CHAPTER 2

Naked in the Sunlight

Privacy Lost, Privacy Abandoned

1984 Is Here, and We Like It

On July 7, 2005, London was shaken as suicide bombers detonated four explosions, three on subways and one on a double-decker bus. The attack on the transit system was carefully timed to occur at rush hour, maximizing its destructive impact. 52 people died and 700 more were injured.

Security in London had already been tight. The city was hosting the G8 Summit, and the trial of fundamentalist cleric Abu Hamza al-Masri had just begun. Hundreds of thousands of surveillance cameras hadn't deterred the terrorist act, but the perpetrators were caught on camera. Their pictures were sent around the world instantly. Working from 80,000 seized tapes, police were able to reconstruct a reconnaissance trip the bombers had made two weeks earlier.

George Orwell's *1984* was published in 1948. Over the 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 ...

The real 1984 came and went nearly a quarter century ago. Today, Big Brother's two-way telescreens would be amateurish toys. Orwell's imagined

London had cameras everywhere. His actual city now has at least half a million. Across the UK, there is one surveillance camera for every dozen people. 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. 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 his 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, while computers of staggering capacity reconstruct our movements from the tracks. Computers re-assemble 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. Yet 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. According to a 2007 Pew/Internet Project report, "60% of Internet users say they are not worried about how much information is available about them online." Many of us publish and broadcast the most intimate moments of our lives for all the world to see, even when no one requires or even asks us to do so. 55% of teenagers and 20% of adults have created profiles on social networking web sites. A third of the teens with profiles, and half the adults, place no restrictions on who can see them.

In Orwell's imagined London, only O'Brien and other members of the Inner Party could escape the gaze of the telescreen. For the rest, the constant gaze was a source of angst and anxiety. Today, we willingly accept the gaze. We either don't think about it, don't know about it, or feel helpless to avoid it except by becoming hermits. We may even judge its benefits to outweigh its risks. In Orwell's imagined London, like Stalin's actual Moscow, citizens spied on their fellow citizens. Today, we can all be Little Brothers, using our search

engines to check up on our children, our spouses, our neighbors, our colleagues, our enemies, and our friends. More than half of all adult Internet users have done exactly that.

The explosive growth in digital technologies has radically altered our expectations about what will be private and shifted our thinking about what *should* be private. Ironically, the notion of privacy has become fuzzier at the same time as the secrecy-enhancing technology of encryption has become widespread. Indeed, it is remarkable that we no longer blink at intrusions that a decade ago would have seemed shocking. Unlike the story of secrecy, there was no single technological event that caused the change, no privacy-shattering breakthrough—only a steady advance on several technological fronts that ultimately passed a tipping point.

Public Organizations Involved in Defending Privacy

Existing organizations have focused on privacy issues in recent years, and new ones have sprung up. In the U.S., important forces are the American Civil Liberties Union (ACLU, `www.aclu.org`), the Electronic Privacy Information Center (`EPIC`, `epic.org`), the Center for Democracy and Technology (CDT, `www.cdt.org`), and the Electronic Frontier Foundation (`www.eff.org`).

Many devices got cheaper, better, and smaller. Once they became useful consumer goods, we stopped worrying about their uses as surveillance devices. For example, if the police were the only ones who had cameras in their cell phones, we would be alarmed. But as long as we have them too, so we can send our friends funny pictures from parties, we don't mind so much that others are taking pictures of us. The social evolution that was supported by consumer technologies in turn made us more accepting of new enabling technologies; the social and technological evolutions have proceeded hand in hand. Meanwhile, international terrorism has made the public in most democracies more sympathetic to intrusive measures intended to protect our security. With corporations trying to make money from us and the government trying to protect us, civil libertarians are a weak third voice when they warn that we may not want others to know so much about us.

So we tell the story of privacy in stages. First, we detail the enabling technologies, the devices and computational processes that have made it easy and convenient for us to lose our privacy—some of them familiar technologies, and some a bit more mysterious. We then turn to an analysis of how we have lost our privacy, or simply abandoned it. Many privacy-shattering things have happened to us, some with our cooperation and some not. As a result, the sense of personal privacy is very different today than it was two decades ago. Next, we discuss the social changes that have occurred—cultural shifts

that were facilitated by the technological diffusion, which in turn made new technologies easier to deploy. And finally we turn to the big question: What does privacy even mean in the digitally exploded world? Is there any hope of keeping anything private when everything is bits, and the bits are stored, copied, and moved around the world in an instant? And if we can't—or won't—keep our personal information to ourselves anymore, how can we make ourselves less vulnerable to the downsides of living in such an exposed world? Standing naked in the sunlight, is it still possible to protect ourselves against ills and evils from which our privacy used to protect us?

Footprints and Fingerprints

As we do our daily business and lead our private lives, we leave footprints and fingerprints. We can see our footprints in mud on the floor and in the sand and snow outdoors. We would not be surprised that anyone who went to the trouble to match our shoes to our footprints could determine, or guess, where we had been. Fingerprints are different. It doesn't even occur to us that we are leaving them as we open doors and drink out of tumblers. Those who have guilty consciences may think about fingerprints and worry about where they are leaving them, but the rest of us don't.

THE UNWANTED GAZE

The Unwanted Gaze by Jeffrey Rosen (Vintage, 2001) details many ways in which the legal system has contributed to our loss of privacy.

In the digital world, we all leave both electronic footprints and electronic fingerprints—data trails we leave intentionally, and data trails of which we are unaware or unconscious. The identifying data may be useful for forensic purposes. Because most of us don't consider ourselves criminals, however, we tend not to worry about that. What we don't think about is that the various small smudges we leave on the digital landscape may be useful to someone else—someone who wants to use the data we left behind to make money or to get something from us. It is therefore important to understand how and where we leave these digital footprints and fingerprints.

Smile While We Snap!

Big Brother had his legions of cameras, and the City of London has theirs today. But for sheer photographic pervasiveness, nothing beats the cameras in the cell phones in the hands of the world's teenagers. Consider the alleged misjudgment of Jeffrey Berman. In early December 2007, a man about

60 years old committed a series of assaults on the Boston public transit system, groping girls and exposing himself. After one of the assaults, a victim took out her cell phone. Click! Within hours, a good head shot was up on the Web and was shown on all the Boston area television stations. Within a day, Berman was under arrest and charged with several crimes. "Obviously we, from time to time, have plainclothes officers on the trolley, but that's a very difficult job to do," said the chief of the Transit Police. "The fact that this girl had the wherewithal to snap a picture to identify him was invaluable."

That is, it would seem, a story with a happy ending, for the victim at least. But the massive dissemination of cheap cameras coupled with universal access to the Web also enables a kind of vigilante justice—a ubiquitous Little-Brotherism, in which we can all be detectives, judges, and corrections officers. Mr. Berman claims he is innocent; perhaps the speed at which the teenager's snapshot was disseminated unfairly created a presumption of his guilt. Bloggers can bring global disgrace to ordinary citizens.

In June 2005, a woman allowed her dog to relieve himself on a Korean subway, and subsequently refused to clean up his mess, despite offers from others to help. The incident was captured by a fellow passenger and posted online. She soon became known as "gae-ttong-nyue" (Korean for "puppy poo girl"). She was identified along with her family, was shamed, and quit school. There is now a Wikipedia entry about the incident. Before the digital explosion—before bits made it possible to convey information instantaneously, everywhere—her actions would have been embarrassing and would have been known to those who were there at the time. It is unlikely that the story would have made it around the world, and that it would have achieved such notoriety and permanence.

There are many free webcam sites, at which you can watch what's happening right now at places all over the world. Here are a few:

```
www.camvista.com
www.earthcam.com
www.webcamworld.com
www.webworldcam.com
```

Still, in these cases, at least someone thought someone did something wrong. The camera just happened to be in the right hands at just the right moment. But looking at images on the Web is now a leisure activity that anyone can do at any time, anywhere in the world. Using Google Street View, you can sit in a café in Tajikistan and identify a car that was parked in my driveway when Google's camera came by (perhaps months ago). From Seoul, you can see what's happening right now, updated every few seconds, in Picadilly Circus or on the strip in Las Vegas. These views were always available to the public, but cameras plus the Web changed the meaning of "public."

And an electronic camera is not just a camera. *Harry Potter and the Deathly Hallows* is, as far as anyone knows, the last book in the Harry Potter series. Its arrival was eagerly awaited, with lines of anxious Harry fans stretching around the block at bookstores everywhere. One fan got a pre-release copy, painstakingly photographed every page, and posted the entire book online before the official release. A labor of love, no doubt, but a blatant copyright violation as well. He doubtless figured he was just posting the pixels, which could not be traced back to him. If that was his presumption, he was wrong. His digital fingerprints were all over the images.

Digital cameras encode metadata along with the image. This data, known as the Exchangeable Image File Format (EXIF), includes camera settings (shutter speed, aperture, compression, make, model, orientation), date and time, and, in the case of our Harry Potter fan, the make, model, and serial number of his camera (a Canon Rebel 350D, serial number 560151117). If he registered his camera, bought it with a credit card, or sent it in for service, his identity could be known as well.

Knowing Where You Are

Global Position Systems (GPSs) have improved the marital lives of countless males too stubborn to ask directions. Put a Garmin or a Tom Tom in a car, and it will listen to precisely timed signals from satellites reporting their positions in space. The GPS calculates its own location from 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 that makes these systems popular birthday presents.

If you carry a GPS-enabled cell phone, your friends can find you, if that its what you want. If your GPS-enabled rental car has a radio transmitter, you can be found whether you want it or not. In 2004, Ron Lee rented a car from Payless in San Francisco. He headed east to Las Vegas, then back to Los Angeles, and finally home. He was expecting to pay $150 for his little vacation, but Payless made him pay more—$1,400, to be precise. Mr. Lee forgot to read the fine print in his rental contract. He had not gone too far; his contract was for unlimited mileage. He had missed the fine print that said, "Don't leave California." When he went out of state, the unlimited mileage clause was invalidated. The fine print said that Payless would charge him $1 per Nevada mile, and that is exactly what the company did. They knew where he was, every minute he was on the road.

