*By Viktor Mayer-Schonberger and Kenneth Cukier*


242 pp., $27.00

Big data is all the buzz! Anyone connected with the business world understands that they need to be harnessing big data. A recent Senate report estimated that private companies collect, mine, and sell upward to 75,000 individual data points on each US consumer. But just what is big data, and what differentiates it from small data or more traditional statistical analysis? These are some of the questions answered in *Big Data: A Revolution That Will Transform How We Live, Work, and Think*.

Authors Viktor Mayer-Schonberger and Kenneth Cukier state that they are not “so much big data’s evangelists, but merely its messengers.” They might also describe themselves as big-data chroniclers, for they tell the story of big data today, and how we transitioned to this point from the “small data” era. Mayer-Schonberger is a professor of Internet governance and regulation at the Oxford Internet Institute, Oxford University, in the United Kingdom. Cukier is the data editor of the *Economist*. Together they present the evolution of thinking and science that makes big data such an important part of our world today.

Much of the book focuses on three primary “shifts” in the way in which information is analyzed. They label these shifts as *more, messy, and correlations.* More speaks to the relatively newfound ability to analyze vast amounts of data. Until the recent computer revolution, our ability to collect, organize, store, and analyze data was limited. To keep it manageable and useful, organizations were careful to work with only the most important data, as they perceived it. With increased computing power came the ability to manage and analyze data sets of enormous size—and ultimately the freedom to view data from different angles and explore selected aspects from a closer perspective.

Next up is the *messy* shift. Before the advent of modern computing and storage and analysis with advanced algorithms, organizations were very careful with their data. Prior to the big-data revolution, the rules of statistical analysis meant that data were only useful if they were complete and accurate. Big data opened the door to messy data. Mayer-Schonberger and Cukier state that “big data transforms figures into something more probabilistic than precise. Big data opened the door from exactitude to inexactitude, and random sampling has given way to *N = All.*”

The third shift can be summarized as follows: “Knowing what, not why, is good enough.” This shift, *correlations*, reflects a move away from the age-old search for causality. While correlations may be of interest in a small-data world, they glow in the world of big data. Previously, a hunch was needed in order to start the data gathering and analysis process. With so much data and computing power readily at hand, correlations are discovered more quickly and at lower costs. One no longer needs a hypothesis before testing begins. One runs the data and studies the correlations that are found to determine which ones may be useful.

Many interesting case studies illuminate the three shifts and other key points in the book. Everything from the tracking of international money transfers, to political polling, and DNA analysis is included in the discussion of the first shift. In examining how big data deals with the problem of messiness, the authors discuss the multiple efforts toward machine translation that led to the successful development of Google Translate. Who would have guessed that throwing more messy data into the mix would increase accuracy? And as for understanding correlations in practice, the authors point to Amazon’s recommendation system, a well-known example.

In addition to the three shifts, the authors discuss important principles that can shape big data, including values, risks, and control. These are key to the high-level understanding of big data and its implications for society today. The *Health Affairs* reader will not find in this book a lot in the way of big data applied specifically in health care, although there is some discussion. The reader will, however, appreciate the high-level understanding of big data that comes from the book’s many real-life examples and anecdotes illustrating how small-data thinking evolved to big-data realizations. *Big Data* is not a “how to” for data scientists, but it is a “how did we get here and where are we headed” for those who want a deeper understanding of the big-data revolution.

Jeff Deal (deal@datamininglab.com) is the vice president of operations for Elder Research, a data mining and predictive analytics consultancy, in Charlottesville, Virginia. He is also the program chair for the inaugural Predictive Analytics World for Healthcare conference, in Boston, Massachusetts, October 6–7, 2014.