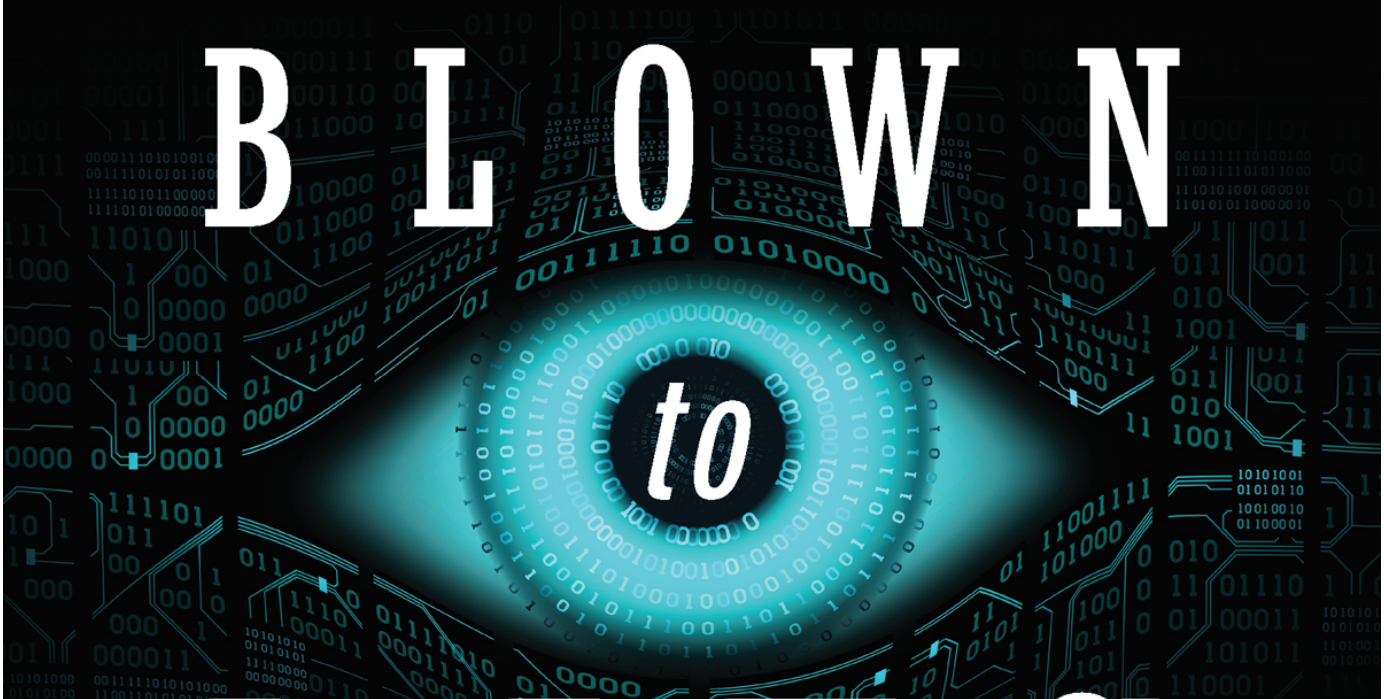
HAL ABELSON • KEN LEDEEN
HARRY LEWIS • WENDY SELTZER

BLOWN to BITS



[YOUR Life, Liberty, and Happiness After the Digital Explosion]

SECOND EDITION

Blown To Bits

Your Life, Liberty, and Happiness After the Digital Explosion

Second Edition

Hal Abelson
Ken Ledeen
Harry Lewis
Wendy Seltzer

CHAPTER 1

Digital Explosion

Why Is It Happening, and What Is at Stake?

This book isn't about computers. It's about your life and mine. It's about how the ground underneath us has shifted in fundamental ways. We all know it is happening. We see it all around us, every day. We all need to understand it more.

The digital explosion is changing everything. In this book we talk about both what is happening and how. We explain the technology itself—why it creates so many surprises and why things often don't work the way we expect them to. It is also about things the information explosion is destroying: old assumptions about our privacy, about our identity, and about who is in control of our lives. It's about how we got this way, what we are losing, and what remains that society still has a chance to put right.

The digital explosion is creating both opportunities and risks. Many of both will be gone in a decade, settled one way or another. Governments, corporations, and other authorities are taking advantage of the chaos, and most of us don't even see it happening. Yet we all have a stake in the outcome. Beyond the science, the history, the law, and the politics, this book is a wake-up call. The forces shaping your future are digital, and you need to understand them.

This book is about the stories we hear and read every day. Stories that are about the profound, often unexpected impact that digital technology is having on our lives. Let's begin with the story of Nicolette Vartuli.

Nicolette couldn't figure out why she didn't get the job. A college senior with a 3.5 GPA, she had prepared for her interview with the investment bank and stayed positive throughout. She kept her head up, smiled, and spoke with confidence. But when the company followed up, it was bad news. She would not be moving on in the hiring process.[1]

Nicolette wanted to know what she had done wrong, but no one could explain why she was rejected—because no one actually knew. She had been interviewed by a computer that used AI software from HireVue to assess her suitability. That software rejected her not because she didn't have some particular qualification but because, as it claimed, the software could detect patterns in people who were successful in the job—and what it observed in Nicolette didn't match. It is easy to understand being rejected because you don't have three years of required experience or some particular skill. This is different. And scary—especially because no explanation was offered for what the software was looking for. And it may be that no explanation could be offered, even if HireVue were willing to disclose its proprietary algorithms. (It is not.)

Companies like this new technology. It is cheaper and more efficient than human interviews. In fact, HireVue, just one of many providers, has completed more than 10 million interviews. Many applicants, by contrast, don't like these automated hiring assistants. It's not just that it feels dehumanizing to be judged by a machine. The companies that offer the service counter that by using technology, more people can get interviews now, and the likelihood of inherent bias on the part of interviewers is diminished. They claim the technology is opening up opportunities, not limiting them—but how do we know?

The instinctive antipathy to automated job screening can't really be because people don't want computers making life-critical decisions. Many such decisions are made by computers today; airplanes and radiation therapy machines are now largely automated systems, for example. Computers now beat highly trained radiologists at spotting cancer tumors in breast X-rays.[2] Would anyone prefer less accurate human screeners? But HireVue's judgments are of a different kind. The program made a decision about Nicolette's humanity. It decided that she was not the sort of person the company should hire, and it did so without explaining to her or anyone else what sort of person would be a good hire and how Nicolette fell short.

Many other systems are today making similar judgments in other human domains. Judges consult computers to assess the risk that criminal defendants will fail to show up for their trials—again by comparing the individuals with others who have been arrested in the past and have been given the benefit of avoiding pretrial detention.[3] Real-estate agents use computers to judge which prospective renters are likely to be deadbeats.[4]

Most of these systems are proprietary, and the companies that make them don't have to disclose how they work. And after all, they argue, human interviewers are no gold standard of impartial judgment. They are prone to all sorts of unfortunate biases and prejudices. That is why tryouts for instrumental musicians are now commonly held out of view of the listeners: When the performers could be seen, women were systematically judged more harshly than men.[5] By matching candidates' interview skills to those of existing workers, HireVue claims, it is

eliminating the most fallible part of the system. It's the human recruiters, HireVue says, who are the "ultimate black box." Maybe—except that HireVue says it is matching candidates to the profile of the best of the bank's current employees. How would anyone know if the software is simply replicating, now automatically, all the prejudices that gave the bank the workforce it now has?

What makes this whole story particularly important is not only that Nicolette was judged by a machine to be unsuitable but that no one—not a human resource manager, not even a programmer—told the HireVue software what criteria to use. It determined those all by itself. The software watched videos of existing employees and picked its own criteria.

The tale of Nicolette's rejected job application is what we call "a bits story." That is, it is not just a job search story; it is a story about the collection, storage, analysis, transmission, and use of trillions of trillions of trillions of individual 0s and 1s. By looking carefully at these stories, we can understand not only the technology behind them, but the implications and risks as well.

Bits represented Nicolette's image as it flowed from her own computer to HireVue's, over wires and cables and probably several kinds of radio waves. The bits were reassembled, taken apart, and analyzed by HireVue's programs. They were somehow compared to trillions of trillions of trillions of bits representing videos of other people, and then a single bit, a single yes or no, came out: continue to the next stage of the hiring process or reject immediately. That bit was a 0 for Nicolette, and that is all she heard back from the company. But HireVue kept all the bits of Nicolette's failed interview; she had to sign over her rights to them in order to get the interview in the first place.

"Algorithmic transparency" is the principle that we should know how computers are making decisions about us. In the words of EPIC (the Electronic Privacy Information Center), "The public has a right to know the data processes that impact their lives so they can correct errors and contest decisions made by algorithms."[6]

New technologies interact in odd ways with evolving standards of privacy, communications practices, and criminal law. Nicolette's story, while important to her, is just one of thousands of bits stories that could be told about any one of us. Every day we encounter unexpected consequences of data flows that could not have happened a few years ago.

When you have finished reading this book, you should see the world in a different way. You should hear a story from a friend or on a newscast and say to yourself, "That's really a bits story," even if no one mentions anything digital. The movements of physical objects and the actions of flesh-and-blood human beings are only the surface. To understand what is really going on, you have to see the virtual world, the eerie flow of bits steering the events of life.

This book is your guide to this new world.

The Explosion of Bits, and Everything Else

The world changed very suddenly. Almost everything is stored in a computer somewhere. Court records, grocery purchases, precious family photos and priceless Hollywood movies, pointless television shows....Computers contain a lot of stuff that isn't useful today but somebody thinks might someday come in handy. It is all being reduced to 0s and 1s—"bits." The bits are stashed on disks of home computers, in the data centers of big corporations and government agencies. Many of the disks aren't even round, spinning things—they are a different kind of storage media, called "disks" for historical reasons. Most of the disks these days are "in the cloud"—just a fancy name for disks owned by a big company such as Amazon and rented out to whoever needs space to store stuff. The disks can hold so many bits that there is no need to pick and choose what gets remembered.

"Bit" is shorthand for "binary digit." The binary number system uses just two digits, 0 and 1, instead of the ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 used in the decimal number system. The first clear statement of the principles of binary notation was given by Gottfried Wilhelm Leibniz in 1679.

So much digital information, misinformation, data, and garbage is being squirreled away that most of it will be seen only by computers, never by human eyes. And computers are getting better and better at extracting meaning from all those bits—finding patterns that sometimes solve crimes, diagnose diseases, and make useful suggestions—and sometimes reveal things about us we did not expect others to know.

The tale of Edward Snowden, who leaked thousands of highly secret government documents in 2013, is a bits story. He brought the documents out of the United States on his laptop; only a few years earlier, he would have needed to carry hundreds of pounds of paper. And everything he disclosed was about government electronic surveillance, raising fundamental questions about trade-offs between privacy and security.

The grounding of the 737 Max in 2019 was not just an airplane story. It was also a bits story. The engines of earlier model 737s had been moved, changing the airplane's weight distribution; software written to process

sensor data and automatically control the airplane's movements did not work as intended.[7]

But it is not just events of global significance that are bits stories; it's the day-to-day stories of ordinary life. The creepy experience of recreational runner Rosie Spinks is a bits story. Spinks used an app on her phone to keep track of her routes and times, and she thought her whereabouts were being kept secret because she had the app's so-called "Enhanced Privacy" setting on. Only when strangers started "liking" her workouts while she was traveling abroad did she realize that "Enhanced Privacy" actually meant "tell random men about my runs if I'm on the leader board." The fitness app was also a social network app, and Rosie's data was being commercialized.[8]

Once something is on a computer, it can replicate and move around the world in a heartbeat. Making a million perfect copies takes but an instant—copies of things we want everyone in the world to see and also copies of things that weren't meant to be copied at all.

The digital explosion is changing the world as much as printing once did—and some of the changes are catching us unaware, blowing to bits our assumptions about the way the world works.

The digital explosion can seem benign, amusing, or even utopian. Instead of sending prints through the mail to Grandma, we share pictures of our children on Instagram. Then not only can Grandma see them but so can Grandma's friends and anyone else. We enjoy the benefits, but what are the risks? The photos are cute and harmless. But suppose a tourist takes a vacation snapshot, and you just happen to appear in the background, at a restaurant where no one knew you were dining. If the tourist uploads his photo and makes it public, the whole world could know where you were and when you were there. Face recognition, which only a few years ago was beyond the capabilities of computers, is now good enough that the tourist photo could even get tagged with your name. It might not happen because a policy or a law prohibits it, but technological limitations won't. Identifying faces in crowds automatically is now a solved problem, and software to do this is being used in China and other authoritarian regimes to discourage public protests and generally to keep track of citizens' whereabouts. And this technology is being used in the United States, too: With the aid of billions of labeled photos gathered from Facebook and other social media, a small company named Clearview AI suddenly became a tool of many law enforcement agencies and even security-conscious private companies.[9] It wasn't even very hard to do—for an entrepreneurial company willing to stretch the limits of the appropriate use

of the massive photographic databases that Facebook and other companies had gathered.

And before we leave the topic of family photos—remember when they were all printed on paper and lasted for decades? Not so anymore. The wondrous benefits of digital images also make them inaccessible. You can't put digital images in a box under your bed for your grandchildren to find. All those family memories may well be lost in the future. There is a good and bad side to pretty much everything in the digital world.

Data leaks. Credit card records are supposed to stay locked up in a data warehouse, but they escape into the hands of identity thieves. And we give information away just because we get something back for doing so. A company will give you free phone calls to anywhere in the world—if you don't mind watching ads for the products its computers hear you talking about. Google will suggest restaurants you might like—if you will leave location tracking on so Google knows what restaurants you already frequent. If you have a meal, Google will ask you if you enjoyed it—no software is yet able to figure *that* out by itself—and into the data maw goes your answer to help Google make recommendations to you and others (and make a little money along the way).

