The fruits of the information society are easy to see, with a cellphone in every pocket, a computer in every backpack, and big information technology systems in back offices everywhere. But less noticeable is the information itself. Half a century after computers entered mainstream society, the data has begun to accumulate to the point where something new and special is taking place. Not only is the world awash with more information than ever before, but that information is growing faster. The change of scale has led to a change of state. The quantitative change has led to a qualitative one. The sciences like astronomy and genomics, which first experienced the explosion in the 2000s, coined the term 'big data'. The concept is now migrating to all areas of human endeavor.

There is no rigorous definition of big data. Initially the idea was that the volume of information had grown so large that the quantity being examined no longer fit into the memory that computers use for processing, so engineers needed to revamp the tools they used for analyzing it all. That is the origin of new processing technologies like Google's MapReduce and its open-source equivalent, Hadoop, which came out of Yahoo. These let one manage far larger quantities of data than before, and the data – importantly – need not be placed in tidy rows or classic database tables. Other data-crunching technologies that dispense with the rigid hierarchies and homogeneity of yore are also on the horizon. At the same time, because internet companies could collect vast troves of data and had a burning financial incentive to make sense of them, they became the leading users of the latest processing technologies, superseding offline companies that had, in some cases, decades more experience.

One way to think about the issue today – and the way we do in the book – is this: big data refers to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more.

But this is just the start. The era of big data challenges the way we live and interact with the world. Most strikingly, society will need to shed some of its obsession for causality in exchange for simple correlations: not knowing *why* but only *what.* This overturns centuries of established practices and challenges our most basic understanding of how to make decisions and comprehend reality.

Big data marks the beginning of a major transformation. Like so many new technologies, big data will surely become a victim of Silicon Valley's notorious hype cycle: after being feted on the cover of magazines and at industry conferences, the trend will be dismissed and many of the data-smitten startups will flounder. But both the infatuation and the damnation profoundly misunderstand the importance of what is taking place. Just as the telescope enabled us to comprehend the universe and the microscope allowed us to understand germs, the new techniques for collecting and analyzing huge bodies of data will help us make sense of our world in ways we are just starting to appreciate. In this book we are not so much big data's evangelists, but merely its messengers. And, again, the real revolution is not in the machines that calculate data but in data itself and how we use it.

To appreciate the degree to which an information revolution is already under way, consider trends from across the spectrum of society. Our digital universe is constantly expanding. Take astronomy. When the Sloan Digital Sky Survey began in 2000, its telescope in New Mexico collected more data in its first few weeks than had been amassed in the entire history of astronomy. By 2010 the survey's archive teemed with a whopping 140 terabytes of information. But a successor, the Large Synoptic Survey Telescope in Chile, due to come on stream in 2016, will acquire that quantity of data every five days.

Such astronomical quantities are found closer to home as well. When scientists first decoded the human genome in 2003, it took them a decade of intensive work to sequence the three billion base pairs. Now, a decade later, a single facility can sequence that much DNA in a day. In finance, about seven billion shares change hands every day on U.S. equity markets, of which around two-thirds is traded by computer algorithms based on mathematical models that crunch mountains of data to predict gains while trying to reduce risk.

From the sciences to healthcare, from banking to the Internet, the sectors may be diverse yet together they tell a similar story: the amount of data in the world is growing fast, outstripping not just our machines but our imaginations. [...]

Take the long view, by comparing the current data deluge with an earlier information revolution, that of the Gutenberg printing press, which was invented around 1439. In the fifty years from 1453 to 1503 about eight million books were printed, according to the historian Elizabeth Eisenstein. This is considered to be more than all the scribes of Europe had produced since the founding of Constantinople some 1,200 years earlier. In other words, it took 50 years for the stock of information to roughly double in Europe, compared with around every three years today.

What does this increase mean? Peter Norvig, an artificial intelligence expert at Google, likes to think about it with an analogy to images. First, he asks us to consider the iconic horse from the cave paintings in Lascaux, France, which date to the Paleolithic era some 17,000 years ago. Then think of a photograph of a horse. […] Yet now, Norvig implores, consider capturing the image of a horse and speeding it up to 24 frames per second. Now, the quantitative change has produced a qualitative change. A movie is fundamentally different from a frozen photograph. It's the same with big data: by changing the amount, we change the essence.