



When the Sample Is the Population: Big Data

A Review of

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

by Viktor Mayer-Schönberger and Kenneth Cukier

New York, NY: Houghton Mifflin Harcourt, 2013. 242 pp. ISBN

978-0-544-00269-2. \$27.00

<http://dx.doi.org/10.1037/a0035106>

Reviewed by

David J. Pittenger 

In September 1854, John Snow plotted on a map the location of cholera deaths in a portion of London, England, and saw a cluster of points that radiated from one of 12 water pumps in the area. When the pump was disabled, forcing the residents to use the other pumps, the cholera epidemic abated (Tuft, 1983). In 1936, Allport and Odbert published a monograph examining 17,953 adjectives, selected from a then-current dictionary of well over 400,000 words, that describe human characteristics, including the colorful *agathokakological*. Their review of the terms foreshadowed the development of the Big-Five personality model. In their new book, Mayer-Schönberger and Cukier describe how researchers using Google queries developed a model that can accurately predict the spread of a flu virus long before the Centers for Disease Control and Prevention are aware of a problem.

An obvious commonality among these examples is that “public” data provided the researchers useful insights. Another common theme is that the researchers did not use sampling techniques, as the data represented all cholera deaths in the neighborhood, all adjectives English-speaking people use to describe one another, and all Google searches related to flu symptoms. The Google example is unique, however, given the magnitude of the data collected and the ability the researchers had to analyze the data in real time. These characteristics are the focus of *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Readers will find that this well-written book offers much for psychologists who are dedicated to using the empirical method to examine human behavior and who remain attentive to their civic responsibilities.

The first message about big data is that interesting, practical, and vexing problems have been addressed with big data analysis. Throughout the book, the authors provide convincing anecdotes of how the judicious use of big data has eliminated the use of hunches for selecting players for major league baseball teams, improved spell-checking and language translation software, and reduced operating costs for large corporations.

A second message about big data is that everything we psychologists have been taught about conducting research using the scientific method is not relevant when one is dealing with big data. There is no concern about statistical sampling techniques or participant selection as big data researchers use all the data—the sample is the population. Gone, too, is the concern about the precision with which the data are measured. Although experimental psychologists are continuing to refine their measurement techniques to reduce random or measurement error and to increase the fidelity of measurement, big data researchers are content with the data that they have and look for the larger trends rather than worry about the minutia of measurement precision.

In essence, traditional research data are expensive to gather. Accordingly, there has been an evolution of rituals for crafting unambiguous operational definitions, rituals for precise measurement of the independent and dependent variables, and rituals for creating samples that support internal, statistical, and external validity. By contrast, the brave new world of big data uses data that are inexpensive to acquire and store, and nearly limitless in size. A modern-day Allport and Odbert can now access all dictionaries of all languages, translate the terms into English, and then classify the terms by traits. Indeed, they can access all digitized texts and examine patterns of adjective use to characterize people.

Big data research is also atheoretical. Such a perspective is reminiscent of Skinner's 1950 article asking whether psychology needed theories of learning. According to Mayer-Schönberger and Cukier, the goal of big data is not to affirm or refute critical theory. Rather, the goal of big data is to find interesting patterns of data that serve some useful goal.

One of the interesting features of big data is the types of data available to researchers. What big data researchers now call *data exhaust*, an earlier generation of psychologists called *unobtrusive measures*. The searches that people perform, the pages they view, and how long they linger at a site become ideographic information, thus allowing the clever analyst to use online behavior as a reliable predictor of a medical condition, as was the case when the retail store chain Target detected an unmarried teen's pregnancy before she shared the information with her family (Hill, 2012). Might contemporary experimental psychologists be able to use big data to examine modern racism, obedience and conformity, the effectiveness of various forms of psychotherapy, or mechanisms that control the construction of sentences? The potential is intriguing.

The authors examine how big data has brought about the demise of sole reliance upon the disciplinary expert. According to this account, those with expertise in data analysis and cognition are best able to capture useful insights from big data. The authors do concede that subject area experts will continue to exist, but they will need to work collaboratively within a larger interdisciplinary team whose members view questions from multiple perspectives. All this is to say that psychologists, with their background in studying cognition, social behavior, and other behavioral phenomena, may find themselves comfortable in such a venue.

The brave new world of big data portends solutions to the many problems of the day. Although IBM's Watson may have demonstrated that computers can be programmed to win *Jeopardy*, this proof-of-concept exercise illustrated how algorithmic searchers of near-limitless data sets can yield decisive answers (Thompson, 2010). The brave new world of big data also portends a dystopia that eliminates reasonable expectations of privacy, localizes

control to only those who have access to data, and is characterized by algorithmic decision making. As the authors note, there is an ample history of technology illustrating how new technology affects social culture and necessitates certain checks. Indeed, as empiricists and citizens, psychologists will need to consider their ethical responsibilities when working with big data; they will be assisted in this if they read Chapter 8, which examines the risks of a “data above all else” perspective of big data.

The book is an important read for the psychologist qua psychologist who may be able to make effective use of big data. The authors foresee an earth covered with sensors that will allow for the detailed study of climate change to reveal methods to correct the effects of pollution. By extension, psychologists may find ample unobtrusive data that will also reveal insights into human behavior not readily found in controlled conditions. The book might well serve as a supplementary book for an experimental psychology course to examine alternative research methods. Finally, *Big Data* is an important read for psychologists concerned about the potential iatrogenic consequences of the applications of big data.

References

- Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, 47, i-171. <http://dx.doi.org/10.1037/h0093360> PsycINFO →
- Hill, K. (2012, February 16). *How Target figured out a teen girl was pregnant before her father did*. Retrieved from <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>
- Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, 57, 193-216. <http://dx.doi.org/10.1037/h0054367> PsycINFO →
- Thompson, C. (2010, June 16). What is I.B.M.'s Watson? *New York Times Magazine*. Retrieved from <http://www.nytimes.com/2010/06/20/magazine/20Computer-t.html>
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.