And these are merely some of the things that are happening today. The explosion, and the social disruption it will create, have barely begun.

We already live in a world in which there is enough memory *just in cell phones* to store every word of every book in the Library of Congress billions of times over. Every day, enough video is uploaded to YouTube to record every moment of an entire human lifetime. The explosive growth is still happening. Every year we can store more information, move it more quickly, and do far more ingenious things with it than we could the year before. Now that refrigerators and vacuum cleaners create data, the increasing rate at which data is created is almost unimaginable. Most of the data that has ever existed was created in the past year—and that will be true again next year and the year after that.

So much disk storage is being produced every year that it could be used to record a page of information, every few seconds, about you *and every other human being on earth*. A remark made long ago can come back to haunt a political candidate, and a letter jotted quickly can be a key discovery for a biographer. Imagine what it would mean to record every word every human being speaks or writes in a lifetime. The technological barrier to that has already been removed: There is enough storage to remember it all. YouTube says that 500 hours of video are uploaded every minute.[10] Should any social barrier stand in the way?

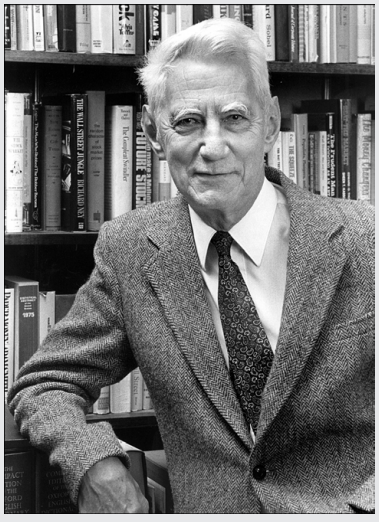
Sometimes things seem to work both better and worse than they used to. A "public record" is now *very* public. Before you get hired in Nashville,

Tennessee, your employer can figure out if you were caught ten years ago taking an illegal left turn in Lubbock, Texas. The old notion of a "sealed court record" is mostly a fantasy in a world where tidbits of information are duplicated, cataloged, and moved around endlessly. In Europe a new "right to be forgotten" has been added to the list of human rights, intended to protect people from having to carry every youthful indiscretion with them forever; but in the United States the right to free speech remains dominant, and the collision between these conflicting rights is inevitable. In the bits world, the Atlantic ocean can be crossed in microseconds.

With hundreds of TV and radio stations and millions of websites, Americans love the variety of news sources, but they have adjusted uncomfortably to the displacement of more authoritative sources. In China, the situation is reversed: The technology creates greater government control of the information its citizens receive, as well as better tools for monitoring their behavior.

The Koans of Bits

Bits behave strangely. They travel almost instantaneously, and they take almost no space to store. We have to use physical metaphors to make them understandable. We liken them to dynamite exploding or water flowing. We even use social metaphors for bits. We talk about two computers agreeing on some bits and about people using burglary tools to steal bits. Getting the right metaphor is important, but so is knowing the limitations of our metaphors. An imperfect metaphor can mislead as much as an apt metaphor can illuminate.



CLAUDE SHANNON

Claude Shannon (1916–2001) is the undisputed founding figure of information and communication theory. While working at Bell Telephone Laboratories after the Second World War, he wrote the seminal paper "A Mathematical Theory of Communication," which foreshadowed much of the subsequent development of digital technologies. Published in 1948, this paper gave birth to the now-universal realization that the bit is the natural unit of information and to the use of the term.

Reused with permission of Nokia Corporation and AT&T Archives. http:www.bell-labs.com/news/2001/february/26/shannon2_lg.jpeg.

We offer seven truths about bits. We call them "koans" because they are paradoxes, like the Zen verbal puzzles that provoke meditation and enlightenment. These koans are oversimplifications and overgeneralizations. They describe a world that is developing but hasn't yet fully emerged. But even today they are truer than we often realize. These themes will echo through our tales of the digital explosion.

Koan 1: It's All Just Bits

Your computer and your smartphone (really just another computer) successfully create the illusion that they contain photographs, letters, songs, and movies. All they really contain is bits—lots of them—patterned in ways you can't see. Your computer was designed to store just bits; all the files and folders and different kinds of data are illusions created by computer programmers. When you send a message containing a photograph, the computers that handle your message as it flows through the Internet have no idea that what they are handling is part text and part graphic. Telephone calls are also just bits, and that has helped create competition: Traditional phone companies, cell phone companies, cable TV companies, and voice over IP (VoIP) service providers can shuffle bits around to each other to complete calls. The Internet was designed to handle just bits, not emails or attachments, which are inventions of software engineers. We couldn't live without those more intuitive concepts, but they are artifices. Underneath, it's all just bits.

This koan is more consequential than you might think. Consider the story of Naral Pro-Choice America and Verizon Wireless. Naral wanted to form a text messaging group to send alerts to its members. Verizon decided not to allow it, citing the "controversial or unsavory" things the messages might contain.[11] Text message alert groups for political candidates it would allow, but not for political causes it deemed controversial. Had Naral simply wanted telephone service or an 800 number, Verizon would have had no choice. Telephone companies were long ago declared "common carriers." Like railroads, phone companies are legally prohibited from picking and choosing customers from among those who want their services. In the bits world, there is no difference between a text message and a wireless phone call. It's all just bits, traveling through the air by radio waves. But the law hasn't caught up to the technology. It doesn't treat all bits the same, and the common carriage rules for voice bits don't apply to text message bits.

Exclusive and Rivalrous

Economists would say that bits, unless controlled somehow, tend to be non-exclusive (once a few people have them, it is hard to keep them from others) and non-rivalrous (when someone gets them from me, I don't have any less). In a letter he wrote about the nature of ideas, Thomas Jefferson eloquently stated both properties: "If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it."[12]

Verizon backed down in the case of Naral, but not on the principle. A phone company can do whatever it thinks will maximize its profits in deciding whose messages to distribute. Yet no sensible engineering distinction can be drawn between text messages, phone calls, and any other bits traveling through the digital airwaves.

Koan 2: Perfection Is Normal

To err is human. When books were laboriously transcribed by hand in ancient scriptoria and medieval monasteries, errors crept in with every copy. Computers and networks work differently. Every copy is perfect. If you email a photograph to a friend, the friend won't receive a fuzzier version than the original. The copy will be identical, down to levels of detail too small for the eye to see.

Computers do fail, of course. Networks break down, too. If the power goes out with no battery backup, nothing works at all. So the statement that copies are normally perfect is only relatively true. Digital copies are perfect only to the extent that they can be communicated at all. And yes, it is possible in theory that a single bit of a big message will arrive incorrectly—but it's also possible that a volcano will erupt under you, and you won't get the message at all. The odds of an erroneous bit are lower than the odds of a physical catastrophe, and that is good enough for all practical purposes.

Networks don't just pass bits from one place to another. They check to see if the bits seem to have been damaged in transit and correct them or retransmit them if they seem incorrect. As a result of these error detection and correction mechanisms, the odds of an actual error—a character being wrong in an email, for example—are so low that we would be wiser to worry instead about a meteor hitting our computer, improbable though precision meteor strikes may be.

The phenomenon of perfect copies has drastically changed the law, a story told in Chapter 6, "Balance Toppled." In the days when music was distributed on audio tape, teenagers were not prosecuted for making copies of songs because the copies weren't as good as the originals, and copies of copies would be even worse. The reason that people today more often subscribe to music services than own their own copies of recordings is that copies are perfect—not just as good as the original but identical to the original so that even the notion of "original" is meaningless. The consequences of digital disruption of "intellectual property" are not over yet. Bits are an odd kind of property. Once I release them, everybody has them. And if I give you my bits, I don't have any fewer.

Koan 3: There Is Want in the Midst of Plenty

Vast as worldwide data storage is today, two years from now, it will be twice as large. Yet the information explosion means, paradoxically, the loss of information that is not online. One of us saw a new doctor at a clinic he had been using for decades. She showed him dense charts of his blood chemistry, data transferred from his home medical device to the clinic's computer—more data than any specialist could have had at her disposal five years ago. The doctor then asked whether he had ever had a stress test and what the test had shown. Those records should be all there, the patient explained, in the medical file. But the information was in the *paper* file, to which the doctor did not have access. It wasn't in the *computer's* memory, and the patient's memory was being used as a poor substitute. The old data might as well not have existed at all since it wasn't digital.

Even information that exists in digital form is useless if there are no devices to read it. The rapid progress of storage engineering has meant that data stored on obsolete devices effectively ceases to exist. A twentieth-century digital update of the eleventh-century British *Domesday Book* shown in Figure 1.1 was useless by the time it was only one-sixtieth the age of the original.

FIGURE 1.1 The Domesday Book of 1086. A 900th anniversary digital update was unreadable 15 years later.[13]

Or consider search, among the subjects of Chapter 4, "Gatekeepers." At first, search engines such as Google were interesting conveniences that a few people used for special purposes. With the growth of the World Wide Web and the explosion of online information, search engines became the first place many people look for information—even before they look in books or ask friends. Appearing prominently in search results has become a matter of life or death for businesses. We may move on to purchase from a competitor if we can't find the site we wanted in the first page or two of results. We may assume something didn't happen if we can't find it quickly in an online news source. If it can't be found—quickly—it's just as though it doesn't exist at all.

And some information isn't true. All of the mechanisms that enable the communication and storage of facts also work for falsehoods. Ugliness and cruelty are as easily captured in bits as beauty and kindness. The market economics of information change when everyone can be a publisher and no one needs an editor. Floods of misinformation, disinformation, and garbage can overwhelm truth and beauty. Authoritarian societies may be able to manage the flow of bits more efficiently than free societies, which risk being undercut by their own principles of information freedom.

Koan 4: Processing Is Power

Moore's Law

Gordon Moore, founder of Intel Corporation, observed that the density of integrated circuits seemed to double every couple of years. This observation is referred to as "Moore's Law." Of course, it is not a natural law, like the law of gravity. Instead, it is an empirical observation of the progress of engineering and a challenge to engineers to continue their innovation. In 1965, Moore predicted that this exponential growth would continue for quite some time.[14] That it has continued for more than 40 years is one of the great marvels of engineering. No other effort in history has sustained a growth rate anywhere close to this.

The speed of a computer is usually measured by the number of basic operations, such as additions, that can be performed in one second. The fastest computers available in the early 1940s could perform about five operations per second. The fastest today can perform about a trillion. Buyers of personal computers know that a machine that seems fast today will seem slow in a year or two.

For at least three decades, the increase in processor speeds was exponential. Computers became twice as fast every couple of years. These increases were one consequence of Moore's Law (see sidebar).

Since 2001, processor speed has not followed Moore's Law; in fact, processors have hardly grown faster at all. But that doesn't mean that computers

won't continue to get faster. New chip designs include multiple processors on the same chip so the work can be split up and performed in parallel. Such design innovations promise to achieve the same effect as continued increases in raw processor speed. And the same technology improvements that make computers faster also make them less expensive.

The rapid increase in processing power means that inventions move out of labs and into consumer goods very quickly. Robot vacuum cleaners and self-parking vehicles were possible in theory a decade ago, but now they have become consumer items. Tasks that today seem to require uniquely human skills are no longer just the subject of research projects in corporate or academic laboratories; they are incorporated in consumer products. Face recognition and voice recognition are here and now; telephones know who is calling, and surveillance cameras don't need humans to watch them. The power comes not just from the bits but from being able to do things with the bits.

Koan 5: More of the Same Can Be a Whole New Thing

The explosive growth is exponential growth—doubling at a steady rate. Imagine earning 100% annual interest on your savings account: In 10 years, your money would have increased more than a thousandfold, and in 20 years, more than a millionfold. A more reasonable interest rate of 5% will hit the same growth points, but it will do so 14 times more slowly. Epidemics initially spread exponentially, as each infected individual infects several others.

When something grows exponentially, for a long time it may seem not to be changing at all. If we don't watch it steadily, it will seem as though something discontinuous and radical occurred while we weren't looking.

That is why epidemics at first go unnoticed, no matter how catastrophic they may be when full-blown. Imagine 1 sick person infecting 2 healthy people, and the next day each of those 2 infects 2 others, and the next day after that each of those 4 infects 2 others, and so on. The number of newly infected each day grows from 2 to 4 to 8. In a week, 128 people come down with the disease in a single day, and twice that number are now sick, but in a population of 10 million, no one notices. Even after two weeks, barely 3 people in a 1,000 are sick. But after another week, 40% of the population is sick, and society collapses. The 2019–2020 coronavirus pandemic followed pretty much this pattern in parts of the world where societies did not react quickly. At the start of the epidemic in Wuhan, the number of cases doubled about every three days.[15]

Exponential growth is actually smooth and steady; it just takes very little time to pass from unnoticeable change to highly visible. Exponential growth of anything can suddenly make the world look utterly different than it had been. When that threshold is passed, changes that are "just" quantitative can look qualitative.

Another way of looking at the apparent abruptness of exponential growth—its explosive force—is to think about how little lead time we have to respond to it. Our hypothetical epidemic took three weeks to overwhelm the population. At what point was it only half as devastating? The answer is *not* "a week and a half." The answer is *on the next-to-last day*. Suppose it took a week to develop and administer a vaccine. Then noticing the epidemic after a week and a half would have left ample time to prevent the disaster. But that would have required understanding that there *was* an epidemic when only 2,000 people out of 10 million were infected.

The information story is full of examples of unperceived changes followed by dislocating explosions. Those with the foresight to notice the explosion just a little earlier than everyone else can reap huge benefits. Those who move a little too slowly may be overwhelmed by the time they try to respond. Take the case of digital photography.