A GPS will locate you anywhere on earth; that is why mountain climbers carry them. They will locate you not just on the map but in three dimensions, telling you how high up the mountain you are. But even an ordinary cell phone will serve as a rudimentary positioning system. If you are traveling in

settled territory—any place where you can get cell phone coverage—the signals from the cell phone towers can be used to locate you. That is how Tanya Rider was found (see Chapter 1 for details). The location is not as precise as that supplied by a GPS—only within ten city blocks or so—but the fact that it is possible at all means that photos can be stamped with identifying information about where they were shot, as well as when and with what camera.

Knowing Even Where Your Shoes Are

A Radio Frequency Identification tag—RFID, for short—can be read from a distance of a few feet. Radio Frequency Identification is like a more elaborate version of the familiar bar codes that identify products. Bar codes typically identify what kind of thing an item is—the make and model, as it were. Because RFID tags have the capacity for much larger numbers, they can provide a unique serial number for each item: not just "Coke, 12 oz. can" but "Coke can #12345123514002." And because RFID data is transferred by radio waves rather than visible light, the tags need not be visible to be read, and the sensor need not be visible to do the reading.

RFIDs are silicon chips, typically embedded in plastic. They can be used to tag almost anything (see Figure 2.1). "Prox cards," which you wave near a sensor to open a door, are RFID tags; a few bits of information identifying you are transmitted from the card to the sensor. Mobil's "Speedpass" is a little RFID on a keychain; wave it near a gas pump and the pump knows whom to charge for the gasoline. For a decade, cattle have had RFIDs implanted in their flesh, so individual animals can be tracked. Modern dairy farms log the milk production of individual cows, automatically relating the cow's identity to its daily milk output. Pets are commonly RFID-tagged so they can be reunited with their owners if the animals go missing for some reason. The possibility of tagging humans is obvious, and has been proposed for certain high-security applications, such as controlling access to nuclear plants.

But the interesting part of the RFID story is more mundane—putting tags in shoes, for example. RFID can be the basis for powerful inventory tracking systems.

RFID tags are simple devices. They store a few dozen bits of information, usually unique to a particular tag. Most are passive devices, with no batteries, and are quite small. The RFID includes a tiny electronic chip and a small coil, which acts as a two-way antenna. A weak

SPYCHIPS

This aptly named book by Katherine Albrecht and Liz McIntyre (Plume, 2006) includes many stories of actual and proposed RFID uses by consumer goods manufacturers and retailers.

current flows through the coil when the RFID passes through an electromagnetic field—for example, from a scanner in the frame of a store, under the carpet, or in someone's hand. This feeble current is just strong enough to power the chip and induce it to transmit the identifying information. Because RFIDs are tiny and require no connected power source, they are easily hidden. We see them often as labels affixed to products; the one in Figure 2.1 was between the pages of a book bought from a bookstore. They can be almost undetectable.

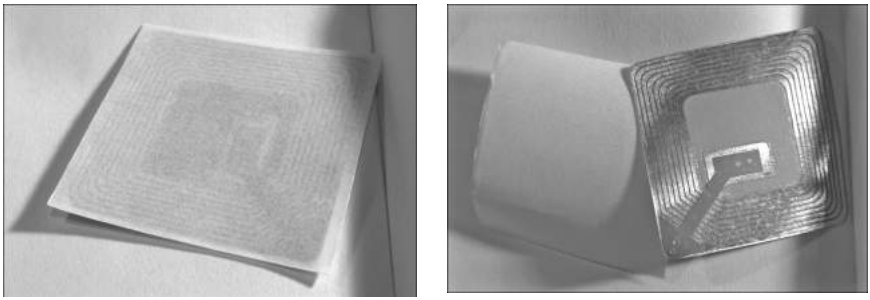


FIGURE 2.1 An RFID found between the pages of a book. A bookstore receiving a box of RFID-tagged books can check the incoming shipment against the order without opening the carton. If the books and shelves are scanned during stocking, the cash register can identify the section of the store from which each purchased copy was sold.

RFIDs are generally used to improve record-keeping, not for snooping. Manufacturers and merchants want to get more information, more reliably, so they naturally think of tagging merchandise. But only a little imagination is required to come up with some disturbing scenarios. Suppose, for example, that you buy a pair of red shoes at a chain store in New York City, and the shoes have an embedded RFID. If you pay with a credit card, the store knows your name, and a good deal more about you from your purchasing history. If you wear those shoes when you walk into a branch store in Los Angeles a month later, and that branch has an RFID reader under the rug at the entrance, the clerk could greet you by name. She might offer you a scarf to match the shoes—or to match anything else you bought recently from any other branch of the store. On the other hand, the store might know that you have a habit of returning almost everything you buy—in that case, you might find yourself having trouble finding anyone to wait on you!

The technology is there to do it. We know of no store that has gone quite this far, but in September 2007, the Galeria Kaufhof in Essen, Germany equipped the dressing rooms in the men's clothing department with RFID readers. When a customer tries on garments, a screen informs him of available sizes and colors. The system may be improved to offer suggestions about accessories. The store keeps track of what items are tried on together and what combinations turn into purchases. The store will remove the RFID tags from the clothes after they are purchased—if the customer asks; otherwise, they remain unobtrusively and could be scanned if the garment is returned to the store. Creative retailers everywhere dream of such ways to use devices to make money, to save money, and to give them small advantages over their competitors. Though Galeria Kaufhof is open about its high-tech men's department, the fear that customers won't like their clever ideas sometimes holds back retailers—and sometimes simply causes them to keep quiet about what they are doing.

Black Boxes Are 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.

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 had not fastened his seatbelt, and 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 being tracked his speed. We know his exact speed at the moment of impact because his car, like 30 million 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. By federal law, they will be mandatory in the United States beginning in 2011. If you are driving a new GM, Ford, Isuzu, Mazda, Mitsubishi, or Subaru, your car has one—whether anyone told you that or not. So do about half of new Toyotas. Your insurance company is probably entitled to its data if you have an accident. Yet most people do not realize that they exist.

EDRs capture information about speed, braking time, turn signal status, seat belts: things 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 MPH 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 against unreasonable searches and seizures, because it had not asked his permission or obtained a search warrant before retrieving the data. That 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.

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. Yet even if we live our lives without walking, we would unsuspectingly be leaving fingerprints anyway.

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

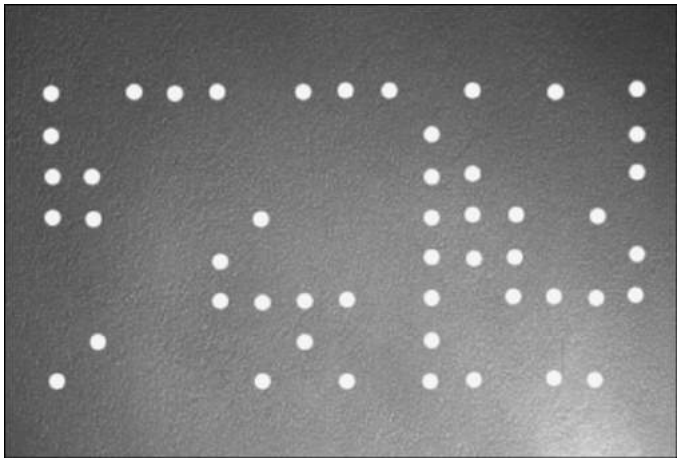
Some of the intrusions into our privacy come because of the unexpected, unseen side effects of things we do quite voluntarily. We painted the hypothetical picture of the shopper with the RFID-tagged shoes, who is either welcomed or shunned on her subsequent visits to the store, depending on her shopping history. Similar surprises can lurk almost anywhere that bits are exchanged. That is, for practical purposes, pretty much everywhere in daily life.

Tracing Paper

If I send an email or download a web page, it should come as no surprise that I've left some digital footprints. After all, the bits have to get to me, so some part of the system knows where I am. In the old days, if I wanted to be anonymous, I could write a note, but my handwriting might be recognizable, and I might leave fingerprints (the oily kind) on the paper. I might have typed, but Perry Mason regularly solved crimes by matching a typewritten note with the unique signature of the suspect's typewriter. More fingerprints.

So, today I would laserprint the letter and wear gloves. But even that may not suffice to disguise me. Researchers at Purdue have developed techniques for matching laser-printed output to a particular printer. They analyze printed sheets and detect unique characteristics of each manufacturer and each individual printer—fingerprints that can be used, like the smudges of old typewriter hammers, to match output with source. It may be unnecessary to put the microscope on individual letters to identify what printer produced a page.

The Electronic Frontier Foundation has demonstrated that many color printers secretly encode the printer serial number, date, and time on every page that they print (see Figure 2.2). Therefore, when you print a report, you should not assume that no one can tell who printed it.



Source: Laser fingerprint. Electronic Frontier Foundation. `http://w.2.eff.org/Privacy/printers/docucolor/`.

FIGURE 2.2 Fingerprint left by a Xerox DocuColor 12 color laser printer. The dots are very hard to see with the naked eye; the photograph was taken under blue light. The dot pattern encodes the date (2005-05-21), time (12:50), and the serial number of the printer (21052857).

There was a sensible rationale behind this technology. The government wanted to make sure that office printers could not be used to turn out sets of hundred dollar bills. The technology that was intended to frustrate counterfeiters makes it possible to trace every page printed on color laser printers back to the source. Useful technologies often have unintended consequences.

Many people, for perfectly legal and valid reasons, would like to protect their anonymity. They may be whistleblowers or dissidents. Perhaps they are merely railing against injustice in their workplace. Will technologies that undermine anonymity in political discourse also stifle free expression? A measure of anonymity is essential in a healthy democracy—and in the U.S., has been a weapon used to advance free speech since the time of the Revolution. We may regret a complete abandonment of anonymity in favor of communication technologies that leave fingerprints.

The problem is not just the existence of fingerprints, but that no one told us that we are creating them.

The problem is not just the existence of fingerprints, but that no one told us that we are creating them.

The Parking Garage Knows More Than You Think

One day in the spring of 2006, Anthony and his wife drove to Logan Airport to pick up some friends. They took two cars, which they parked in the garage. Later in the evening, they paid at the kiosk inside the terminal, and left—or tried to. One car got out of the garage without a problem, but Anthony's was held up for more than an hour, in the middle of the night, and was not allowed to leave. Why? Because his ticket did not match his license plate.

It turns out that every car entering the airport garage has its license plate photographed at the same time as the ticket is being taken. Anthony had held both tickets while he and his wife were waiting for their friends, and then he gave her back one—the "wrong" one, as it turned out. It was the one he had taken when he drove in. When he tried to leave, he had the ticket that matched his wife's license plate number. A no-no.

Who knew that if two cars arrive and try to leave at the same time, they may not be able to exit if the tickets are swapped? In fact, who knew that every license plate is photographed as it enters the garage?

There is a perfectly sensible explanation. People with big parking bills sometimes try to duck them by picking up a second ticket at the end of their trip. When they drive out, they try to turn in the one for which they would have to pay only a small fee. Auto thieves sometimes try the same trick. So the system makes sense, but it raises many questions. Who else gets access to the license plate numbers? If the police are looking for a particular car, can they search the scanned license plate numbers of the cars in the garage? How long is the data retained? Does it say anywhere, even in the fine print, that your visit to the garage is not at all anonymous?

All in Your Pocket

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.

On October 18, 2007, a junior staff member at the British national tax agency sent a small package to the government's auditing agency via TNT, a private delivery service. Three weeks later, it had not arrived at its destination and was reported missing. Because the sender had not used TNT's "registered mail" option, it couldn't be traced, and as of this writing has not been found. Perhaps it was discarded by mistake and never made it out of the mailroom; perhaps it is in the hands of criminals.

The mishap rocked the nation. As a result of the data loss, every bank and millions of individuals checked account activity for signs of fraud or identity theft. On November 20, the head of the tax agency resigned. Prime Minister Gordon Brown apologized to the nation, and the opposition party accused the Brown administration of having "failed in its first duty—to protect the public."

The package contained two computer disks. The data on the disks included names, addresses, birth dates, national insurance numbers (the British equivalent of U.S. Social Security Numbers), and bank account numbers of 25 million people—nearly 40% of the British population, and almost every child in the land. The tax office had all this data because every British child receives weekly government payments, and most families have the money deposited directly into bank accounts. Ten years ago, that much data would have required a truck to transport, not two small disks. Fifty years ago, it would have filled a building.

This was a preventable catastrophe. Many mistakes were made; quite ordinary mistakes. The package should have been registered. The disks should have been encrypted. It should not have taken three weeks for someone to speak up. But those are all age-old mistakes. Offices have been sending packages for centuries, and even Julius Caesar knew enough to encrypt information if he had to use intermediaries to deliver it. What happened in 2007 that could not have happened in 1984 was the assembly of such a massive database in a form that allowed it to be easily searched, processed, analyzed, connected to other databases, transported—and "lost."

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. The Prime Minister had to apologize to the British nation because among the things that have been blown to bits is the presumption that no junior staffer could do that much damage by mailing a small parcel.

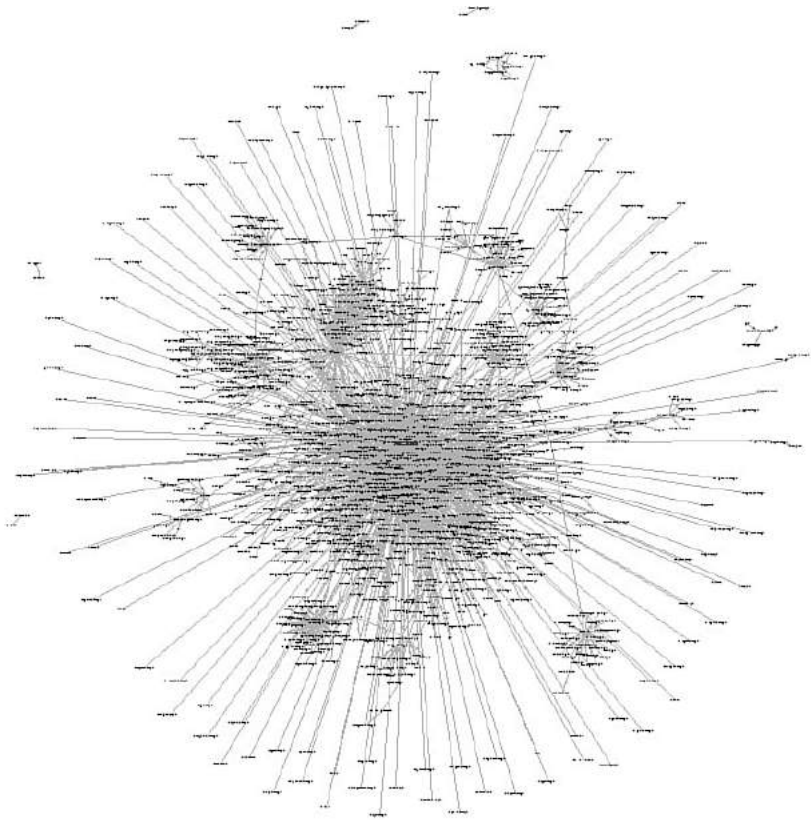
Connecting the Dots

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 have been 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, 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 2.3). 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 time of the communications and can correlate them with the time of other events.

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.



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

FIGURE 2.3 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.

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 "de-identified"; 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 de-identified 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 de-identified, 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 de-identified really was the material?

Latanya Sweeney was at the time a researcher at MIT (she went on to become a computer science professor at Carnegie Mellon University). She wondered how hard it would be for those who had received the de-identified data to "re-identify" 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 municipalities, makes its voter lists publicly available, for a charge of $15, and free for candidates and political organizations. If you know the precinct, they are available for only $.75. Sweeney spent a few dollars and got the voter lists for Cambridge. Anyone 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—and also those for members of his family, since the data was organized by employee. This type of re-identification 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."

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 de-identify the data before releasing it. Arguably the GIC might not have released the data to other state agencies, but that would be like saying that every department of government should acquire its heating oil independently. Data is a valuable resource, and once someone has collected it, the government is entirely correct in wanting it used for the public good. Some might object to selling the data to an outside business, but only in retrospect; had the data really been better de-identified, whoever made the decision to sell the data might well have been rewarded for helping to hold down the cost of government.

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.

Perhaps the mistake was the ease with which voter lists can be obtained. However, it is a tradition deeply engrained 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 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 de-identified, 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.

How can we solve a problem that results from many developments, no one of which is really a problem in itself?

Why We Lost Our Privacy, or Gave It Away

Information technology did not cause the end of privacy, any more than automotive technology caused teen sex. Technology creates opportunities and risks, and people, as individuals and as societies, decide how to live in the changed landscape of new possibilities. To understand why we have less privacy today than in the past, we must look not just at the gadgets. To be sure, we should be wary of spies and thieves, but we should also look at those who protect us and help us—and we should also take a good look in the mirror.

We are most conscious of our personal information winding up in the hands of strangers when we think about data loss or theft. Reports like the one about the British tax office have become fairly common. The theft of information about 45 million customers of TJX stores, described in Chapter 5, "Secret Bits," was even larger than the British catastrophe. In 2003, Scott Levine, owner of a mass email business named Snipermail, stole more than a billion personal information records from Acxiom. Millions of Americans are victimized by identity theft every year, at a total cost in the tens of billions of dollars annually. Many more of us harbor daily fears that just "a little bit" of our financial information has leaked out, and could be a personal time bomb if it falls into the wrong hands.

Why can't we just keep our personal information to ourselves? Why do so many other people have it in the first place, so that there is an opportunity for it to go astray, and an incentive for creative crooks to try to steal it?

We lose control of our personal information because of things we do to ourselves, and things others do to us. Of things we do to be ahead of the curve, and things we do because everyone else is doing them. Of things we do to save money, and things we do to save time. Of things we do to be safe from our enemies, and things we do because we feel invulnerable. Our loss of privacy is a problem, but there is no one answer to it, because there is no one reason why it is happening. It is a messy problem, and we first have to think about it one piece at a time.

We give away information about ourselves—voluntarily leave visible footprints of our daily lives—because we judge, perhaps without thinking about it very much, that the benefits outweigh the costs. To be sure, the benefits are many.

Saving Time

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 is also 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 from inside the car, is an RFID, powered with a battery so identifying information can be sent to the 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. All that 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—when, accurate to the second. It also shows where you got off and how far you went. Informing you of the mileage 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 as it was charging for 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 five 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 five 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!" These records can be subpoenaed, and have been, 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!

Saving Money

Sometimes it's money, not time, which motivates us to leave footprints. Such is the case with supermarket loyalty cards. If you do not want Safeway to keep track of the fact that you bought the 12-pack of Yodels despite your recent cholesterol results, you can make sure it doesn't know. You simply pay the "privacy tax"—the surcharge for customers not presenting a loyalty card. The purpose of loyalty cards is to enable merchants to track individual item purchases. (Item-level transactions are typically not tracked by credit card companies, which do not care if you bought Yodels instead of granola, so long as you pay the bill.) With loyalty cards, stores can capture details of cash transactions as well. They can process all the transaction data, and draw inferences about shoppers' habits. Then, if a lot of people who buy Yodels also buy Bison Brew Beer, the store's automated cash register can automatically spit out a discount coupon for Bison Brew as your Yodels are being bagged. A "discount" for you, and more sales for Safeway. Everybody wins. Don't they?

As grocery stores expand their web-based business, it is even easier for them to collect personal information about you. Reading the fine print when you sign up is a nuisance, but it is worth doing, so you understand what you are giving and what you are getting in return. Here are a few sentences of Safeway's privacy policy for customers who use its web site:

> Safeway may use personal information to provide you with newsletters, articles, product or service alerts, new product or service announcements, saving awards, event invitations, personally tailored coupons, program and promotional information and offers, and other information, which may be provided to Safeway by other companies. ... We may provide personal information to our partners and suppliers for customer support services and processing of personal information on behalf of Safeway. We may also share personal information with our affiliate companies, or in the course of an actual or potential sale, re-organization, consolidation, merger, or amalgamation of our business or businesses.