In 1983, Christmas shoppers could buy digital cameras to hook up to their IBM PC and Apple II home computers. The potential was there for anyone to see; it was not hidden in secret corporate laboratories. But digital photography did not take off. Economically and practically, it couldn't. Cameras were too bulky to put in your pocket, and digital memories were too small to hold many images. Even 14 years later, film photography was still a robust industry. In early 1997, Kodak stock hit a record price, with a 22% increase in quarterly profit, "fueled by healthy film and paper sales...[and] its motion picture film business," according to a news report.[16] The company raised its dividend for the first time in eight years. But by 2007, digital memories had become huge, digital processors had become fast and compact, and both were cheap. As a result, cameras had become little computers. The company that was once synonymous with photography was a shadow of its former self. Kodak announced that its employee force would be cut to 30,000, barely a fifth the size it was during the good times of the late 1980s.[17] By 2018, that number was down to about 5,400. Bits took away 90% of the jobs. Moore's Law moved faster than Kodak did.

In the rapidly changing world of bits, it pays to notice even small changes—and to do something about them.

Koan 6: Nothing Goes Away

25,000,000,000,000,000,000.

That is the number of bits that were created and stored every day of 2019, according to one industry estimate. The capacity of disks has followed its own version of Moore's Law, doubling every two or three years. Far more data is created by every manner of device but not stored.

In financial industries, federal laws now *require* massive data retention to assist in audits and investigations of corruption. In many other businesses, economic competitiveness drives companies to save everything they collect and to seek out new data to retain. Tens of millions of transactions occur in Walmart stores every day, and every one of them is saved: date, time, item, store, price, who made the purchase, and how—credit, debit, cash, or gift card. Such data is so valuable to planning the supply chain that stores will pay money to get more of it from their customers. That is really what loyalty programs at supermarkets and other stores provide: Shoppers are supposed to think that the store is granting them a discount in appreciation for their steady business, but actually the store is paying them for information about their buying patterns. We might better think of a privacy tax: We pay the regular price *unless* we want to keep information about our food, alcohol, and pharmaceutical purchases from the market; to keep our habits to ourselves, we pay extra.

The massive databases challenge our expectations about what will happen to the data about us. Take something as simple as a stay in a hotel. When you check in, you are given a keycard, not a mechanical key. In fact, some hotels have gone one step further, having you use your own cell phone as the room key. Because the keycards can be deactivated instantly, there is no longer any great risk associated with losing your key, as long as you report it missing quickly. On the other hand, the hotel now has a record, accurate to the second, of every time you entered your room, used the gym or the business center, or went in the back door after-hours. The same database could identify every cocktail and steak you charged to the room, which other rooms you phoned and when, and the brands of tampons and laxatives you charged at the hotel's gift shop. This data might be merged with billions of bits' worth of other data, analyzed, and transferred to the parent company, which owns restaurants and fitness centers as well as hotels. It might also be lost, or stolen, or subpoenaed in a court case.

The ease of storing information has meant asking for more of it. Birth certificates used to include just the information about the child's and parents' names, birthplaces, and birthdates, plus the parents' occupations. Now the electronic birth record includes how much the mother drank and smoked during her pregnancy, whether she had genital herpes or a variety of other medical conditions, and both parents' Social Security numbers. Opportunities for research are plentiful, and so are opportunities for mischief and catastrophic accidental data loss.

The data will all be kept forever, unless there are policies to get rid of it.

And the data will all be kept forever unless there are policies to get rid of it. For the time being at least, the data sticks around. And because databases are intentionally

duplicated—backed up for security or shared while pursuing useful analyses—it is far from certain that data can ever be permanently expunged, even if we wish that to happen. The Internet consists of millions of interconnected computers; once data gets out, there is no getting it back. Victims of identity theft experience daily the distress of having to remove misinformation from the record. It seems never to go away.

Koan 7: Bits Move Faster Than Thought

The Internet existed before there were personal computers. It predates the fiber-optic communication cables that now hold it together. When it started around 1970, ARPANET, as it was called, was designed to connect a handful of university and military computers. No one imagined a network connecting tens of millions of computers and shipping information around the world in the blink of an eye. (In fact, no one imagined that so many computers would even exist.) Even the engineer who was charged with designing the gateways that would connect the computers together remembers his reaction to the idea of a computer network: "Looks like a straightforward engineering job; we could certainly do it, but I can't imagine why anyone would want such a thing."[18] Along with processing power and storage capacity, networking has experienced its own exponential growth in the number of computers interconnected and the rate at which data can be shipped over long distances, from space to earth and from service providers into private homes.

The Internet has caused drastic shifts in business practice. Customer service calls are outsourced to India today not just because labor costs are low there. Labor costs have *always* been low in India, but international telephone calls used to be expensive. Calls about airline reservations and lingerie returns are answered in India today because it now takes almost no time and costs almost no money to send to India the bits representing your voice. The same principle holds for professional services. When you are X-rayed at your local hospital in Iowa, the radiologist who reads the X-ray may be half a world away. The digital X-ray moves around the world and back faster than a physical X-ray could be moved between floors of a hospital. When you place an order at a drive-through station at a fast food restaurant, the person taking the order may be in another state. She keys the order so it appears on a computer screen in the kitchen, a few feet from your car, and you are none the wiser. Such developments are causing massive changes to the global economy, as industries figure out how to keep their workers in one place and ship their business as bits.

In the bits world, in which messages flow instantaneously, it sometimes seems that distance doesn't matter at all. The consequences can be startling. One of us, while dean of an American college, witnessed the shock of a father

receiving condolences on his daughter's death. The story was sad but familiar, except that this version had a startling twist. Father and daughter were both in Massachusetts, but the condolences arrived from halfway around the world before the father had learned that his daughter had died. News, even the most intimate news, travels fast in the bits world once it gets out.

When everyone is connected all the time, people can organize themselves as never before. Those afflicted with rare diseases or inspired by idiosyncratic interests can stroke a few keys and share their experiences, though they are separated by oceans and will never meet in person. And those united by a common cause can organize to air their grievances, as the digital-savvy youth of Hong Kong did with great skill during the pro-democracy protests of 2014. But in the hands of the authorities, the bits the protesters exchanged became evidence against them. By the time of the 2019 Hong Kong protests, organizers had abandoned Facebook and were resorting to less convenient encrypted texting apps—and were wearing face masks to confuse the government's face surveillance systems.[19]

And if surveillance fails, governments can simply shut down the Internet. That happened in the Muslim-majority Indian state of Kashmir in 2019; there was no Internet for 7 months, in the interest of "public safety."[20] Similar shutdowns happened in 2019 in Iran, Congo, Bangladesh, and more than a dozen other countries.[21] And in the United States, Section 706 of the Communications Act of 1934 authorizes the president to shut down "a facility for wire communication" in case of "a state or threat of war"—a very sweeping authorization, thus far never invoked to gain control of the Internet.

The instantaneous communication of massive amounts of information has created the misimpression that there is a place called "cyberspace," a land without frontiers where all the world's people can be interconnected as though they were residents of the same small town. That concept has been decisively refuted by the actions of the world's courts. National and state borders still count—and count a lot. If a book is bought online in England, the publisher and author are subject to British libel laws rather than those of the homeland of the author or publisher. Under British law, defendants have to prove their innocence; in the United States, plaintiffs have to prove the guilt of the defendants. An ugly downside to the explosion of digital information and its movement around the world is that information may become less available even where it would be legally protected. (We return to this subject in Chapter 7, "You Can't Say That on the Internet.") "Right to be forgotten" laws may require information to disappear—not just in the country where an individual has asked for some past misdeed to be struck from the electronic record but everywhere. Such a law might seem to be unenforceable, but the companies making the information available—Google, for example—operate internationally, and if they violate the law anywhere, they risk having

employees harassed or arrested any time they are within a jurisdiction where the law has been violated or ignored. Similarly, the publishing world has been blown to bits. It used to be possible to publish a bowdlerized edition of a book or an edited edition of a newspaper in countries with strict speech codes, but now the bits can readily flow from less censorious regions to more. It may prove simpler to publish only a single version of a work for sale everywhere, an edition omitting information that might somewhere excite a lawsuit.

Good and Ill, Promise and Peril

The digital explosion has thrown a lot of things up for grabs, and we all have a stake in who does the grabbing. The way the technology is offered to us, the way we use it, and the consequences of the vast dissemination of digital information are matters not in the hands of technology experts alone. Governments and corporations and universities and other social institutions have a say. And ordinary citizens, to whom these institutions are accountable, can influence their decisions. Important choices are made every year, in government offices and legislatures, in town meetings and police stations, in the corporate offices of banks and insurance companies, in the purchasing departments of chain stores and pharmacies. We can all help raise the level of discourse and understanding. We can all help ensure that technical decisions are made in a context of ethical standards.

We offer two basic morals. First, information technology is inherently neither good nor bad; it can be used for good or ill, to free us or to shackle us. Second, new technologies bring social change, and change comes with both risks and opportunities. All of us, and all of our public agencies and private institutions, have a say in whether technology will be used for good or ill and whether we will fall prey to its risks or prosper from the opportunities it creates.

Technology Is Neither Good nor Bad

Any technology can be used for good or ill; digital technologies, in particular, can be simultaneously good and bad. Nuclear reactions create electric power and weapons of mass destruction. These two uses share a common core but are otherwise quite distinct. Not so in the world after the digital explosion.

The same encryption technology that makes it possible for you to email your friends with confidence that no eavesdropper will be able to decipher your message also makes it possible for terrorists to plan their attacks undiscovered. The same Internet technology that facilitates the widespread distribution of educational works to impoverished students in remote locations

also enables massive copyright infringement. The photomanipulation tools that enhance your snapshots are used by child pornographers to escape prosecution.

The same technologies can be used to monitor individuals, to track their behaviors, and to control what information they receive. Search engines need not return unbiased results. Many users of web browsers do not realize that the sites they visit archive their actions. There is probably a record of exactly what you have been accessing and when, as you browse a clothing or book store catalog, a site selling pharmaceuticals, or a service offering advice on contraception or drug overdose. There are vast opportunities to use this information for invasive but relatively benign purposes, such as marketing, and also for more questionable purposes, such as blacklisting and blackmail.

The key to managing the ethical and moral consequences of technology while nourishing economic growth is to *regulate the use* of technology without *banning or restricting its creation*.

Few regulations mandate disclosure that the information is being collected or restrict the use to which the data can be put. The USA PATRIOT Act and other federal laws give government agencies sweeping authority to sift through mostly innocent data looking for signs of "suspicious activity" by potential terrorists—and to notice lesser transgressions in the process. Although the World Wide Web reaches into millions of households, the rules and regulations governing it are not much better than those of a lawless frontier town of the Old West.

New Technologies Bring Both Risks and Opportunities

The same large storage media that enable anyone to analyze millions of baseball statistics also allow anyone with access to confidential information to jeopardize its security. Access to aerial maps via the Internet makes it possible for criminals to plan burglaries of upscale houses, but technologically sophisticated police know that records of such queries can also be used to solve crimes.

Social networking tools such as Facebook and Twitter have made their founders quite wealthy and have given birth to many thousands of new friendships, marriages, and other ventures. But interconnectivity has unexpected side effects, too. A woman in England discovered that her fiancé was married when Facebook suggested his wife as someone she might want as a friend.[22] And in 2019 a Massachusetts college student committed suicide by jumping from the fourth floor of a parking garage, having received some 47,000 text messages, many allegedly abusive, from his girlfriend in the previous two months. She was charged with involuntary manslaughter—the same crime she might have been charged with had she instead struck him with her car while driving and texting.[23] In a nation deeply committed to free expression as a legal right, which Internet evils should be crimes, and which are just wrong?

Vast data networks have made it possible to move work to where the people are, not people to the work. The results are enormous business opportunities for entrepreneurs who take advantage of these technologies and new enterprises around the globe, and also the other side of the coin: jobs lost to outsourcing.

The difference every one of us can make, to our workplace or to another institution, can be to ask a question at the right time about the risks of some new technological innovation—or to point out the possibility of doing something in the near future that a few years ago would have been utterly impossible.

We begin our tour of the digital landscape with a look at our privacy, a social structure that the explosion has left in shambles. While we enjoy the benefits of ubiquitous information, we also sense the loss of the shelter that privacy once gave us. And we don't know what we want to build in its place. The good and ill of technology, and its promise and peril, are all thrown together when information about us is spread everywhere. In the post-privacy world, we stand exposed to the glare of noonday sunlight—and sometimes it feels strangely pleasant.

Endnotes

1 Drew Harwell, "A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job," *Washington Post*, November 6, 2019, https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/.

2 Scott Mayer McKinney et al., "International Evaluation of an AI System for Breast Cancer Screening," *Nature* 577, no. 7788 (January 2020): 89–94, https://doi.org/10.1038/s41586-019-1799-6.

3 Julia Angwin et al., "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

4 Elizabeth Fernandez, "Will Machine Learning Algorithms Erase the Progress of the Fair Housing Act?" *Forbes*, November 17, 2019, https://www.forbes.com/sites/fernandezelizabeth/2019/11/17/will-machine-learning-algorithms-erase-the-progress-of-the-fair-housing-act/.

5 Claudia Goldin and Cecilia Rouse, "Orchestrating Impartiality: The Impact of 'Blind' Auditions on Female Musicians," *The American Economic Review* 90, no. 4 (September 2000), https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.90.4.715.

6 "Algorithmic Transparency: End Secret Profiling," Electronic Privacy Information Center, March 1, 2020, https://epic.org/algorithmic-transparency/.

7 Benjamin Zhang, "The Boeing 737 Max Is Likely to Be the Last Version of the Best-Selling Airliner of All Time," *Business Insider*, March 19, 2019, https://www.businessinsider.com/boeing-737-max-design-pushed-to-limit-2019-3.

8 Rosie Spinks, "Confused About How to Use Strava Safely? You Are Not Alone," *Quartz*, January 29, 2018, https://qz.com/1191431/strava-privacy-concerns-here-is-how-to-safely-use-the-app/.