Dreary reading, but the language gives Safeway lots of leeway. Maybe you don't care about getting the junk mail. Not everyone thinks it is junk, and the

company does let you "opt out" of receiving it (although in general, few people bother to exercise opt-out rights). But Safeway has lots of "affiliates," and who knows how many companies with which it *might* be involved in a merger or sale of part of its business. Despite privacy concerns voiced by groups like C.A.S.P.I.A.N. (Consumers Against Supermarket Privacy Invasion and Numbering, `www.nocards.org`), most shoppers readily agree to have the data collected. The financial incentives are too hard to resist, and most consumers just don't worry about marketers knowing their purchases. But whenever purchases can be linked to your name, there is a record, somewhere in a huge database, of whether you use regular or super tampons, lubricated or unlubricated condoms, and whether you like regular beer or lite. You have authorized the company to share it, and even if you hadn't, the company could lose it accidentally, have it stolen, or have it subpoenaed.

Convenience of the Customer

The most obvious reason not to worry about giving information to a company is that you do business with them, and it is in your interest to see that they do their business with you better. You have no interest in whether they make more money from you, but you do have a strong interest in making it easier and faster for you to shop with them, and in cutting down the amount of stuff they may try to sell you that you would have no interest in buying. So your interests and theirs are, to a degree, aligned, not in opposition. Safeway's privacy policy states this explicitly: "Safeway Club Card information and other information may be used to help make Safeway's products, services, and programs more useful to its customers." Fair enough.

No company has been more progressive in trying to sell customers what they might want than the online store Amazon. Amazon suggests products to repeat customers, based on what they have bought before—or what they have simply looked at during previous visits to Amazon's web site. The algorithms are not perfect; Amazon's computers are drawing inferences from data, not being clairvoyant. But Amazon's guesses are pretty good, and recommending the wrong book every now and then is a very low-cost mistake. If Amazon does it too often, I might switch to Barnes and Noble, but there is no injury to me. So again: Why should anyone care that Amazon knows so much about me? On the surface, it seems benign. Of course, we don't want the credit card information to go astray, but who cares about knowing what books I have looked at online?

Our indifference is another marker of the fact that we are living in an exposed world, and that it feels very different to live here. In 1988, when a

How Sites Know Who You Are

1. **You tell them**. Log in to Gmail, Amazon, or eBay, and you are letting them know exactly who you are.
2. **They've left cookies on one of your previous visits**. A *cookie* is a small text file stored on your local hard drive that contains information that a particular web site wants to have available during your current session (like your shopping cart), or from one session to the next. Cookies give sites persistent information for tracking and personalization. Your browser has a command for showing cookies—you may be surprised how many web sites have left them!
3. **They have your IP address**. The web server has to know where you are so that it can ship its web pages to you. Your IP address is a number like 66.82.9.88 that locates your computer in the Internet (see the Appendix for details). That address may change from one day to the next. But in a residential setting, your Internet Service Provider (your *ISP*—typically your phone or cable company) knows who was assigned each IP address at any time. Those records are often subpoenaed in court cases.

If you are curious about who is using a particular IP address, you can check the American Registry of Internet Numbers (`www.arin.net`). Services such as `whatismyip.com`, `whatismyip.org`, and `ipchicken.com` also allow you to check your own IP address. And `www.whois.net` allows you to check who owns a domain name such as `harvard.com`—which turns out to be the Harvard Bookstore, a privately owned bookstore right across the street from the university. Unfortunately, that information won't reveal who is sending you spam, since spammers routinely forge the source of email they send you.

videotape rental store clerk turned over Robert Bork's movie rental records to a Washington, DC newspaper during Bork's Supreme Court confirmation hearings, Congress was so outraged that it quickly passed a tough privacy protection bill, The Video Privacy Protection Act. Videotape stores, if any still exist, can be fined simply for keeping rental records too long. Twenty years later, few seem to care much what Amazon does with its millions upon millions of detailed, fine-grained views into the brains of all its customers.

It's Just Fun to Be Exposed

Sometimes, there can be no explanation for our willing surrender of our privacy except that we take joy in the very act of exposing ourselves to public

view. Exhibitionism is not a new phenomenon. Its practice today, as in the past, tends to be in the province of the young and the drunk, and those wishing to pretend they are one or the other. That correlation is by no means perfect, however. A university president had to apologize when an image of her threatening a Hispanic male with a stick leaked out from her MySpace page, with a caption indicating that she had to "beat off the Mexicans because they were constantly flirting with my daughter."

And there is a continuum of outrageousness. The less wild of the party photo postings blend seamlessly with the more personal of the blogs, where the bloggers are chatting mostly about their personal feelings. Here there is not exuberance, but some simpler urge for human connectedness. That passion, too, is not new. What *is* new is that a photo or video or diary entry, once posted, is visible to the entire world, and that there is no taking it back. Bits don't fade and they don't yellow. Bits are forever. And we don't know how to live with that.

Bits don't fade and they don't yellow. Bits are forever. And we don't know how to live with that.

For example, a blog selected with no great design begins:

> This is the personal web site of Sarah McAuley. ... I think sharing my life with strangers is odd and narcissistic, which of course is why I'm addicted to it and have been doing it for several years now. Need more? You can read the "About Me" section, drop me an email, or you know, just read the drivel that I pour out on an almost-daily basis.

No thank you, but be our guest. Or consider that there is a Facebook group just for women who want to upload pictures of themselves uncontrollably drunk. Or the Jennicam, through which Jennifer Kay Ringley opened her life to the world for seven years, setting a standard for exposure that many since have surpassed in explicitness, but few have approached in its endless ordinariness. We are still experimenting, both the voyeurs and viewed.

Because You Can't Live Any Other Way

Finally, we give up data about ourselves because we don't have the time, patience, or single-mindedness about privacy that would be required to live our daily lives in another way. In the U.S., the number of credit, debit, and bank cards is in the billions. Every time one is used, an electronic handshake records a few bits of information about who is using it, when, where, and for what. It is now virtually unheard of for people to make large purchases of

ordinary consumer goods with cash. Personal checks are going the way of cassette tape drives, rendered irrelevant by newer technologies. Even if you could pay cash for everything you buy, the tax authorities would have you in their databases anyway. There even have been proposals to put RFIDs in currency notes, so that the movement of cash could be tracked.

Only sects such as the Amish still live without electricity. It will soon be almost that unusual to live without Internet connectivity, with all the fingerprints it leaves of your daily searches and logins and downloads. Even the old dumb TV is rapidly disappearing in favor of digital communications. Digital TV will bring the advantages of video on demand—no more trips to rent movies or waits for them to arrive in the mail—at a price: Your television service provider will record what movies you have ordered. It will be so attractive to be able to watch what we want when we want to watch it, that we won't miss either the inconvenience or the anonymity of the days when all the TV stations washed your house with their airwaves. You couldn't pick the broadcast times, but at least no one knew which waves you were grabbing out of the air.

Little Brother Is Watching

So far, we have discussed losses of privacy due to things for which we could, in principle anyway, blame ourselves. None of us really needs a loyalty card, we should always read the fine print when we rent a car, and so on. We would all be better off saying "no" a little more often to these privacy-busters, but few of us would choose to live the life of constant vigilance that such resolute denial would entail. And even if we were willing to make those sacrifices, there are plenty of other privacy problems caused by things others do to us.

The snoopy neighbor is a classic American stock figure—the busybody who watches how many liquor bottles are in your trash, or tries to figure out whose Mercedes is regularly parked in your driveway, or always seems to know whose children were disorderly last Saturday night. But in Cyberspace, we are all neighbors. We can all check up on each other, without even opening the curtains a crack.

Public Documents Become VERY Public

Some of the snooping is simply what anyone could have done in the past by paying a visit to the Town Hall. Details that were always public—but inaccessible—are quite accessible now.

In 1975, Congress created the Federal Election Commission to administer the Federal Election Campaign Act. Since then, all political contributions have been public information. There is a difference, though, between "public" and "readily accessible." Making public data available on the Web shattered the veil of privacy that came from inaccessibility.

Want to know who gave money to Al Franken for Senate? Lorne Michaels from Saturday Night Live, Leonard Nimoy, Paul Newman, Craig Newmark (the "craig" of craigslist.com), and Ginnie W., who works with us and may not have wanted us to know her political leanings. Paul B., and Henry G., friends of ours, covered their bases by giving to both Obama and Clinton.

The point of the law was to make it easy to look up big donors. But since data is data, what about checking on your next-door neighbors? Ours definitely leaned toward Obama over Clinton, with no one in the Huckabee camp. Or your clients? One of ours gave heartily to Dennis Kucinich. Or your daughter's boyfriend? You can find out for yourself, at www.fec.gov or fundrace.huffingtonpost.com. We're not telling about our own.

Hosts of other facts are now available for armchair browsing—facts that in the past were nominally public but required a trip to the Registrar of Deeds. If you want to know what you neighbor paid for their house, or what it's worth today, many communities put all of their real estate tax rolls online. It was always public; now it's accessible. It was never wrong that people could get this information, but it feels very different now that people can browse through it from the privacy of their home.

If you are curious about someone, you can try to find him or her on Facebook, MySpace, or just using an ordinary search engine. A college would not peek at the stupid Facebook page of an applicant, would it? Absolutely not, says the Brown Dean of Admissions, "unless someone says there's something we should look at."

New participatory websites create even bigger opportunities for information-sharing. If you are about to go on a blind date, there are special sites just for that. Take a look at www.dontdatehimgirl.com, a social networking site with a self-explanatory focus. When we checked, this warning about one man had just been posted, along with his name and photograph: "Compulsive womanizer, liar, internet cheater; pathological liar who can't be trusted as a friend much less a boyfriend. Total creep! Twisted and sick—needs mental help. Keep your daughter away from this guy!" Of course, such information may be worth exactly what we paid for it. There is a similar site, www.platewire.com, for reports about bad drivers. If you are not dating or driving, perhaps you'd like to check out a neighborhood before you move in, or just register a public warning about the obnoxious revelers who live next door to you. If so, www.rottenneighbor.com is the site for you. When we

typed in the zip code in which one of us lives, a nice Google map appeared with a house near ours marked in red. When we clicked on it, we got this report on our neighbor:

> you're a pretty blonde, slim and gorgeous. hey, i'd come on to you if i weren't gay. you probably have the world handed to you like most pretty women. is that why you think that you are too good to pick up after your dog? you know that you are breaking the law as well as being disrespectful of your neighbors. well, i hope that you step in your own dogs poop on your way to work, or on your way to dinner. i hope that the smell of your self importance follows you all day.

For a little money, you can get a lot more information. In January 2006, John Aravosis, creator of `Americablog.com`, purchased the detailed cell phone records of General Wesley Clark. For $89.95, he received a listing of all of Clark's calls for a three-day period. There are dozens of online sources for this kind of information. You might think you'd have to be in the police or the FBI to find out who people are calling on their cell phones, but there are handy services that promise to provide anyone with that kind of information for a modest fee. The *Chicago Sun Times* decided to put those claims to a test, so it paid $110 to `locatecell.com` and asked for a month's worth of cell phone records of one Frank Main, who happened to be one of its own reporters. The *Sun Times* did it all with a few keystrokes—provided the telephone number, the dates, and a credit card number. The request went in on Friday of a long weekend, and on Tuesday morning, a list came back in an email. The list included 78 telephone numbers the reporter had called—sources in law enforcement, people he was writing stories about, and editors in the newspaper. It was a great service for law enforcement—except that criminals can use it too, to find out whom the detectives are calling. These incidents stimulated passage of the Telephone Records and Privacy Act of 2006, but in early 2008, links on `locatecell.com` were still offering to help "find cell phone records in seconds," and more.

If cell phone records are not enough information, consider doing a proper background check. For $175, you can sign up as an "employer" with ChoicePoint and gain access to reporting services including criminal records, credit history, motor vehicle records, educational verification, employment verification, Interpol, sexual offender registries, and warrants searchers—they are all there to be ordered, with *a la carte* pricing. Before we moved from paper to bits, this information was publicly available, but largely inaccessible. Now, all it takes is an Internet connection and a credit card. This is one

PERSONAL COMPUTER MONITORING SOFTWARE

PC Pandora (www.pcpandora.com) enables you to "know everything they do on your PC," such as "using secret email accounts, chatting with unknown friends, accessing secret dating profiles or even your private records." Using it, you can "find out about secret email accounts, chat partners, dating site memberships, and more."

Actual Spy (www.actualspy.com) is a "keylogger which allows you to find out what other users do on your computer in your absence. It is designed for the hidden computer monitoring and the monitoring of the computer activity. Keylogger Actual Spy is capable of catching all keystrokes, capturing the screen, logging the programs being run and closed, monitoring the clipboard contents."

of the most important privacy transformations. Information that was previously available only to professionals with specialized access or a legion of local workers is now available to everyone.

Then there is real spying. Beverly O'Brien suspected her husband was having an affair. If not a physical one, at a minimum she thought he was engaging in inappropriate behavior online. So, she installed some monitoring software. Not hard to do on the family computer, these packages are promoted as "parental control software"—a tool to monitor your child's activities, along with such other uses as employee monitoring, law enforcement, and to "catch a cheating spouse." Beverly installed the software, and discovered that her hapless hubby, Kevin, was chatting away while playing Yahoo! Dominoes. She was an instant spy, a domestic wire-tapper. The marketing materials for her software neglected to tell her that installing spyware that intercepts communications traffic was a direct violation of Florida's Security of Communications Act, and the trial court refused to admit any of the evidence in their divorce proceeding. The legal system worked, but that didn't change the fact that spying has become a relatively commonplace activity, the domain of spouses and employers, jilted lovers, and business competitors.

Idle Curiosity

There is another form of Little Brother-ism, where amateurs can sit at a computer connected to the Internet and just look for something interesting—not about their neighbors or husbands, but about anyone at all. With so much data out there, anyone can discover interesting personal facts, with the

investment of a little time and a little imagination. To take a different kind of example, imagine having your family's medical history re-identified from a paper in an online medical journal.

Figure 2.4 shows a map of the incidence of a disease, let's say syphilis, in a part of Boston. The "syphilis epidemic" in this illustration is actually a simulation. The data was just made up, but maps exactly like this have been common in journals for decades. Because the area depicted is more than 10 square kilometers, there is no way to figure out which house corresponds to a dot, only which neighborhood.

Source: John S. Brownstein, Christopher A. Cassa, Kenneth D. Mandl, No place to hide—reverse identification of patients from published maps, *New England Journal of Medicine*, 355:16, October 19, 2007, 1741-1742.

FIGURE 2.4 Map of part of Boston as from a publication in a medical journal, showing where a disease has occurred. (Simulated data.)

At least that was true in the days when journals were only print documents. Now journals are available online, and authors have to submit their

figures as high-resolution JPEGs. Figure 2.5 shows what happens if you download the published journal article from the journal's web site, blow up a small part of the image, and superimpose it on an easily available map of the corresponding city blocks. For each of the seven disease locations, there is only a single house to which it could correspond. Anyone could figure out where the people with syphilis live.

Source: John S. Brownstein, Christopher A. Cassa, Kenneth D. Mandl, No place to hide—reverse identification of patients from published maps, *New England Journal of Medicine*, 355:16, October 19, 2007, 1741-1742.

FIGURE 2.5 Enlargement of Figure 2.4 superimposed on a housing map of a few blocks of the city, showing that individual households can be identified to online readers, who have access to the high-resolution version of the epidemiology map.

This is a re-identification problem, like the one Latanya Sweeney noted when she showed how to get Governor Weld's medical records. There are things that can be done to solve this one. Perhaps the journal should not use such high-resolution images (although that could cause a loss of crispness, or even visibility—one of the nice things about online journals is that the visually impaired can magnify them, to produce crisp images at a very large scale). Perhaps the data should be "jittered" or "blurred" so what appears on the screen for illustrative purposes is intentionally incorrect in its fine details. There are always specific policy responses to specific re-identification scenarios.

Every scenario is a little different, however, and it is often hard to articulate sensible principles to describe what should be fixed.

In 2001, four MIT students attempted to re-identify Chicago homicide victims for a course project. They had extremely limited resources: no proprietary databases such as the companies that check credit ratings possess, no access to government data, and very limited computing power. Yet they were able to identify nearly 8,000 individuals from a target set of 11,000.

The source of the data was a free download from the Illinois Criminal Justice Authority. The primary reference data source was also free. The Social Security Administration provides a comprehensive death index including name, birth date, Social Security Number, zip code of last residence, date of death, and more. Rather than paying the nominal fee for the data (after all, they were students), these researchers used one of the popular genealogy web sites, RootsWeb.com, as a free source for the Social Security Death Index (SSDI) data. They might also have used municipal birth and death records, which are also publicly available.

The SSDI did not include gender, which was important to completing an accurate match. But more public records came to the rescue. They found a database published by the census bureau that enabled them to infer gender from first names—most people named "Robert" are male, and most named "Susan" are female. That, and some clever data manipulation, was all it took. It is far from clear that it was wrong for any particular part of these data sets to be publicly available, but the combination revealed more than was intended.

The more re-identification problems we see, and the more *ad hoc* solutions we develop, the more we develop a deep-set fear that our problems may never end. These problems arise because there is a great deal of public data, no one piece of which is problematic, but which creates privacy violations in combination. It is the opposite of what we know about salt—that the component elements, sodium and chlorine, are both toxic, but the compound itself is safe. Here we have toxic compounds arising from the clever combination of harmless components. What can possibly be done about *that?*

Big Brother, Abroad and in the U.S.

Big Brother really is watching today, and his job has gotten much easier because of 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*, the cards will 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 case of street protests or any individual activity 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. Chinese tracking is far more detailed and pervasive than Britain's ubiquitous surveillance cameras.

You Pay for the Mike, We'll Just Listen In

Planting tiny microphones where they might pick up conversations of underworld figures used to be risky work for federal authorities. There are much safer alternatives now that many people carry their own radio-equipped microphones with them all the time.

Many cell phones can be reprogrammed remotely so that the microphone is always on and the phone is transmitting, even if you think you have powered it off. The FBI used this technique in 2004 to listen to John Tomero's conversations with other members of his organized crime family. A federal court ruled that this "roving bug," installed after due authorization, constituted a legal from of wiretapping. Tomero could have prevented it by removing the battery, and now some nervous business executives routinely do exactly that.

The microphone in a General Motors car equipped with the OnStar system can also be activated remotely, a feature that can save lives when OnStar operators contact the driver after receiving a crash signal. OnStar warns, "OnStar will cooperate with official court orders regarding criminal investigations from law enforcement and other agencies," and indeed, the FBI has used this method to eavesdrop on conversations held inside cars. In one case, a federal court ruled against this way of collecting evidence—but not on privacy grounds. The roving bug disabled the normal operation of OnStar, and the court simply thought that the FBI had interfered with the vehicle owner's contractual right to chat with the OnStar operators!

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 causes a furious libertarian reaction from parties not usually outspoken in defense of individual freedom. Under the REAL ID act of 2005, uniform federal standards are being implemented for state-issued drivers' licenses. Although it passed through Congress without debate, the law is opposed by at least 18 states. Resistance pushed back the implementation timetable first to 2009, and then, in early 2008, to 2011. 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 U.S., 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 the shape of their earlobes. 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 U.S., 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 55 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 is there 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.'" No REAL ID controversies, goes the theory; all the data being gathered would, in some sense at least, be offered voluntarily.