9 Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It," *The New York Times*, January 18, 2020, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html.

10 J. Clement, "Hours of Video Uploaded to YouTube Every Minute, 2007–2019," *Statista*, August 9, 2019, https://www.statista.com/statistics/259477/hours-of-video-uploaded-to-youtube-every-minute/.

11 Adam Liptak, "Verizon Blocks Messages of Abortion Rights Group," *The New York Times*, September 27, 2007, https://www.nytimes.com/2007/09/27/us/27verizon.html.

12 "Article 1, Section 8, Clause 8: Thomas Jefferson to Isaac McPherson," in Andrew A. Lipscomb and Albert Ellery Bergh, eds., *The Writings of Thomas Jefferson* (Thomas Jefferson Memorial Association, 1905), http://press-pubs.uchicago.edu/founders/documents/a1_8_8s12.html.

13 Robin McKie and Vanessa Thorpe, "Digital Domesday Book lasts 15 years not 1000," *Guardian Unlimited*, March 3, 2002.

14 G. E. Moore, "Cramming More Components onto Integrated Circuits," *Proceedings of the IEEE* 86, no. 1 (January 1998): 82–85, https://doi.org/10.1109/JPROC.1998.658762.

15 Steven Sanche et al., "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2," *Emerging Infectious Diseases* 26, no. 7 (July 2020): 1470–1477, https://dx.doi.org/10.3201/eid2607.200282.

16 "Kodak, GE, Digital Report Strong Quarterly Results," *Atlanta Constitution*, January 17, 1997.

17 Claudia H. Deutsch, "Shrinking Pains at Kodak," *The New York Times*, February 9, 2007.

18 Harry R. Lewis, "A Science Is Born," *Harvard Magazine*, September-October 2020: 42, https://harvardmagazine.com/2020/09/features-a-science-is-born/

19 Lily Kuo, "Hong Kong's Digital Battle: Tech That Helped Protesters Now Used Against Them," *The Guardian*, June 14, 2019, https://www.theguardian.com/world/2019/jun/14/hong-kongs-digital-battle-technology-that-helped-protesters-now-used-against-them.

20 Billy Perrigo, "India's Supreme Court Orders Review of Internet Shutdown in Kashmir. But for Now, It Continues," *Time*, January 10, 2020, https://time.com/5762751/internet-kashmir-supreme-court/.

21 Samuel Woodhams and Simon Migliano, "The Global Cost of Internet Shutdowns in 2019," Top10VPN, January 7, 2020, https://www.top10vpn.com/cost-of-internet-shutdowns/.

22 "Mum-of-Three Uncovered her Cheating Fiance's Double Life After His Wife Came Up as a Friend Suggestion on Facebook," *The Sun*, September 7, 2017, https://www.thesun.co.uk/fabulous/4411305/mum-of-three-uncovered-her-cheating-fiances-double-life-after-his-wife-came-up-as-a-friend-suggestion-on-facebook/.

23 Julia Jones, "Girlfriend Charged in Boston College Student's Death After Telling Him 'Hundreds of Times' to Kill Himself, prosecutors say," CNN, October 29, 2019, https://www.cnn.com/2019/10/28/us/boston-college-student-suicide-charges/index.html.

CHAPTER 2

Naked in the Sunlight

Privacy Lost, Privacy Abandoned

1984 Is Here, and We Like It

Fans attending Taylor Swift's packed Rose Bowl concert in the spring of 2018 saw her take the stage in a cloud of fog to sing hits from *Reputation*. As they entered or mingled between sets, some of those fans visited video kiosks to watch clips of the star's earlier performances and rehearsals, to get a behind-the-scenes glimpse of a favorite artist. What they didn't know was that the kiosk was watching them, too. The video booth was fitted with a camera that sent its visitors' images back to a "command post" in Nashville, where facial recognition software scanned them, reportedly looking for matches against a database of people who had stalked Swift in the past.[1] Were these images kept, or were they deleted securely? We don't know, just as we don't know how many other cameras capture us every day. Scanners like Swift's have been spotted at entrances to sports arenas, concert halls, and other entertainment venues. The public is often in the dark about their existence—and about policies related to how the images and other captured data are to be used, stored, or shared.

George Orwell's *1984* was published in 1948. Over subsequent years, the book became synonymous with a world of permanent surveillance, a society devoid of both privacy and freedom:

> ...there seemed to be no color in anything except the posters that were plastered everywhere. The black-mustachio'd face gazed down from every commanding corner. There was one on the house front immediately opposite. BIG BROTHER IS WATCHING YOU.[2]

The real 1984 came and went decades ago. Today, Big Brother's two-way telescreens would be amateurish toys today. Orwell's imagined London had cameras everywhere. His actual city now has at least half a million. Across the United Kingdom, there is one surveillance camera for every ten people.[3] The average Londoner is photographed hundreds of times a day by electronic eyes on the sides of buildings and on utility poles.

Yet there is much about the digital world that Orwell did not imagine. He did not anticipate that cameras are far from the most pervasive of today's tracking technologies. There are dozens of other kinds of data sources, and the data they produce is retained and analyzed. Cell phone companies know not only what numbers you call but where you have carried your phone. Credit card companies know not only where you spent your money but what you spent it on. Your friendly bank keeps electronic records of your transactions not only to keep your balance right but because it has to tell the government if you make huge withdrawals. When you go to a restaurant or a store, an app that has been quietly tracking your location asks you how you liked it, to feed your response into its recommendation-making engine.

The digital explosion has scattered the bits of our lives everywhere: records of the clothes we wear, the soaps we wash with, the streets we walk, and the cars we drive and where we drive them. And although Orwell's Big Brother had cameras, he didn't have search engines to piece the bits together, to find the needles in the haystacks. Wherever we go, we leave digital footprints, and computers of staggering capacity reconstruct our movements from those tracks. Computers reassemble the clues to form a comprehensive image of who we are, what we do, where we are doing it, and whom we are discussing it with.

Perhaps none of this would have surprised Orwell. Had he known about electronic miniaturization, he might have guessed that we would develop an astonishing array of tracking technologies. But there is something more fundamental that distinguishes the world of *1984* from the actual world of today. We have fallen in love with this always-on world. We accept our loss of privacy in exchange for efficiency, convenience, and small price discounts.

Attitudes have changed in the past decade. In a 2007 Pew/Internet Project report, 60% of Internet users were "not worried about how much information is available about them online," but by 2018, the ratio had flipped, and more than 60% "would like to do more to protect their privacy"; just 9% believe they have "a lot of control" over the information that is collected about them.[4] Although we're getting more worried about the loss of control over personal information, we're not sure there's much we can do about it.

In the world of bits, Big Brother has gotten both bigger and smaller. Technologically sophisticated nations like the United States and China have unprecedented ability to watch us, and they exercise that ability more often

than we might like. Companies do, too. They've built new businesses around ubiquitous data collection, much of it geared toward marketing directly to us. Commercial data also forms a rich lode for government to mine, a public–private surveillance partnership.

We, too, are a part of the surveillance networks, keeping tabs on ourselves and one another. We invite apps to track our movements and smart assistants to listen in to our conversations. We record our changes of mood and chitchat with friends, and we snap photos of friends and strangers. About seven in ten adults have created profiles on social networking websites. Yet most are dissatisfied with the level of control they have over what happens to the data they post there.[5]

There are hints that the privacy tide may be changing, that we're not willing to trade privacy for the benefits of the digital world. Regulators are giving us new protections (although often not from government surveillance), and companies are now marketing privacy as a feature.

Bits Cubed: The Snowden Files

When a 29-year-old Edward Snowden met with journalists in the lobby of Hong Kong's Mira Hotel in June 2013, he told them to look for the guy with a Rubik's Cube.[6] They eventually did so and got a trove of classified documents and PowerPoint presentations describing massive U.S. government communications surveillance: a series of front-page stories for the journalists. Snowden, as a systems administrator for the National Security Agency (NSA), extracted gigabytes of material, copying it to micro-SD cards smaller than the stickers on his Rubik's Cube.

The Snowden revelations fueled a series of front-page stories in the *New York Times*, *Washington Post*, and *Guardian* in 2013.[7] They showed the NSA engaged in pervasive communications surveillance—not just of foreigners and suspected terrorists but of law-abiding American citizens. If you used Yahoo! Mail or Google Search or dozens of other popular services, you were swept up in the dragnet. While the U.S. Constitution and laws make a sharp distinction between U.S. citizens and "foreign persons" that limit the government's ability to spy on its citizens, the bits carried no such distinction, and citizens ended up in the same buckets.

After the September 11, 2001, terrorist attacks, Congress passed new laws increasing spying powers. Notably, the USA PATRIOT Act authorized national security letters, which are secret demands for communications records; warrantless wiretaps of foreigners suspected of terrorist activity; and increasing ability to collect information on citizens any time obtaining foreign intelligence information is "a significant purpose" of the surveillance. Civil liberties groups expressed concern at the time that the act

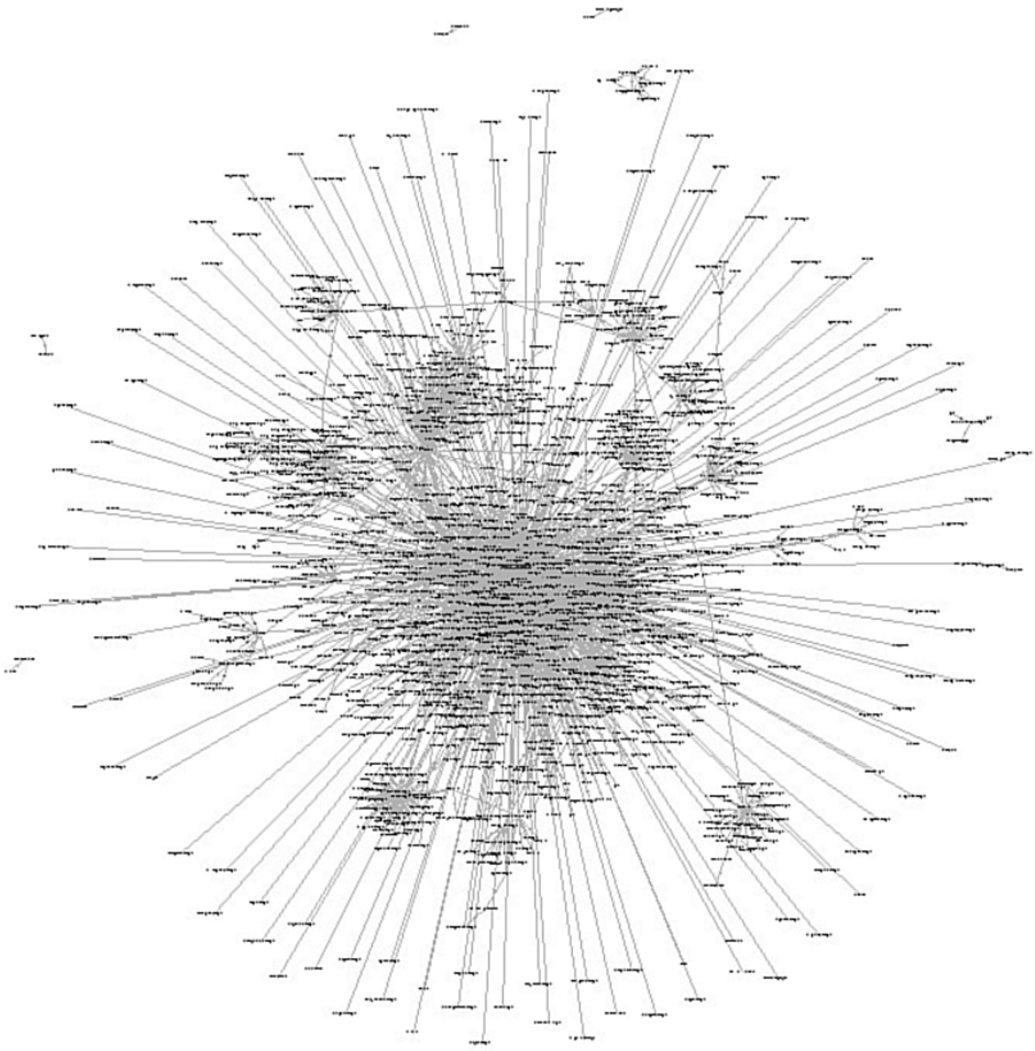
eliminated judicial checks and balances on surveillance,[8] but the act passed the Senate 98:1. Snowden's documents showed how far the NSA was pushing these new authorities.

The NSA exploited several properties of electronic communications. The popularity of centralized services for phone, email, search, and storage meant that taps at these corporate networks captured significant activity. The global nature of the Internet meant these taps could reach around the world from a few implant sites. A single request to Verizon for "business records" enabled the collection of millions of Americans' telephone call activity.[9] The Upstream program made full copies of everything carried along major domestic fiber-optic cable networks. Other top-secret warrantless data collection tools included XKEYSCORE[10] and EGOTISTICALGIRAFFE.[11]

U.S. officials defending the programs said they were only collecting metadata, not the contents of communications—the envelopes and addresses, not the letters inside. However, the web of contacts itself is tremendously informative. "We kill people based on metadata," said General Michael Hayden, former director of the NSA and the CIA.[12] William Binney, another ex-NSA whistleblower, left after the agency cut a program to conduct privacy-preserving searches.

The way we leave fingerprints and footprints is only part of what is new. We have always left a trail of information behind us—in our tax records, hotel reservations, and long-distance telephone bills. True, the footprints are far clearer and more complete today than ever before. But something else has changed: the harnessing of computing power to correlate data, to connect the dots, to put pieces together, and to create cohesive, detailed pictures from what would otherwise be meaningless fragments. The digital explosion does not just blow things apart. Like the explosion at the core of an atomic bomb, it blows things together as well. Gather up the details, connect the dots, and assemble the parts of the puzzle, and a clear picture will emerge.