Friendly Cooperation Between Big Siblings

In fact, there are two Big Brothers, who often work together. And 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 governmental data sources together with private and commercial ones, creating comprehensive digital dossiers on millions of people. With their massive data storage and processing power, they can make connections in the data, like the clever connections the MIT students made with the Chicago homicide data, but 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 our insurance rates, or simply help us be sure that 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—in 2004, this amounted to more than a billion transactions a day. The company uses its massive 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 they gather and analyze 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, simply because of 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. Thirty 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, while 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 or 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 web site would be illegal. Yet 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 congressmen, 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 it 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.

In spite of its 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 inter-agency 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 they can both buy it 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. Yet 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 their survey of 128 federal departments. They 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 twelve 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.

The Transportation Security Administration (TSA) is responsible for airline passenger screening. One proposed system, CAPPS II, which was ultimately terminated over privacy concerns, sought to bring together disparate data sources to determine whether a particular individual might pose a transportation threat. Color-coded assessment tags would determine whether you could board quickly, be subject to further screening, or denied access to air travel.

The government creates projects, the media and civil liberties groups raise serious privacy concerns, the projects are cancelled, and new ones 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.

Technology Change and Lifestyle Change

New technologies enable new kinds of social interactions. There were no suburban shopping malls before private automobiles became cheap and widely used. Thirty years ago, many people getting off an airplane reached for cigarettes; today, they reach for cell phones. As Heraclitus is reported to have said 2,500 years ago, "all is flux"—everything keeps changing. The reach-for-your-cell phone gesture may not last much longer, since airlines are starting to provide onboard cell phone coverage.

The more people use a new technology, the more useful it becomes. (This is called a "network effect"; see Chapter 4, "Needles in the Haystack.") When one of us got the email address `lewis@harvard` as a second-year graduate student, it was a vainglorious joke; all the people he knew who had email addresses were students in the same office with him. Email culture could not develop until a lot of people had email, but there wasn't much point in having email if no one else did.

Technology changes and social changes reinforce each other. Another way of looking at the technological reasons for our privacy loss is to recognize that the social institutions enabled by the technology are now more important than the practical uses for which the technology was originally conceived. Once a lifestyle change catches on, we don't even think about what it depends on.

Credit Card Culture

The usefulness of the data aggregated by Acxiom and its kindred data aggregation services rises as the number of people in their databases goes up, and as larger parts of their lives leave traces in those databases. When credit cards were mostly short-term loans taken out for large purchases, the credit card data was mostly useful for determining your creditworthiness. It is still useful for that, but now that many people buy virtually everything with credit cards, from new cars to fast-food hamburgers, the credit card transaction database can be mined for a detailed image of our lifestyles. The information is there, for example, to determine if you usually eat dinner out, how much traveling you do, and how much liquor you tend to consume. Credit card companies do in fact analyze this sort of information, and we are glad they do. If you don't seem to have been outside Montana in your entire life and you turn up buying a diamond bracelet in Rio de Janeiro, the credit card company's computer notices the deviation from the norm, and someone may call to be sure it is really you.

The credit card culture is an economic problem for many Americans, who accept more credit card offers than they need, and accumulate more debt than they should. But it is hard to imagine the end of the little plastic cards, unless even smaller RFID tags replace them. Many people carry almost no cash today, and with every easy swipe, a few more bits go into the databases.

Email Culture

Email is culturally in between telephoning and writing a letter. It is quick, like telephoning (and instant messaging is even quicker). It is permanent, like a letter. And like a letter, it waits for the recipient to read it. Email has, to a great extent, replaced both of the other media for person-to-person communication, because it has advantages of both. But it has the problems that other communication methods have, and some new ones of its own.

Phone calls are not intended to last forever, or to be copied and redistributed to dozens of other people, or to turn up in court cases. When we use email as though it were a telephone, we tend to forget about what else might happen to it, other than the telephone-style use, that the recipient will read it and throw it away. Even Bill Gates probably wishes that he had written his corporate emails in a less telephonic voice. After testifying in an antitrust lawsuit that he had not contemplated cutting a deal to divide the web browser market with a competitor, the government produced a candid email he had sent, seeming to contradict his denial: "We could even pay them money as part of the deal, buying a piece of them or something."

Email is as public as postcards, unless it is encrypted, which it usually is not.

Email is bits, traveling within an ISP and through the Internet, using email software that may keep copies, filter it for spam, or submit it to any other form of inspection the ISP may choose. If your email service provider is Google, the point of the inspection is to attach some appropriate advertising. If you are working within a financial services corporation, your emails are probably logged—even the ones to your grandmother—because the company has to be able to go back and do a thorough audit if something inappropriate happens.

Email is as public as postcards, unless it is encrypted, which it usually is not. Employers typically reserve the right to read what is sent through company email. Check the policy of your own employer; it may be hard to find, and it may not say what you expect. Here is Harvard's policy, for example:

> Employees must have no expectation or right of privacy in anything they create, store, send, or receive on Harvard's computers, networks, or telecommunications systems. .... Electronic files, e-mail, data files, images, software, and voice mail may be accessed at any time by management or by other authorized personnel for any business purpose. Access may be requested and arranged through the system(s) user, however, this is not required.

Employers have good reason to retain such sweeping rights; they have to be able to investigate wrongdoing for which the employer would be liable. As a result, such policies are often less important than the good judgment and ethics of those who administer them. Happily, Harvard's are generally good. But as a general principle, the more people who have the authority to snoop, the more likely it is that someone will succumb to the temptation.

Commercial email sites can retain copies of messages even after they have been deleted. And yet, there is very broad acceptance of public, free, email services such as Google's Gmail, Yahoo! Mail, or Microsoft's Hotmail. The technology is readily available to make email private: whether you use encryption tools, or secure email services such as Hushmail, a free, web-based email service that incorporates PGP-based encryption (see Chapter 5). The usage of these services, though, is an insignificant fraction of their unencrypted counterparts. Google gives us free, reliable email service and we, in return, give up some space on our computer screen for ads. Convenience and cost trump privacy. By and large, users don't worry that Google, or its competitors, have all their mail. It's a bit like letting the post office keep a copy of every letter you send, but we are so used to it, we don't even think about it.

Web Culture

When we send an email, we think at least a *little* bit about the impression we are making, because we are sending it to a human being. We may well say things we would not say face-to-face, and live to regret that. Because we can't see anyone's eyes or hear anyone's voice, we are more likely to overreact and be hurtful, angry, or just too smart for our own good. But because email is directed, we don't send email thinking that no one else will ever read what we say.

The Web is different. Its social sites inherit their communication culture not from the letter or telephone call, but from the wall in the public square, littered with broadsides and scribbled notes, some of them signed and some not. Type a comment on a blog, or post a photo on a photo album, and your action can be as anonymous as you wish it to be—you do not know to whom your message is going. YouTube has millions of personal videos. Photo-archiving sites are the shoeboxes and photo albums of the twenty-first century. Online backup now provides easy access to permanent storage for the contents of our personal computers. We entrust commercial entities with much of our most private information, without apparent concern. The generation that has grown up with the Web has embraced social networking in all its varied forms: MySpace, YouTube, LiveJournal, Facebook, Xanga, Classmates.com, Flickr, dozens more, and blogs of every shape and size. More than being taken, personal privacy has been given away quite freely, because everyone else is doing it—the surrender of privacy is more than a way to social connectedness, it is a social institution in its own right. There are 70 million bloggers sharing everything from mindless blather to intimate personal details. Sites like `www.loopt.com` let you find your friends, while `twitter.com` lets you tell the entire world where you are and what you are doing. The Web is a confused, disorganized, chaotic realm, rich in both gold and garbage.

The "old" web, "Web 1.0," as we now refer to it, was just an information resource. You asked to see something, and you got to see it. Part of the disinhibition that happens on the new "Web 2.0" social networking sites is due to the fact that they still allow the movie-screen illusion—that we are "just looking," or if we are contributing, we are not leaving footprints or fingerprints if we use pseudonyms. (See Chapter 4 for more on Web 1.0 and Web 2.0.)

But of course, that is not really the way the Web ever worked. It is important to remember that even Web 1.0 was never anonymous, and even "just looking" leaves fingerprints.

In July 2006, a *New York Times* reporter called Thelma Arnold of Lilburn, Georgia. Thelma wasn't expecting the call. She wasn't famous, nor was she involved in anything particularly noteworthy. She enjoyed her hobbies, helped her friends, and from time to time looked up things on the Web—stuff about her dogs, and her friends' ailments.

Then AOL, the search engine she used, decided to release some "anonymous" query data. Thelma, like most Internet users, may not have known that AOL had kept every single topic that she, and every other one of their users, had asked about. But it did. In a moment of unenlightened generosity, AOL released for research use a small sample: about 20 million queries from 658,000 different users. That is actually not a lot of data by today's standards. For example, in July 2007, there were about 5.6 billion search engine queries, of which roughly 340 million were AOL queries. So, 20 million queries comprise only a couple of days' worth of search queries. In an effort to protect their clients' privacy, AOL "de-identified" the queries. AOL never mentioned anyone by name; they used random numbers instead. Thelma was 4417149. AOL mistakenly presumed that removing a single piece of personal identification would make it hard to figure out who the users were. It turned out that for some of the users, it wasn't hard at all.

It didn't take much effort to match Thelma with her queries. She had searched for "landscapers in Lilburn, GA" and several people with the last name "Arnold," leading to the obvious question of whether there were any Arnolds in Lilburn. Many of Thelma's queries were not particularly useful for identifying her, but were revealing nonetheless: "dry mouth," "thyroid," "dogs that urinate on everything," and "swing sets."

Thelma was not the only person to be identified. User 22690686 (Terri) likes astrology, and the Edison National Bank, Primerica, and Budweiser. 5779844 (Lawanna) was interested in credit reports, and schools. From what he searched for, user 356693 seems to have been an aide to Chris Shays, Congressman from Connecticut.

One of the privacy challenges that we confront as we rummage through the rubble of the digital explosion is that information exists without context. Was Thelma Arnold suffering from a wide range of ailments? One might readily conclude that from her searches. The fact is that she often tried to help her friends by understanding their medical problems.

Or consider AOL user 17556639, whose search history was released along with Thelma Arnold's. He searched for the following:

how to kill your wife 23 Mar, 22:09
wife killer 23 Mar, 22:11
poop 23 Mar, 22:12
dead people 23 Mar, 22:13
pictures of dead people 23 Mar, 22:15
killed people 23 Mar, 22:16
dead pictures 23 Mar, 22:17
murder photo 23 Mar, 22:20
steak and cheese 23 Mar, 22:22
photo of death 23 Mar, 22:30
death 23 Mar, 22:33
dead people photos 23 Mar, 22:33
photo of dead people 23 Mar, 22:35
www.murderdpeople.com 23 Mar, 22:37
decapitated photos 23 Mar, 22:39
car crashes3 23 Mar, 22:40
car crash photo 23 Mar, 22:41

Is this AOL user a potential criminal? Should AOL have called the police? Is 17556639 about to kill his wife? Is he (or she) a researcher with a spelling problem and an interest in Philly cheese steak? Is reporting him to the police doing a public service, or is it an invasion of privacy?

There is no way to tell just from these queries if this user was contemplating some heinous act or doing research for a novel that involves some grisly scenes. When information is incomplete and decontextualized, it is hard to judge meaning and intent.

In this particular case, we happen to know the answer. The user, Jason from New Jersey, was just fooling around, trying to see if Big Brother was watching. He wasn't planning to kill his wife at all. Inference from incomplete data has the problem of false positives—thinking you have something that you don't, because there are other patterns that fit the same data.

Information without context often leads to erroneous conclusions. Because our digital trails are so often retrieved outside the context within which they were created, they sometimes suggest incorrect interpretations. Data interpretation comes with balanced social responsibilities, to protect society when there is evidence of criminal behavior or intent, and also to protect the individual when such evidence is too limited to be reliable. Of course, for every example of misleading and ambiguous data, someone will want to solve the problems it creates by collecting more data, rather than less.

Beyond Privacy

There is nothing new under the sun, and the struggles to define and enforce privacy are no exception. Yet history shows that our concept of privacy has evolved, and the law has evolved with it. With the digital explosion, we have arrived at a moment where further evolution will have to take place rather quickly.

Leave Me Alone

More than a century ago, two lawyers raised the alarm about the impact technology and the media were having on personal privacy:

> Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops."

This statement is from the seminal law review article on privacy, published in 1890 by Boston attorney Samuel Warren and his law partner, Louis Brandeis, later to be a justice of the U.S. Supreme Court. Warren and Brandeis went on, "Gossip is no longer the resource of the idle and of the vicious, but has become a trade, which is pursued with industry as well as effrontery. To satisfy a prurient taste the details of sexual relations are spread broadcast in the columns of the daily papers. To occupy the indolent, column upon column is filled with idle gossip, which can only be procured by intrusion upon the domestic circle." New technologies made this garbage easy to produce, and then "the supply creates the demand."

And those candid photographs and gossip columns were not merely tasteless; they were bad. Sounding like modern critics of mindless reality TV, Warren and Brandeis raged that society was going to hell in a handbasket because of all that stuff that was being spread about.

> Even gossip apparently harmless, when widely and persistently circulated, is potent for evil. It both belittles and perverts. It belittles by inverting the relative importance of things, thus dwarfing the thoughts and aspirations of a people. When personal gossip attains the dignity of print, and crowds the space available for matters of

> real interest to the community, what wonder that the ignorant and thoughtless mistake its relative importance. Easy of comprehension, appealing to that weak side of human nature which is never wholly cast down by the misfortunes and frailties of our neighbors, no one can be surprised that it usurps the place of interest in brains capable of other things. Triviality destroys at once robustness of thought and delicacy of feeling. No enthusiasm can flourish, no generous impulse can survive under its blighting influence.

The problem they perceived was that it was hard to say just why such invasions of privacy should be unlawful. In individual cases, you could say something sensible, but the individual legal decisions were not part of a general regime. The courts had certainly applied legal sanctions for defamation—publishing malicious gossip that was false—but then what about malicious gossip that was true? Other courts had imposed penalties for publishing an individual's private letters—but on the basis of property law, just as though the individual's horse had been stolen rather than the words in his letters. That did not seem to be the right analogy either. No, they concluded, such rationales didn't get to the nub. When something private is published about you, something has been taken from you, you are a victim of theft—but the thing stolen from you is part of your identity as a person. In fact, privacy was a right, they said, a "general right of the individual to be let alone." That right had long been in the background of court decisions, but the new technologies had brought this matter to a head. In articulating this new right, Warren and Brandeis were, they asserted, grounding it in the principle of "inviolate personhood," the sanctity of individual identity.

Privacy and Freedom

The Warren-Brandeis articulation of privacy as a right to be left alone was influential, but it was never really satisfactory. Throughout the twentieth century, there were simply too many good reasons for *not* leaving people alone, and too many ways in which people *preferred* not to be left alone. And in the U.S., First Amendment rights stood in the way of privacy rights. As a general rule, the government simply cannot stop me from saying *anything*. In particular, it usually cannot stop me from saying what I want about your private affairs. Yet the Warren-Brandeis definition worked well enough for a long time, because, as Robert Fano put it, "The pace of technological progress was for a long time sufficiently slow as to enable society to learn pragmatically how to exploit new technology and prevent its abuse, with society maintaining its equilibrium most of the time." By the late 1950s, the emerging

electronic technologies, both computers and communication, had destroyed that balance. Society could no longer adjust pragmatically, because surveillance technologies were developing too quickly.

The result was a landmark study of privacy by the Association of the Bar of the City of New York, which culminated in the publication, in 1967, of a book by Alan Westin, entitled *Privacy and Freedom*. (Fano was reviewing Westin's book when he painted the picture of social disequilibrium caused by rapid technological change.) Westin proposed a crucial shift of focus.

Brandeis and Warren had seen a loss of privacy as a form of personal injury, which might be so severe as to cause "mental pain and distress, far greater than could be inflicted by mere bodily injury." Individuals had to take responsibility for protecting themselves. "Each man is responsible for his own acts and omissions only." But the law had to provide the weapons with which to resist invasions of privacy.

Westin recognized that the Brandeis-Warren formulation was too absolute, in the face of the speech rights of other individuals and society's legitimate data-gathering practices. Protection might come not from protective shields, but from control over the uses to which personal information could be put. "Privacy," wrote Westin, "is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."

> ... what is needed is a structured and rational weighing process, with definite criteria that public and private authorities can apply in comparing the claim for disclosure or surveillance through new devices with the claim to privacy. The following are suggested as the basic steps of such a process: measuring the seriousness of the need to conduct surveillance; deciding whether there are alternative methods to meet the need; deciding what degree of reliability will be required of the surveillance instrument; determining whether true consent to surveillance has been given; and measuring the capacity for limitation and control of the surveillance if it is allowed.

So even if there were a legitimate reason why the government, or some other party, might know something about you, your right to privacy might limit what the knowing party could do with that information.

This more nuanced understanding of privacy emerged from the important social roles that privacy plays. Privacy is not, as Warren and Brandeis had it, the right to be isolated from society—privacy is a right that makes society work. Fano mentioned three social roles of privacy. First, "the right to maintain the privacy of one's personality can be regarded as part of the right of

self-preservation"—the right to keep your adolescent misjudgments and personal conflicts to yourself, as long as they are of no lasting significance to your ultimate position in society. Second, privacy is the way society allows deviations from prevailing social norms, given that no one set of social norms is universally and permanently satisfactory—and indeed, given that social progress requires social experimentation. And third, privacy is essential to the development of independent thought—it enables some decoupling of the individual from society, so that thoughts can be shared in limited circles and rehearsed before public exposure.

Privacy is the way society allows deviations from prevailing social norms, given that social progress requires social experimentation.

Privacy and Freedom, and the rooms full of disk drives that sprouted in government and corporate buildings in the 1960s, set off a round of soul-searching about the operational significance of privacy rights. What, in practice, should those holding a big data bank think about when collecting the data, handling it, and giving it to others?

Fair Information Practice Principles

In 1973, the Department of Health, Education, and Welfare issued "Fair Information Practice Principles" (FIPP), as follows:

- **Openness.** There must be no personal data record-keeping systems whose very existence is secret.
- **Disclosure.** There must be a way for a person to find out what information about the person is in a record and how it is used.
- **Secondary use.** There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent.
- **Correction.** There must be a way for a person to correct or amend a record of identifiable information about the person.
- **Security.** Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for its intended use and must take precautions to prevent misuses of the data.

These principles were proposed for U.S. medical data, but were never adopted. Nevertheless, they have been the foundation for many corporate privacy policies. Variations on these principles have been codified in international trade agreements by the Organization of Economic Cooperation and Development (OECD) in 1980, and within the European Union (EU) in 1995. In the United States, echoes of these principles can be found in some state laws, but federal laws generally treat privacy on a case by case or "sectorial" basis. The 1974 Privacy Act applies to interagency data transfers within the federal government, but places no limitations on data handling in the private sector. The Fair Credit Reporting Act applies only to consumer credit data, but does not apply to medical data. The Video Privacy Act applies only to videotape rentals, but not to "On Demand" movie downloads, which did not exist when the Act was passed! Finally, few federal or state laws apply to the huge data banks in the file cabinets and computer systems of cities and towns. American government is decentralized, and authority over government data is decentralized as well.

The U.S. is not lacking in privacy laws. But privacy has been legislated inconsistently and confusingly, and in terms dependent on technological contingencies. There is no national consensus on what should be protected, and how protections should be enforced. Without a more deeply informed collective judgment on the benefits and costs of privacy, the current legislative hodgepodge may well get worse in the United States.

U.S. Privacy Laws

The Council of Better Business Bureaus has compiled a "Review of Federal and State Privacy Laws":

`www.bbbonline.org/UnderstandingPrivacy/library/fed_statePrivLaws.pdf`

The state of Texas has also compiled a succinct summary of major privacy laws:

`www.oag.state.tx.us/notice/privacy_table.htm`.