Computers can sort through databases too massive and too boring to be examined with human eyes. They can assemble colorful pointillist paintings out of millions of tiny dots when any few dots would reveal nothing. When a federal court released half a million Enron emails obtained during the corruption trial, computer scientists quickly identified the subcommunities, and perhaps conspiracies, among Enron employees, using no data other than the pattern of who was emailing whom (see Figure 2.1). The same kinds of clustering algorithms work on patterns of telephone calls. You can learn a lot by knowing who is calling or emailing whom, even if you don't know what they are saying to each other—especially if you know the times of the communications and can correlate them with the times of other events.



Source: Enron, Jeffrey Heer, Figure 3 from http://jheer.org/enron/v1/

FIGURE 2.1 Diagram showing clusters of Enron emailers, indicating which employees carried on heavy correspondence with which others. The evident "blobs" may be the outlines of conspiratorial cliques.

The tale of Snowden and the NSA is two bits stories at once. Digital communication made it possible for the NSA to collect vast quantities of information, millions and millions of calls and emails, from just a few locations—something that would have been impossible if we were still communicating with regular phones and paper letters. And when Snowden took copies of everything, he could fit the equivalent of thousands of file cabinets of information into his pocket.

What can we do in the face of such government-directed surveillance? Snowden chose exposure, aiming for his disclosures to help support lawsuits against the programs and public pressure on lawmakers to rein in the NSA.

When he finished opening his Pandora's memory card of documents, he left us with a cause for hope: Math works. The NSA may have the world's best cryptographers and cryptanalysts, but the fundamental mathematics of encryption are still effective. The years since Snowden's disclosures have seen a dramatic increase in the use of encryption in basic Internet and web protocols[13] and in the applications that run on them.[14] End-to-end encryption enables us to reclaim some of the privacy that pervasive monitoring of unencrypted traffic unraveled.

"Reasonable Expectations of Privacy" Technology and the Fourth Amendment

Technological change has stood in tension with privacy before. When the Supreme Court first encountered the telephone wiretap in 1928, the president did not yet have a phone on his desk, although traffickers in illegal liquor (this was during Prohibition) had found the technology, and law enforcement wanted to listen in.[15] When the bootleggers challenged the tapping of their phone lines—alligator clips on physical wires outside homes and offices—the Court's majority put the telephone, which was high-tech at the time, in a frame they recognized, of physical intrusion and trespass. Without trespass, the Court held, there was no "search" or "seizure" and therefore no need for a warrant:

> The reasonable view is that one who installs in his house a telephone instrument with connecting wires intends to project his voice to those quite outside, that the wires beyond his house, and messages while passing over them, are not within the protection of the Fourth Amendment. Here those who intercepted the projected voices were not in the house of either party to the conversation.

Justice Brandeis, who did not agree, wrote in his dissent:

> Whenever a telephone line is tapped, the privacy of the persons at both ends of the line is invaded, and all conversations between them upon any subject, and although proper, confidential, and privileged, may be overheard. Moreover, the tapping of one man's telephone line involves the tapping of the telephone of every other person whom he may call, or who may call him."

But he was in the minority; for decades, warrantless wiretapping was lawful.

The Court's ruling in *Olmstead v. United States* increased the vulnerability of telephonic communications to police snooping, but it also publicly exposed

that lack of privacy. Criminals, judges, and the general public learned that their conversations were liable to be tapped. As the telephone itself became more widely used, the legal rule triggered responses. States passed wiretap acts to protect by statute what the Constitution would not, and in 1934, Congress included anti-interception prohibitions in The Communications Act, section 605.

When Charles Katz came before the Court in 1967 to challenge the wiretapping of his (illegal wagering) conversation from a public telephone booth, the times, technologies, and legal norms had all changed. The telephone was part of everyday life, for personal and intimate communications as well as businesses both lawful and unlawful. The public and the justices themselves had experience to color their views of the technology. Asked again "whether a public telephone booth is a constitutionally protected area," the Court said that was the wrong formulation: It wasn't place but context. Telephone calls now demanded greater protection, even when conducted from the relative publicity of a glass-walled "public" phone booth. Justice Harlan, concurring in the judgment throwing out Katz's wiretap, articulated the test that still defines the Fourth Amendment's privacy protection: a "twofold requirement, first that a person have exhibited an actual (subjective) expectation of privacy and, second, that the expectation be one that society is prepared to recognize as 'reasonable.'"[16]

Location, Location, Location

Buy a navigation-equipped car, and it will listen to precisely timed signals from satellites reporting their positions in space. The Global Positioning System (GPS) calculates locations based on the satellites' locations and the times their signals are received. The 24 satellites spinning 12,500 miles above the earth enable your car to locate itself within 25 feet, at a price so low that most new cars have it as a standard feature. What was once a military secret now comes free in every smartphone.

If you carry a GPS-enabled cell phone, your friends can find you if that is what you want. If your GPS-enabled rental car has a radio transmitter, you can be found whether you want to be or not. Car leasing companies are adding transponders, including auto-immobilizers, to enable remote repossession, without even sending a repo man to the site. Those who fall behind in their car payments may suddenly find themselves unable to get to or from work.

GPS enables you to determine your location anywhere on earth, and even a low-end cell phone serves as a rudimentary positioning system. If you are traveling in settled territory—anyplace where you can get cell coverage—you

move from the range of one cell tower to the range others, pinging the towers as you go. Triangulation among these signals can be used to locate you. The location is less precise than that supplied by GPS—only within ten city blocks or so—but the fact that it is possible at all means that a pattern of your behavior can be built, or photos can be stamped with identifying information about where they were shot, as well as when and with what camera.

Timothy Carpenter was given away by the bits from his cell phone, and then he was given a second chance by the law. A string of robberies of Radio Shack and T-Mobile stores in the Detroit area led to the arrest of four men in April 2011.[17] One of them confessed and gave the FBI the cell phone numbers of his accomplices; he also let law enforcement collect recently called numbers from his phone. With that evidence, prosecutors obtained an order directing wireless phone carriers to disclose information and cell-site location histories on the called numbers. They concluded that a phone registered to Timothy Carpenter had been near four store locations at the times when the stores were robbed. Carpenter was taken to court, and at trial, several confederates testified that he had been the leader of the robbery operation. With the corroboration of the cell-site mapping data, he was sentenced to more than 100 years in prison.

Carpenter appealed his case to the U.S. Supreme Court, arguing that the use of cell-site location data amounted to a "search,"[18] which could be conducted only with a warrant based on probable cause—not the mere order prosecutors had used to obtain records from the wireless carriers.

In 2018, the Court agreed with Carpenter: Because "cell phone location information is detailed, encyclopedic, and effortlessly compiled," the equivalent of a long-term digital "tail," individuals should have a reasonable expectation of privacy that their location history will not be exposed without a search warrant. Just because we're all carrying detailed location trackers, and those devices locate us to third parties in order to function, doesn't mean law enforcement gets automatic access to our map history. As it had earlier in *Katz*, the Court said that new technological capabilities shouldn't upend the balance between law enforcement and the public. Bits might trace our every move, but police need the judicial oversight of a search warrant to see them.

Lawyers and technologists argue about the balance between their disciplines. Post-Snowden revelations, they wonder whether we can trust the government to check its own power, or if records supposed to be available only by application to a neutral magistrate for a search warrant will instead be handed over or gathered in bulk. In 2013, after Edward Snowden revealed the existence of a secret, and extensive, data collection program code-named "PRISM," the NSA argued that records aren't effectively "collected" until they are searched, even once they are gathered in data banks. But while you can

encrypt your conversations, it's much harder to hide the metadata of your digital footprints. (Tor Project's onion routing, https://www.torproject.org, is the best option.) To protect the privacy of those activities that must be public to be effective or that depend on interactions with others we don't necessarily trust to keep our secrets, we need the force of law and social norms.

Black Boxes: Not Just for Airplanes Anymore

On April 12, 2007, John Corzine, governor of New Jersey, was heading back to the governor's mansion in Princeton to mediate a discussion between Don Imus, the controversial radio personality, and the Rutgers University women's basketball team.[19]

His driver, 34-year-old state trooper Robert Rasinski, headed north on the Garden State Parkway. He swerved to avoid another car and flipped the governor's Chevy Suburban. Governor Corzine, who had not fastened his seatbelt, broke 12 ribs, a femur, his collarbone, and his sternum. The details of exactly what happened were unclear. When questioned, Trooper Rasinski said he was not sure how fast they were going—but we *do* know. He was going 91 in a 65-mile-per-hour zone. There were no police with radar guns around; no human was tracking his speed. We know his exact speed at the moment of impact because his car, like 30 million other cars in America, had a black box—an event data recorder (EDR) that captured every detail about what was going on just before the crash. An EDR is an automotive "black box" like the ones recovered from airplane crashes.

EDRs started appearing in cars around 1995, and they now appear in almost all models. Your insurance company is probably entitled to its data if you have an accident. Yet most people do not realize that EDRs exist, unless they've gotten an offer from their insurance company to give up real-time data rather than pay higher premiums.

EDRs capture information about speed, braking time, turn signal status, seat belts: information needed for accident reconstruction, to establish responsibility, or to prove innocence. CSX Railroad was exonerated of all liability in the death of the occupants of a car when its EDR showed that the car was stopped on the train tracks when it was hit. Police generally obtain search warrants before downloading EDR data—but not always; in some cases, they do not have to. When Robert Christmann struck and killed a pedestrian on October 18, 2003, Trooper Robert Frost of the New York State Police downloaded data from the car at the accident scene. The EDR revealed that Christmann had been going 38 miles per hour in an area where the speed limit was 30. When the data was introduced at trial, Christmann claimed that the state had violated his Fourth Amendment rights because it had not asked his permission or obtained a search warrant before retrieving the data. That

It is almost as hard to avoid leaving digital footprints as it is to avoid touching the ground when we walk.

was not necessary, ruled a New York court. Taking bits from the car was not like taking something out of a house, and no search warrant was necessary.[20]

Bits mediate our daily lives. It is almost as hard to avoid leaving digital footprints as it is to avoid touching the ground when we walk, and even if we lived our lives without walking, we would unsuspectingly be leaving fingerprints.

SEARCHING LOCATION

Download your location history from Google or Facebook and look at the picture it paints. Does anything there make you nervous (whether it should or not)? What would you have difficulty explaining? Have you ever changed the settings from their account defaults? Should you?

Saving Time: Electronic Tolling and License Plate Readers

For commuters who use toll roads or bridges, the risk–reward calculation is not even close. Time is money, and time spent waiting in a car also means anxiety and frustration. If there is an option to get a toll booth transponder, many commuters will get one, even if the device costs a few dollars up front. Cruising past the cars waiting to pay with dollar bills is not just a relief; it actually brings the driver a certain satisfied glow.

The transponder, which the driver attaches to the windshield inside the car, is an RFID-enabled device powered with a battery that sends information to a sensor several feet away as the driver whizzes past. The sensor can be mounted in a constricted travel lane, where a toll booth for a human toll taker might have been. Or it can be mounted on a boom above traffic so the driver doesn't even need to change lanes or slow down. And what is the possible harm? Of course, the state is recording the fact that the car has passed the sensor; that is how the proper account balance can be debited to pay the toll. When the balance gets too low, the driver's credit card may get billed automatically to replenish the balance, which only makes the system better—no fumbling for change or doing anything else to pay for your travels.

The monthly bill—for the Massachusetts Fast Lane, for example—shows where and when you got on the highway, accurate to the second. It also shows how far you traveled on the highway and where you got off. Fast Lane also informs you of the mileage, which is another useful service because Massachusetts drivers can get a refund on certain fuel taxes if the fuel was used on the state toll road. Of course, you do not need a PhD to

figure out that the state also knows when you got off the road, to the second, and that with one subtraction and one division, its computers could figure out if you were speeding. Technically, in fact, it would be trivial for the state to print the appropriate speeding fine at the bottom of the statement and to bill your credit card for that amount at the same time it charges you for the tolls. That would be taking convenience a bit too far, and no state does it—yet.

What does happen right now, however, is that toll transponder records are introduced into divorce and child custody cases. You've never been within 5 miles of that lady's house? Really? Why have you gotten off the highway at the exit near it so many times? You say you can be the better custodial parent for your children, but the facts suggest otherwise. As one lawyer put it, "When a guy says, 'Oh, I'm home every day at 5, and I have dinner with my kids every single night,' you subpoena his E-ZPass and you find out he's crossing that bridge every night at 8:30. Oops!" Such records have been subpoenaed hundreds of times in family law cases. They have also been used in employment cases, to prove that the car of a worker who said he was working was actually far from the workplace.

But most of us aren't planning to cheat on our spouses or our bosses, so the loss of privacy seems like no loss at all—at least compared to the time saved. Of course, if we actually *were* cheating, we *would* be in a big hurry and might take some risks to save a few minutes!

Massachusetts toll roads eliminated toll takers in 2017. Drivers can save some money by equipping their cars with transponders, but if they don't have a transponder, never fear: "We will bill you," the state announces in billboards along the highway. There are no cash lanes now. Gantries equipped with both transponder antennas and automated license plate readers record every car or truck that passes through. To stay anonymous, you need to take the slow road.

The License Plate Tells More Than You Think

In June 2018, southern California mall operator Irvine Company was found to be collecting the license plate numbers of vehicles entering its parking areas. When a visitor, 14-year-old Zoe Wheatcroft, dug deeper into the company's "privacy policy," she found that Irvine was not only collecting license plate information but sharing it with law enforcement, in a database that might be accessed by agents from Immigration and Customs Enforcement (ICE).[21] When word got out, Irvine and Vigilant, the database company, responded that their policy was in fact narrower and more restrictive but gave customers no way to know that a shopping trip wouldn't put them in surveillance crosshairs.

Automated license plate recognition is a form of mass surveillance enabled by cheaper and more sophisticated cameras, software, and network capabilities. Automatic cameras capture images of license plates, convert the plate numbers into plaintext characters, and annotate the images with time, date, and GPS-derived location before transmitting and storing each instance. The data stream may be queried in real time, as in a search for a wanted criminal or stolen vehicle, or it may be retrieved later to give a picture of shoppers' demographics or a particular shopper's travel pattern.

Loose Fitbits Sink Ships?

The Strava fitness-mapping application offers a connection to users' GPS-enabled smartphones, watches, and Fitbit devices in order to enable athletes to track their runs, cycle routes, and other activities. Strava combined the data into a "heatmap" visualization, aggregating more than a billion activity logs into colored streaks across a map. While the Strava team highlighted a few recreation images on their blog—the Ironman triathlon swim off Hawaii, mountain biking in Whistler, British Columbia—a researcher, noting what appeared to be the outlines of military bases in Afghanistan, posted screenshots to Twitter and reminded people "turning off data sharing is an option."[22] Strava's CEO followed up with a blog post pointing to explanations of the privacy settings and promising to work with military and government officials "to address potentially sensitive data."[23]

Of course, one can say that soldiers in sensitive locations should turn off their location reporting—which means they need to know that their devices and applications have that setting and consider its consequences. But the Strava heatmap may be only the most visible and most easily changed of the places we leave these trails. Cell phones build location maps as they ping nearby towers; frequently accessed websites have logs of the IP addresses from which they are viewed (from which the site operator can map corresponding geolocation); and many mobile apps collect location information to target advertising. Individual data points may seem harmless, but points gathered over time and space can paint a detailed picture of travel patterns or home life—and even secret military strategy.

Big Brother, Abroad and in the United States

Big Brother really is watching today, and his job has gotten much easier, thanks to the digital explosion. In China, which has a long history of tracking individuals as a mechanism of social control, the millions of residents of Shenzhen are being issued identity cards, which record far more than the

bearer's name and address. According to a report in the *New York Times*,[24] the cards document the individual's work history, educational background, religion, ethnicity, police record, medical insurance status, landlord's phone number, and reproductive history. Touted as a crime-fighting measure, the new technology—developed by an American company—will come in handy in dealing with cases of street protests and individual activities deemed suspicious by the authorities. The sort of record keeping that used to be the responsibility of local authorities is becoming automated and nationalized as the country prospers and its citizens become increasingly mobile. The technology makes it easier to know where everyone is, and the government is taking advantage of that opportunity. In Xinjiang, where the Uighur minority faces especially strict scrutiny, police have an app that can flag when someone has stopped using a smartphone or avoids the front door. Facial recognition is targeted at Uighurs, who are made to pass through checkpoints that Han (the ethnic majority elsewhere in China) are permitted to avoid. Chinese tracking is far more detailed and pervasive than Britain's system of ubiquitous surveillance cameras.

Identifying Citizens—Without ID Cards

In the age of global terrorism, democratic nations are resorting to digital surveillance to protect themselves, creating hotly contested conflicts with traditions of individual liberty. In the United States, the idea of a national identification card prompts a furious libertarian reaction from parties not usually outspoken in defense of individual freedom. Under the REAL ID Act of 2005, uniform federal standards were to be implemented for state-issued driver's licenses. Although it passed through Congress without debate, the law is opposed by at least 18 states. Resistance pushed back the implementation timetable multiple times. In 2018, 13 years later, only 37 states met the REAL ID rules. Finally, in 2019, states were told their final extension would expire, and only REAL ID–compliant documents would be accepted for federal identification by October 2020. Then COVID-19 hit, and the deadline was extended yet again. Yet even fully implemented, REAL ID would fall far short of the true national ID preferred by those charged with fighting crime and preventing terrorism.

As the national ID card debate continues in the United States, the FBI is making it irrelevant by exploiting emerging technologies. There would be no need for anyone to carry an ID card if the government had enough biometric data on Americans—that is, detailed records of their fingerprints, irises, voices, walking gaits, facial features, scars, and earlobe shapes. Gather a combination of measurements on individuals walking in public places, consult the databases, connect the dots, and—bingo!—their names pop up on the computer screen. No need for them to carry ID cards; the combination of biometric data would pin them down perfectly.

As the national ID card debate continues in the United States, the FBI is making it irrelevant by exploiting emerging technologies.

Well, only imperfectly at this point, but the technology is improving. And the data is already being gathered and deposited in the data vault of the FBI's Criminal Justice Information Services database in Clarksburg, West Virginia. The database already holds some 75 million sets of fingerprints, and the FBI processes 100,000 requests for matches every day. Any of 900,000 federal, state, and local law enforcement officers can send a set of prints and ask the FBI to identify it. If a match comes up, the individual's criminal history can be accessed in the database, too.

But fingerprint data is hard to gather; mostly it is obtained when people are arrested. The goal of the project is to get identifying information on nearly everyone—and to get it without bothering people too much. For example, a simple notice at airport security could advise travelers that, as they pass through airport security, a detailed "snapshot" will be taken as they enter the secure area. The traveler would then know what is happening and could have refused (and stayed home). As an electronic identification researcher puts it, "That's the key. You've chosen it. You have chosen to say, 'Yeah, I want this place to recognize me.'"[25] The project eliminates the issue of REAL ID controversies, as all the data being gathered is, in some sense at least, offered voluntarily.

San Francisco, California, the epicenter of the technology boom, moved in the other direction, banning law enforcement use of facial recognition technology.[26] The Board of Supervisors heard concerns that the technologies were biased, lacked transparency, and could be abused by government. At the same time, however, data-based identification flourishes in private hands. The company PatronScan asserts that its database of IDs swiped at bars contains more than 60 million IDs across 200 cities. PatronScan not only checks that bar-goers are of legal drinking age but maintains a blacklist of patrons flagged for "bad behavior."[27]

Friendly Cooperation Between Big Siblings

In fact, there are two Big Brothers, and they often work together. We are, by and large, glad they are watching—if we are aware of it at all. Only occasionally are we alarmed about their partnership.

The first Big Brother is Orwell's—the government. And the other Big Brother is the industry about which most of us know very little: the business of aggregating, consolidating, analyzing, and reporting on the billions of individual transactions, financial and otherwise, that take place electronically every day. Of course, the commercial data aggregation companies are not in the spying

business; none of their data reaches them illicitly. But they do know a lot about us, and what they know can be extremely valuable, both to businesses and to the government.

The new threat to privacy is that computers can extract significant information from billions of apparently uninteresting pieces of data, in the way that mining technology has made it economically feasible to extract precious metals from low-grade ore. Computers can correlate databases on a massive level, linking government data sources together with private and commercial ones to create comprehensive digital dossiers on millions of people. With their massive data storage and processing power, they can make connections in the data, by using brute force rather than ingenuity. And the computers can discern even very faint traces in the data—traces that may help track payments to terrorists, set insurance rates, or simply help us make sure our new babysitter is not a sex offender.

And so we turn to the story of the government and the aggregators.

Acxiom is the country's biggest customer data company. Its business is to aggregate transaction data from all those swipes of cards in card readers all over the world. This amounted to more than a hundred billion transactions in 2018.[28] The company uses its massive amounts of data about financial activity to support the credit card industry, banks, insurers, and other consumers of information about how people spend money. Unsurprisingly, after the War on Terror began, the Pentagon also got interested in Acxiom's data and the ways the company gathers and analyzes it. Tracking how money gets to terrorists might help find the terrorists and prevent some of their attacks.

ChoicePoint is the other major U.S. data aggregator. ChoicePoint has more than 100,000 clients, which call on it for help in screening employment candidates, for example, or determining whether individuals are good insurance risks.

Acxiom and ChoicePoint are different from older data analysis operations in the scale of their operations. Quantitative differences have qualitative effects, as we said in Chapter 1; what has changed is not the technology but rather the existence of rich data sources. Forty years ago, credit cards had no magnetic stripes. Charging a purchase was a mechanical operation; the raised numerals on the card made an impression through carbon paper so you could have a receipt, and the top copy went to the company that issued the card. Today, if you charge something using your CapitalOne card, the bits go instantly not only to CapitalOne but to Acxiom and other aggregators. The ability to search through huge commercial data sources—including not just credit card transaction data but phone call records, travel tickets, and banking transactions, for example—is another illustration that more of the same can create something new.

Privacy laws do exist, of course. For a bank, or a data aggregator, to post your financial data on its website would be illegal. But privacy is still developing as an area of the law, and it is connected to commercial and government interests in uncertain and surprising ways.

A critical development in privacy law was precipitated by the presidency of Richard Nixon. In what is generally agreed to be an egregious abuse of presidential power, Nixon used his authority as president to gather information on those who opposed him—in the words of his White House counsel at the time, to "use the available federal machinery to screw our political enemies." Among the tactics Nixon used was to have the Internal Revenue Service audit the tax returns of individuals on an "enemies list," which included members of Congress, journalists, and major contributors to Democratic causes. Outrageous as it was to use the IRS for this purpose, it was not illegal, so Congress moved to ban it in the future.

The Privacy Act of 1974 established broad guidelines for when and how the federal government can assemble dossiers on citizens it is not investigating for crimes. The government has to give public notice about what information it wants to collect and why, and it has to use what it collects only for those reasons.

The Privacy Act limits what the government can do to gather information about individuals and what it can do with records it holds. Specifically, it states, "No agency shall disclose any record which is contained in a system of records by any means of communication to any person, or to another agency, except pursuant to a written request by, or with the prior written consent of, the individual to whom the record pertains, unless...." If the government releases information inappropriately, even to another government agency, the affected citizen can sue for damages in civil court. The protections provided by the Privacy Act are sweeping—although not as sweeping as they may seem. Not every government office is in an "agency"; the courts are not, for example. The act requires agencies to give public notice of the uses to which they will put the information, but the notice can be buried in the Federal Register, where the public probably won't see it unless news media happen to report it. Then there is the "unless" clause, which includes significant exclusions. For example, the law does not apply to disclosures for statistical, archival, or historical purposes; civil or criminal law enforcement activities; Congressional investigations; or valid Freedom of Information Act requests.

Despite the act's exclusions, government practices changed significantly because of this law. Then, a quarter century later, came 9/11. *Law enforcement should have seen it all coming*, was the constant refrain as investigations revealed how many unconnected dots were in the hands of different government agencies. *It all could have been prevented if the investigative fiefdoms*

had been talking to each other. They should have been able to connect the dots. But they could not—in part because the Privacy Act restricted interagency data transfers. A response was badly needed. The Department of Homeland Security was created to ease some of the interagency communication problems, but that government reorganization was only a start.

In January 2002, just a few months after the World Trade Center attack, the Defense Advanced Research Projects Agency (DARPA) established the Information Awareness Office (IAO) with a mission to:

> imagine, develop, apply, integrate, demonstrate, and transition information technologies, components and prototype, closed-loop, information systems that will counter asymmetric threats by achieving total information awareness useful for preemption; national security warning; and national security decision making. The most serious asymmetric threat facing the United States is terrorism, a threat characterized by collections of people loosely organized in shadowy networks that are difficult to identify and define. IAO plans to develop technology that will allow understanding of the intent of these networks, their plans, and potentially define opportunities for disrupting or eliminating the threats. To effectively and efficiently carry this out, we must promote sharing, collaborating, and reasoning to convert nebulous data to knowledge and actionable options.

Vice Admiral John Poindexter directed the effort that came to be known as "Total Information Awareness" (TIA). The growth of enormous private data repositories provided a convenient way to avoid many of the prohibitions of the Privacy Act. The Department of Defense can't get data from the Internal Revenue Service because of the 1974 Privacy Act. *But the government can buy the very same data it is barred from collecting from private data aggregators!* In a May 2002 email to Adm. Poindexter, Lt. Col Doug Dyer discussed negotiations with Acxiom:

> Acxiom's Jennifer Barrett is a lawyer and chief privacy officer. She's testified before Congress and offered to provide help. One of the key suggestions she made is that people will object to Big Brother, wide-coverage databases, but they don't object to use of relevant data for specific purposes that we can all agree on. Rather than getting all the data for any purpose, we should start with the goal, tracking terrorists to avoid attacks, and then identify the data needed (although we can't define all of this, we can say that our templates and models of terrorists are good places to start). Already, this guidance has shaped my thinking.

> Ultimately, the U.S. may need huge databases of commercial transactions that cover the world or certain areas outside the U.S. This information provides economic utility, and thus provides two reasons why foreign countries would be interested. Acxiom could build this mega-scale database.

The *New York Times* broke the story in October 2002. As Poindexter had explained in speeches, the government had to "break down the stovepipes" separating agencies and get more sophisticated about how to create a big picture out of a million details, no one of which might be meaningful in itself. The *Times* story set off a sequence of reactions from the Electronic Privacy Information Center and civil libertarians. Congress defunded the office in 2003—but that was not the end of the idea.

The key to TIA was data mining: looking for connections across disparate data repositories, finding patterns, or "signatures," that might identify terrorists or other undesirables. The General Accountability Office report on Data Mining (GAO-04-548) reported on a survey of 128 federal departments.[29] It described 199 separate data mining efforts, of which 122 used personal information.

Although IAO and TIA went away, Project ADVISE at the Department of Homeland Security continued with large-scale profiling system development. Eventually, Congress demanded that the privacy issues concerning this program be reviewed as well. In his June 2007 report (OIG-07-56), Richard Skinner, the DHS inspector general, stated that "program managers did not address privacy impacts before implementing three pilot initiatives," and a few weeks later, the project was shut down. But ADVISE was only one of a dozen data-mining projects going on in DHS at the time.

Similar privacy concerns led to the cancellation of the Pentagon's TALON database project. That project sought to compile a database of reports of suspected threats to defense facilities as part of a larger program of domestic counterintelligence.

Despite these privacy concerns, as Edward Snowden revealed, many surveillance and data mining programs simply carried on under the radar.

The government creates projects, the media and civil liberties groups raise serious privacy concerns, the projects are canceled, and new projects arise to take their place. The cycle seems to be endless. In spite of Americans' traditional suspicions about government surveillance of their private lives, the cycle seems to be almost an inevitable consequence of Americans' concerns about their security and the responsibility that government officials feel to use the best available technologies to protect the nation. Corporate databases often contain the best information on the people about whom the government is curious.

Data Collection, Data Breach

Storage is cheap, but security is difficult. One of the depressingly common events in the digital era is data breach. A customer database is exposed, and user accounts or credit cards are misused until the breach is rectified. Data breach notification laws in many states now provide some transparency, as well as incentive for companies storing data to clean up to avoid class action lawsuits.

Amid numerous breaches, Equifax and OPM stand out. Equifax, one of the big credit-reporting companies, stores records of credit card account payment histories. If you go to take out a car loan or a mortgage, the lender will check your credit score with Equifax. In September 2017, Equifax announced a data breach that exposed the personal information—names and dates of birth, Social Security numbers, physical addresses, and other personal information that could lead to identity theft and fraud—of 147 million people, more than half the adult population of the United States.[30] The Federal Trade Commission complaint alleged that Equifax failed to take basic network security measures, including failing to update database software when notified of an access control vulnerability. This occurred even as the company had a privacy policy promising it implemented "reasonable physical, technical and procedural safeguards" to protect consumer data. Equifax settled the FTC complaint with an agreement to pay at least $575 million and potentially up to $700 million. As part of the settlement, affected consumers were offered free credit monitoring services. Those trying to exclude themselves from future databases, however, were told "You cannot opt out of this data collection."[31]

As the human resources arm of the U.S. government, the Office of Personnel Management collects a great deal of sensitive information: identification, background checks, and fingerprints.[32] Over 21 million of these records were stolen when OPM's data stores were breached in 2014. When people's credit cards are stolen, they get new cards. When their Social Security numbers are taken, they can be enrolled in credit monitoring services. But you can't be issued a new set of fingerprints.

The number of new data sources—and the proliferation and interconnection of old data sources—is part of the story of how the digital explosion shattered privacy. But the other part of the technology story is about how all that data is put together.

Exponential growth—in storage size, processing speed, and communication speed—have changed the same old thing into something new. Blundering, stupidity, curiosity, malice, and thievery are not new. The fact that sensitive data about everyone in a nation could fit on a laptop *is* new. The ability to search for a needle in the haystack of the Internet *is* new. Easily connecting "public" data sources that used to be stored in file drawers in Albuquerque and Atlanta but are now both electronically accessible from Algeria—*that* is new, too.

Training, laws, and software all can help. But the truth of the matter is that, as a society, we don't really know how to deal with these consequences of the digital explosion. The technology revolution is outstripping society's capacity to adjust to the changes in what can be taken for granted.

Sometimes even public information is revealing. In Massachusetts, the Group Insurance Commission (GIC) is responsible for purchasing health insurance for state employees. When the premiums it was paying jumped one year, the GIC asked for detailed information on every patient encounter. And for good reason: All kinds of health care costs had been growing at prodigious rates. In the public interest, the state had a responsibility to understand how it was spending taxpayer money. The GIC did not want to know patients' names; it did not want to track individuals, and it did not want people to *think* they were being tracked. Indeed, tracking the medical visits of individuals would have been illegal.

So, the GIC data had no names, no addresses, no Social Security numbers, no telephone numbers—nothing that would be a "unique identifier" enabling a mischievous junior staffer in the GIC office to see who exactly had a particular ailment or complaint. To use the official lingo, the data was "deidentified"—that is, stripped of identifying information. The data did include the gender, birth date, zip code, and similar facts about individuals making medical claims, along with some information about why they had sought medical attention. That information was gathered not to challenge any particular person but to learn about patterns; if the truckers in Worcester are having lots of back injuries, for example, maybe workers in that region need better training on how to lift heavy items. Most states do pretty much the same kind of analysis of deidentified data about state workers.

Now this was a valuable data set not just for the Insurance Commission but for others studying public health and the medical industry in Massachusetts. Academic researchers, for example, could use such a large inventory of medical data for epidemiological studies. Because it was all deidentified, there was no harm in letting others see it, the GIC figured. In fact, it was such good data that private industry—for example, businesses in the health management sector—might pay money for it. And so the GIC sold the data to businesses. The taxpayers might even benefit doubly from this decision: The data sale would provide a new revenue source to the state, and in the long run, a more informed health care industry might run more efficiently.

But how deidentified was the material—really?

Latanya Sweeney was at the time a researcher at MIT. (She went on to become a professor at Carnegie Mellon University and then Harvard University.) She wondered how hard it would be for those who had received the deidentified data to "reidentify" the records and learn the medical problems of a particular state employee—for example, the governor of the Commonwealth.

Governor Weld lived, at that time, in Cambridge, Massachusetts. Cambridge, like many other municipalities, makes its voter lists publicly available for a charge of $15—and free for candidates and political organizations. For a particular precinct, you can obtain the records for only $.75. Sweeney spent a few dollars and got the voter lists for Cambridge. Anyone else could have done the same.

According to the Cambridge voter registration list, there were only six people in Cambridge with Governor Weld's birth date, only three of those were men, and only one of those lived in Governor Weld's five-digit zip code. Sweeney could use that combination of factors—birth date, gender, and zip code—to recover the Governor's medical records; she could therefore also recover the records of members of his family, since the data was organized by employee. This type of reidentification is straightforward. In Cambridge, in fact, birth date alone was sufficient to identify more than 10% of the population. Nationally, gender, zip code, and date of birth are all it takes to identify 87% of the U.S. population uniquely.

The data set contained far more than gender, zip code, and birth date. In fact, any of the 58 individuals who received the data in 1997 could have identified any of the 135,000 people in the database. "There is no patient confidentiality," said Dr. Joseph Heyman, president of the Massachusetts Medical Society. "It's gone."[33]

It is easy to read a story like this and scream, "Heads should roll!" But it is actually quite hard to figure out *who, if anyone, made a mistake*. Certainly collecting the information was the right thing to do, given that health costs are a major expense for all businesses and institutions. The GIC made an honest effort to deidentify the data before releasing it. Arguably the GIC might not have released the data to other state agencies. Data is a valuable resource, and once someone has collected it, the government is entirely correct in wanting it used for the public good. Forbidding such sharing would be like saying that every department of government should acquire its heating oil independently. Some might object to selling the data to an outside business—but only in retrospect; had the data really been better deidentified, whoever made the decision to sell the data might well have been rewarded for helping to hold down the cost of government.

Perhaps the mistake was the ease with which voter lists can be obtained. However, it is a tradition deeply ingrained in our system of open elections that the public may know who is eligible to vote and, indeed, who has voted. And voter lists are only one source of public data about the U.S. population. How many

It is easy to read a story like this and scream, "Heads should roll!" But it is actually quite hard to figure out who, if anyone, made a mistake.

21-year-old male Native Hawaiians live in Middlesex County, Massachusetts? In the year 2000, there were four. Anyone can browse the U.S. Census data, and sometimes it can help fill in pieces of a personal picture: Just go to factfinder.census.gov.

The mistake was thinking that the GIC data was truly deidentified when it was not. But with so many data sources available, and so much computing power that could be put to work connecting the dots, it is very hard to know just how much information has to be discarded from a database to make it truly anonymous. Aggregating data into larger units certainly helps; releasing data by five-digit zip codes reveals less than releasing it by nine-digit zip codes. But the coarser the data, the less it reveals also of the valuable information for which it was made available.

The Internet of Things

We have already observed that even privacy-conscious people surrender their privacy in exchange for convenience and small cost savings. Nowhere is this principle more evident than in the networking of light switches, refrigerators, and doorbells known as the Internet of Things (IoT). And it turns out that privacy is not the only thing we sacrifice when we let the Internet grow into everything we touch (and everything we no longer need to touch). The security of everything in the network can be compromised.

On October 21, 2016, the U.S. East Coast woke up to a massive Internet outage. Many popular websites for work and play, such as Twitter, Netflix, GitHub, and Reddit, wouldn't load.[34] It turned out that critical Internet services were under attack by hordes of machines elsewhere on the Internet. These machines were sending so many simultaneous requests that nameservers, key components of the Internet's traffic management infrastructure, couldn't keep up with the load. Trying to respond to malicious requests left them unable to answer legitimate users. Without nameservers to give directions, requesting computers couldn't find the sites, and so Twitter was "down" for users even while the service itself was still functioning.

Investigating researchers and engineers found that the requests were coming from an army of "smart" home devices: Internet-connected baby monitors, light bulbs, and routers. The devices' owners hadn't intended this activity and were mostly unaware of it. The gadgets had been infected by malicious software and enlisted in the Mirai botnet. Together, a group of devices operating with weak computing power on home Internet connections

Unfortunately, this kind of mass attack has become common enough to get a name and an acronym: distributed denial of service (DDoS).

combined into a force strong enough to disrupt global Internet services. The malicious software, or *malware*, exploited some common security flaws—default administrative passwords that hadn't been changed and unpatched and outdated software on devices exposed directly to the Internet—to infect one device and then copy itself to other devices (an infection pattern known as a *worm*). Once installed, the malware turned each device into a waiting "bot," listening for commands.

On October 21, the controller directed the cohort to send a rapid stream of requests for domain names, which caused a burst of traffic to publicly accessible nameservers, including those of major nameservice provider Dyn DNS. Dyn reported that, under the attack, it was getting 10 to 20 times the normal volume of requests, which it estimated came from 100,000 malicious or infected end devices.[35] These requests, along with retry efforts from real end users who couldn't get through, overwhelmed Dyn's defenses and left the company's servers unable to respond to legitimate lookups.

What's New Here? Scale, Control, Connectedness, and Interoperability

The Internet of Things promises to connect the physical world much as the Internet of bits connects computers and data. Sometimes that means putting general-purpose connected computers into previously "dumb" devices like refrigerators. A smart refrigerator might not only warn you when you are running out of milk but contact your grocery store and have it delivered to your home and charged to your credit card. Other times, being "smart" means opening an interface by which one can remotely read from and control sensors (devices that see, hear, or otherwise perceive their environment) and actuators (devices that do something, like shut off the dryer). Smart thermostats, for example, can be triggered by motion detectors to turn on the heat or air conditioning when someone is in the room. These connected things enable a vision for automated factories and supply chains, smart homes and cities, and self-driving car fleets.

As sensors, actuators, and chips get cheaper, they grow in number. They also propagate down the value chain. When chips were expensive, it may have made sense to put them in costly equipment like airplanes, but today they are in doorbells. Even low-end smartphones are now smart enough to be at the center of a home appliance network. Capabilities that once were purchased only by factories and run by experts are now available to the general public. Sometimes smart gadgets are capable of doing much more than the purchaser realizes because the functionality is simplified for marketing reasons. At the same time, safety, reliability, and other less marketable features

are given short shrift—and the manufacturers try to excuse their misplaced priorities by saying that devices must be kept small and operate on low power.

A light bulb or thermostat is often "set and forget": Once the device is functioning, its owner thinks of it as an appliance rather than as a small computer in need of security monitoring and software updates. Moreover, budget sellers of the devices might view them as one-time sales with no follow-on support, and even users who want to update software may find themselves with no option to do so. Another alternative is a suite of centrally managed devices, but this option may be more costly in terms of both dollars and customer privacy. Not everyone wants to share their lighting and temperature preferences, much less the audio and video streams from their baby monitors, with a company storing that data somewhere unknown and using it for who knows what.

Many IoT devices are always on, awaiting the moment when their owners will throw the switch to light a room. That makes them attractive hosts for writers of malware—programs designed with evil intent. The cleverest malware doesn't interfere with the devices' normal function; rather, it lurks invisibly, waiting for the "attack" command.

Threats: One-to-One Versus One-to-Many

When one home has a smart refrigerator, its behavior is interesting and important to only a handful of people.[36] An attacker could spoil a gallon of milk and make a mess of the kitchen, drain a bank account by ordering caviar instead of milk (if the right limits aren't set), or use the machine to cause local damage, possibly shorting the whole house's circuitry. Multiply the devices, however, and they can be leveraged to do damage beyond their neighborhood. The first *D* in DDoS, is for *distributed*. Replicating an attack from thousands of distributed devices can have an overwhelming cumulative effect. Denial of service can take many forms: requests for service that look legitimate but that are sent at a high volume; requests that take an unusually long time to fulfill; or requests that are malformed in such a way as to disable or crash the server to which they are sent. For example, consider what happens when all of a town's high schoolers call the local pizza parlor at the same Friday lunch time to ask the price of a slice with the works. A customer genuinely calling to order a pie will probably give up after a few busy signals.

What makes the IoT an "Internet" is standard protocols that enable the devices to communicate with one another and their controllers, Internet protocols, or special-purpose Wi-Fi or Bluetooth-based communications standards. Standard interfaces enable users to address multiple devices together, such as to plug a new light bulb in to an existing setup or add a freezer to a smart refrigerator. Designers might have anticipated that a connected baby

monitor could share updates with grandparents or enable caregivers to listen from the far reaches of their homes; connectivity could enable a refrigerator to consult a weather report and order ice cream when the temperature hits 80 degrees. Connectivity can enable devices to get smarter over time, with software updates and new possibilities for interaction. However, unguarded connectivity can leave openings for intrusions like the Mirai worm, and the common interfaces and underlying software enable malware writers to "break once, run anywhere."

In December 2017, more than a year after attacks took Dyn offline, three men pleaded guilty to charges of computer fraud and abuse, admitting to having written the software behind Mirai: Paras Jha, an undergraduate studying computer science at Rutgers University in New Jersey, and two friends or associates. According to their plea, they first targeted their attack against online gaming servers for the popular Minecraft game, where they were attempting to overwhelm the servers to gain an advantage. Later, Jha started a business selling computer-protection services and launched attacks against Rutgers while taunting the university that it should be buying DDoS protection. Jha and his associates didn't necessarily intend to disrupt Dyn or the Internet at large, but after they posted the software's source code online, others modified and redeployed the malware, pointing it at new targets.

Who's Responsible for IoT Security?

When the Tesla Model 3 electric car was first reviewed by *Consumer Reports*, it got poor marks for braking.[37] "The Tesla's stopping distance of 152 feet from 60 mph was far worse than any contemporary car we've tested," wrote the reviewer. A week after publication of the *Consumer Reports* review, however, the car manufacturer sent an over-the-air software update to cars across the country, including those that had already been sold. The car's braking distance improved by 19 feet, performance comparable to that of other compact cars, prompting *Consumer Reports* to upgrade its review.[38] Tesla told *Consumer Reports* that it had updated software controlling the Model 3's antilock braking system.

This wasn't the first time an over-the-air update changed vehicle performance. While Hurricane Irma was heading for Florida, Tesla overrode software-defined range limitations for cars in the storm's path, enabling owners to escape further.[39] Both cases illustrate the blurred line between software and hardware and murky outlines of product boundaries. Physical features of the car were changed by a remote software update, and a car's owner might not have even been aware of the change or given an opportunity to accept or reject it. Few owners would reject longer range (a feature that was a costly upgrade outside Irma's wake), but what if the more consistent braking came

at the cost of other performance features? Some owners found themselves thinking their cars were slower after the upgrade. What if public safety came at the price of speed? Should owners get to refuse updates?

Driving poses an externality problem. It's not just Tesla drivers who take risks if their cars don't stop in time; their vehicles are more dangerous to everyone who shares the road with them. We impose safety standards and inspection requirements on automobiles to reduce such risks and make roads safer. We might similarly impose a duty to upgrade on software and hardware users. If your software-enabled product is causing risks to others, if a safer alternative is found, you could be required to update, even if doing so would cause you some inconvenience or cost. Yet it's not just obviously dangerous and expensive objects like cars that require this caution; some of the connected devices taken over by the Mirai botnet were cheap toys. Some of their vendors might no longer be in business. Would this requirement change the nature of ownership?

Bruce Schneier speaks of the Internet of Things as a "world-sized robot," with sensors and actuators spanning the globe.[40] As the capabilities of this robot to cause harm—and actual examples of harm—multiply, he predicts that demands for regulation and liability will follow. Unless those who are building the technologies also build safeguards, the political and regulatory responses are likely to be blunt and may include prohibitions on connecting or using devices or broad restrictions on their use. Worse, regulations that do not account for the architectures and incentives of connectedness may fail to protect us.

Smart Cities: Efficiency, Individual Choice, Privacy, and Systemic Risks

An older man in a New York apartment complained that he was virtually imprisoned in his own home after the landlord installed app-controlled "smart locks" at the building's lobby entrance. Tony Mysak, 93 and blind in one eye, was unable to use the smartphone app required to open the lock. Mary Beth McKenzie, Tony's wife, objected to giving a record of her entries to the building and to Latch, the lock provider. The Latch app's privacy policy (since changed) noted that the app collected GPS location information that Latch might use for marketing purposes, as well as providing a record of door accesses and photographs to building management. When she asked for a physical key, the landlord laughed and offered only a smartcard. McKenzie and Mysak and a group of tenants had to sue their landlord to win the right to use keys instead of apps.[41]

The tenants had several complaints about the digitization of their front door. For some, it was the change in usability, from a familiar physical key to

a new application. For others, it was the privacy, the sense that their entries and exits—and perhaps even their travels—would be tracked and compiled, not by a human and fallible doorman but in an impersonal corporate database that wasn't even visible to them. The new affordances of this system, such as the ability to let a guest or super into the building without leaving a key under the doormat, weren't enough to compensate tenants for their loss of control.

Scale this "smart" building up a few orders of magnitude, and you get the "smart city," full of embedded and networked sensors. Traffic lights might coordinate with cars and buses for efficiency; power meters might communicate with the electrical grid in real time to smooth demand.

The city of Toronto planned a revitalization of its industrial waterfront "from the Internet up." The new Quayside would be built as a "smart city" in a partnership between the city and Google/Alphabet's Sidewalk Labs. But as they engaged in giddy futuristic speculation, planners were surprised by the opposition their announcement sparked. People complained about privacy, security, and loss of control. Who gets to see the data generated by the digital infrastructure; who gets to make decisions based on it? Sadly, we won't learn the answer. In May 2020, Toronto and Google scrapped the project. The decision was taken amid the Coronavirus pandemic—but privacy advocates claimed credit.

> "This is a major victory for the responsible citizens who fought to protect Canada's democracy, civil and digital rights," said one opponent of the project. "Toronto will go down in history as one of the more disturbing planned experiments in surveillance capitalism"[42]—referring to the title of a best-selling business book.

Interconnection brings new privacy and security concerns. Who can learn when you're out of town by watching for changes in power usage patterns? Who can learn when you have company or take a hot shower? By monitoring the power signatures of home devices, a watcher could even see when you start the morning coffee pot or turn on the evening news.

The flow of bits, storage capacity, and processing power needed for analysis all tend to heighten the power disadvantage of individuals against governments and corporations. Privacy serves as a way of taking back some control, a zone of autonomy. In Orwell's imagined London, only O'Brien and other members of the Inner Party could escape the gaze of the telescreen. For now, individuals can employ a mix of mathematical and legal protections to shut out the watching eyes of Big Brother—at least most of the time.

Endnotes

1 Sopan Deb and Natasha Singer, "Taylor Swift Keeping An Eye Out For Stalkers," *New York Times*, December 15, 2018, C6, https://www.nytimes.com/2018/12/13/arts/music/taylor-swift-facial-recognition.html

2 George Orwell, *1984* (Signet Classic, 1977), p. 2.

3 Silkie Carlo, "Britain Has More Surveillance Cameras per Person Than Any Country Except China. That's a Massive Risk to Our Free Society," *Time*, May 17, 2019, https://news.yahoo.com/britain-more-surveillance-cameras-per-151641361.html.

4 Lee Rainie, "Americans' Complicated Feelings About Social Media in an Era of Privacy Concerns," Pew Research Center, March 27, 2018, https://www.pewresearch.org/fact-tank/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/.

5 Lee Raine, Americans' complicated feelings about social media in an era of privacy concerns, Pew Research Center, March 27, 2018, https://www.pewresearch.org/fact-tank/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/.

6 Edward Snowden, *Permanent Record* (Metropolitan Books, 2019).

7 https://www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html.

8 Kevin Bankston, "EFF Analysis of the Provisions of the USA PATRIOT Act," Electronic Frontier Foundation, October 27, 2003, https://www.eff.org/deeplinks/2003/10/eff-analysis-provisions-usa-patriot-act.

9 Glenn Greenwald, "NSA Collecting Phone Records of Millions of Verizon Customers Daily," *The Guardian*, June 6, 2013, https://www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order.

10 Micah Lee et al., "A Look at the Inner Workings of NSA's XKEYSCORE," *The Intercept*, July 2, 2015, https://theintercept.com/2015/07/02/look-under-hood-xkeyscore/.

11 Tom Bowman, "Why Does the NSA Keep an EGOTISTICALGIRAFFE? It's Top Secret," NPR, November 10, 2003, https://www.npr.org/2013/11/10/244240199/why-does-the-nsa-keep-an-egotisticalgiraffeits-top-secret.

12 David Cole, "We Kill People Based on Metadata," *The New York Review of Books*, May 10, 2014, https://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/.

13 Stephen Farrell and Hannes Tschofenig, "Pervasive Monitoring Is an Attack," Internet Engineering Task Force, RFC 7258, May 2014, https://tools.ietf.org/html/rfc7258.

14 "HTTPS Encryption on the Web," Google Transparency Report, accessed May 18, 2020, https://transparencyreport.google.com/https/overview;%20https:=.

15 *Olmstead v. United States*, 277 U.S. 438 (1928), https://supreme.justia.com/cases/federal/us/277/438/.

16 *Katz v. United States*, 389 U.S. 347 (1967), https://supreme.justia.com/cases/federal/us/389/347/.

17 Eric Sandy, "Supreme Court Case Has Roots in Radio Shack Robberies in Michigan and Ohio," *Detroit Metro Times*, November 28, 2017, https://www.metrotimes.com/news-hits/archives/2017/11/28/supreme-court-case-has-roots-in-radio-shack-robberies-in-michigan-and-ohio.

18 *Carpenter v. United States*, 138 S. Ct. 2206 (2018), https://www.oyez.org/cases/2017/16-402.

19 Ellen Messmer, "Black Hat: Researcher Claims Hack of Processor Used to Secure Xbox 360, Other Products," *Network World,* February 2, 2010, https://www.networkworld.com/article/2243700/black-hat--researcher-claims-hack-of-processor-used-to-secure-xbox-360--other-products.html.

20 *People v. Christmann*, 861 N.W.2d 18 (2015), https://caselaw.findlaw.com/ny-justice-court/1143124.html.

21 Taylor Hatmaker, "California Malls Are Sharing License Plate Tracking Data with an ICE-Linked Database," TechCrunch, July 10, 2018, https://social.techcrunch.com/2018/07/10/alpr-license-plate-recognition-ice-irvine-company/.

22 Thomas Brewster, "Why Strava's Fitness Tracking Should Really Worry You," *Forbes*, January 29, 2018, https://www.forbes.com/sites/thomasbrewster/2018/01/29/strava-fitness-data-location-privacys-care/#46e488aa55c3.

23 James Quarles, "A Letter to the Strava Community," Strava, January 29, 2018 https://blog.strava.com/press/a-letter-to-the-strava-community/.

24 Chris Buckley and Paul Mozur, "How China Uses High-Tech Surveillance to Subdue Minorities," *The New York Times*, May 22, 2019, https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang. html.

25 Ellen Nakashima, "FBI Prepares Vast Database of Biometrics," *The Washington Post*, December 22, 2007, http://www.washingtonpost.com/wp-dyn/content/article/2007/12/21/AR2007122102544.html.

26 Kate Conger et al., "San Francisco Bans Facial Recognition Technology," *The New York Times*, May 14, 2019, https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html.

27 Susie Cagle, "This ID Scanner Company Is Collecting Sensitive Data on Millions of Bargoers," Medium, May 29, 2019, https://onezero.medium.com/id-at-the-door-meet-the-security-company-building-an-international-database-of-banned-bar-patrons-7c6d4b236fc3.

28 "The 2019 Federal Reserve Payments Study," Board of Governors of the Federal Reserve System, January 6, 2020, https://www.federalreserve.gov/paymentsystems/2019-December-The-Federal-Reserve-Payments-Study.htm.

29 GAO, U.S. Government Accountability Office, https://www.gao.gov/products/GAO-04-548.

30 "Equifax to Pay $575 Million as Part of Settlement with FTC, CFPB, and States Related to 2017 Data Breach," Federal Trade Commission, July 22, 2019, https://www.ftc.gov/news-events/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related.

31 "Equifax Data Breach Settlement," Federal Trade Commission, July 11, 2019, https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement.

32 OPM.GOV, Cybersecurity Resource Center, https://www.opm.gov/cybersecurity/cybersecurity-incidents/

33 Michael Lasalandra, "Panel told releases of med records hurt privacy," *Boston Herald*, March 20, 1997.

34 Manos Antonakakis et al., "Understanding the Mirai Botnet," *Proceedings of 26th USENIX Security Symposium*, April 16, 2017; "Mirai IoT Botnet Co-Authors Plead Guilty," Krebs on Security, December 13, 2017, https://krebsonsecurity.com/2017/12/mirai-iot-botnet-co-authors-plead-guilty/.

35 Scott Hilton, "Dyn Analysis Summary of Friday October 21 Attack," Oracle, 2016. http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack/.

36 C. J. Hughes, "The Latest in Apartment Technology: Fridge Cams and Robotic Valets," *The New York Times*, December 15, 2017, https://www.nytimes.com/2017/12/15/realestate/apartment-technology-fridge-cams-robotic-valets.html.

37 Patrick Olsen, "Tesla Model 3 Falls Short of a CR Recommendation," *Consumer Reports*, May 30, 2018, https://www.consumerreports.org/hybrids-evs/tesla-model-3-review-falls-short-of-consumer-reports-recommendation/.

38 Patrick Olsen, "Tesla Model 3 Gets CR Recommendation After Braking Update," *Consumer Reports*, May 30, 2018, https://www.consumerreports.org/car-safety/tesla-model-3-gets-cr-recommendation-after-braking-update/.

39 Andrew Liptak, "Tesla Extended the Range of Some Florida Vehicles for Drivers to Escape Hurricane Irma," *The Verge*, September 10, 2017, https://www.theverge.com/2017/9/10/16283330/tesla-hurricane-irma-update-florida-extend-range-model-s-x-60-60d.

40 Bruce Schneier, *Click Here to Kill Everybody: Security and Survival in a Hyper-Connected World* (WW Norton & Company, 2018).

41 Alfred Ng, "Tenants Win as Settlement Orders Landlords Give Physical Keys over Smart Locks," CNET, May 7, 2019, https://www.cnet.com/news/tenants-win-rights-to-physical-keys-over-smart-locks-from-landlords/.

42 Rob Gillies, Google Affiliate scraps plan for Toronto Smart City Project, *US News and World Report*, May 7, 2020, https://www.usnews.com/news/business/articles/2020-05-07/google-affiliate-scraps-plan-for-toronto-smart-city-project and Shoshana Zuboff, *The Age of Surveillance Capitalism,* PublicAffairs, 2019.