The discrepancy between American and European data privacy standards threatened U.S. involvement in international trade, because an EU directive would prohibit data transfers to nations, such as the U.S., that do not meet the European "adequacy" standard for privacy protection. Although the U.S. sectorial approach continues to fall short of European requirements, in 2000 the European Commission created a "safe harbor" for American businesses with multinational operations. This allowed individual corporations to establish their practices are adequate with respect to seven principles, covering notice, choice, onward transfer, access, security, data integrity, and enforcement.

It is, unfortunately, too easy to debate whether the European omnibus approach is more principled than the U.S. piecemeal approach, when the real question is whether either approach accomplishes what we want it to achieve. The Privacy Act of 1974 assured us that obscure statements would be buried deep in the Federal Register, providing the required official notice about massive governmental data collection plans—better than nothing, but providing "openness" only in a narrow and technical sense. Most large corporations doing business with the public have privacy notices, and virtually no one reads them. Only 0.3% of Yahoo! users read its privacy notice in 2002, for example. In the midst of massive negative publicity that year when Yahoo! changed its privacy policy to allow advertising messages, the number of users who accessed the privacy policy rose only to 1%. None of the many U.S. privacy laws prevented the warrantless wiretapping program instituted by the Bush administration, nor the cooperation with it by major U.S. telecommunications companies.

Indeed, cooperation between the federal government and private industry seems more essential than ever for gathering information about drug trafficking and international terrorism, because of yet another technological development. Twenty years ago, most long-distance telephone calls spent at least part of their time in the air, traveling by radio waves between microwave antenna towers or between the ground and a communication satellite. Government eavesdroppers could simply listen in (see the discussion of Echelon in Chapter 5). Now many phone calls travel through fiber optic cables instead, and the government is seeking the capacity to tap this privately owned infrastructure.

High privacy standards have a cost. They can limit the public usefulness of data. Public alarm about the release of personal medical information has led to major legislative remedies. The Health Information Portability and Accountability Act (HIPAA) was intended both to encourage the use of electronic data interchange for health information, and to impose severe penalties for the disclosure of "Protected Health Information," a very broad category including not just medical histories but, for example, medical payments. The bill mandates the removal of anything that could be used to re-connect medical records to their source. HIPAA is fraught with problems in an environment of ubiquitous data and powerful computing. Connecting the dots by assembling disparate data sources makes it extremely difficult to achieve the level of anonymity that HIPAA sought to guarantee. But help is available, for a price, from a whole new industry of HIPAA-compliance advisors. If you search for HIPAA online, you will likely see advertisements for services that will help you protect your data, and also keep you out of jail.

Ever Read Those "I Agree" Documents?

Companies can do almost anything they want with your information, as long as you agree. It seems hard to argue with that principle, but the deck can be stacked against the consumer who is "agreeing" to the company's terms. Sears Holding Corporation (SHC), the parent of Sears, Roebuck and Kmart, gave consumers an opportunity to join "My Sears Holding Community," which the company describes as "something new, something different ... a dynamic and highly interactive online community ... where your voice is heard and your opinion matters." When you went online to sign up, the terms appeared in a window on the screen.

The scroll box held only 10 lines of text, and the agreement was 54 boxfuls long. Deep in the terms was a detail: You were allowing Sears to install software on your PC that "monitors all of the Internet behavior that occurs on the computer ..., including ... filling a shopping basket, completing an application form, or checking your ... personal financial or health information." So your computer might send your credit history and AIDS test results to SHC, and you said it was fine!

At the same time as HIPAA and other privacy laws have safeguarded our personal information, they are making medical research costly and sometimes impossible to conduct. It is likely that classic studies such as the Framingham Heart Study, on which much public policy about heart disease was founded, could not be repeated in today's environment of strengthened privacy rules. Dr. Roberta Ness, president of the American College of Epidemiology, reported that "there is a perception that HIPAA may even be having a negative effect on public health surveillance practices."

The European reliance on the Fair Information Practice Principles is often no more useful, in practice, than the American approach. Travel through London, and you will see many signs saying "Warning: CCTV in use" to meet the "Openness" requirement about the surveillance cameras. That kind of notice throughout the city hardly empowers the individual. After all, even Big Brother satisfied the FIPP Openness standard, with the ubiquitous notices that he was watching! And the "Secondary Use" requirement, that European citizens should be asked permission before data collected for one purpose is used for another, is regularly ignored in some countries, although compliance practices are a major administrative burden on European businesses and may cause European businesses at least to pause and think before "repurposing" data they have gathered. Sociologist Amitai Etzioni repeatedly asks European

audiences if they have *ever* been asked for permission to re-use data collected about them, and has gotten only a single positive response—and that was from a gentleman who had been asked by a U.S. company.

The five FIPP principles, and the spirit of transparency and personal control that lay behind them, have doubtless led to better privacy practices. But they have been overwhelmed by the digital explosion, along with the insecurity of the world and all the social and cultural changes that have occurred in daily life. Fred H. Cate, a privacy scholar at the Indiana University, characterizes the FIPP principles as almost a complete bust:

> Modern privacy law is often expensive, bureaucratic, burdensome, and offers surprisingly little protection for privacy. It has substituted individual control of information, which it in fact rarely achieves, for privacy protection. In a world rapidly becoming more global through information technologies, multinational commerce, and rapid travel, data protection laws have grown more fractured and protectionist. Those laws have become unmoored from their principled basis, and the principles on which they are based have become so varied and procedural, that our continued intonation of the FIPPS mantra no longer obscures the fact that this emperor indeed has few if any clothes left.

Privacy as a Right to Control Information

It is time to admit that we don't even really know what we want. The bits are everywhere; there is simply no locking them down, and no one really wants to do that anymore. The meaning of privacy has changed, and we do not have a good way of describing it. It is not the right to be left alone, because not even the most extreme measures will disconnect our digital selves from the rest of the world. It is not the right to keep our private information to ourselves, because the billions of atomic factoids don't any more lend themselves into binary classification, private or public.

The bits are everywhere; there is simply no locking them down, and no one really wants to do that anymore.

Reade Seligmann would probably value his privacy more than most Americans alive today. On Monday, April 17, 2006, Seligmann was indicted in connection with allegations that a 27-year-old performer had been raped at a party at a Duke fraternity house. He and several of his lacrosse teammates instantly became poster children for everything that is wrong with

American society—an example of national over-exposure that would leave even Warren and Brandeis breathless if they were around to observe it. Seligmann denied the charges, and at first it looked like a typical he-said, she-said scenario, which could be judged only on credibility and presumptions about social stereotypes.

But during the evening of that fraternity party, Seligmann had left a trail of digital detritus. His data trail indicated that he could not have been at the party long enough, or at the right time, to have committed the alleged rape. Time-stamped photos from the party showed that the alleged victim of his rape was dancing at 12:02 AM. At 12:24 AM, he used his ATM card at a bank, and the bank's computers kept records of the event. Seligmann used his cell phone at 12:25 AM, and the phone company tracked every call he made, just as your phone company keeps a record of every call you make and receive. Seligmann used his prox card to get into his dormitory room at 12:46 AM, and the university's computer kept track of his comings and goings, just as other computers keep track of every card swipe or RFID wave you and I make in our daily lives. Even during the ordinary movements of a college student going to a fraternity party, every step along the way was captured in digital detail. If Seligmann had gone to the extraordinary lengths necessary to avoid leaving digital fingerprints—not using a modern camera, a cell phone, or a bank, and living off campus to avoid electronic locks—his defense would have lacked important exculpatory evidence.

Which would we prefer—the new world with digital fingerprints everywhere and the constant awareness that we are being tracked, or the old world with few digital footprints and a stronger sense of security from prying eyes? And what is the point of even asking the question, when the world cannot be restored to its old information lock-down?

In a world that has moved beyond the old notion of privacy as a wall around the individual, we could instead regulate those who would inappropriately *use* information about us. If I post a YouTube video of myself dancing in the nude, I should expect to suffer some personal consequences. Ultimately, as Warren and Brandeis said, individuals have to take responsibility for their actions. But society has drawn lines in the past around which facts are relevant to certain decisions, and which are not. Perhaps, the border of privacy having become so porous, the border of relevancy could be stronger. As Daniel Weitzner explains:

> New privacy laws should emphasize usage restrictions to guard against unfair discrimination based on personal information, even if it's publicly available. For instance, a prospective employer might be able to find a video of a job applicant entering an AIDS clinic or a

> mosque. Although the individual might have already made such facts public, new privacy protections would preclude the employer from making a hiring decision based on that information and attach real penalties for such abuse.

In the same vein, it is not intrinsically wrong that voting lists and political contributions are a matter of public record. Arguably, they are essential to the good functioning of the American democracy. Denying someone a promotion because of his or her political inclinations *would be* wrong, at least for most jobs. Perhaps a nuanced classification of the ways in which others are allowed to use information about us would relieve some of our legitimate fears about the effects of the digital explosion.

In *The Transparent Society*, David Brin wrote:

> Transparency is not about eliminating privacy. It's about giving us the power to hold accountable those who would *violate* it. Privacy implies serenity at home and the right to be let alone. It may be irksome how much other people know about me, but I have no right to police their minds. On the other hand I care very deeply about what others *do* to me and to those I love. We all have a right to some place where we can feel safe.

Despite the very best efforts, and the most sophisticated technologies, we cannot control the spread of our private information. And we often want information to be made public to serve our own, or society's purposes.

Yet there can still be principles of accountability for the *misuse* of information. Some ongoing research is outlining a possible new web technology, which would help ensure that information is used appropriately even if it is known. Perhaps automated classification and reasoning tools, developed to help connect the dots in networked information systems, can be retargeted to limit inappropriate use of networked information. A continuing border war is likely to be waged, however, along an existing free speech front: the line separating my right to tell the truth about you from your right not to have that information used against you. In the realm of privacy, the digital explosion has left matters deeply unsettled.

Always On

In *1984,* the pervasive, intrusive technology could be turned off:

> As O'Brien passed the telescreen a thought seemed to strike him. He stopped, turned aside and pressed a switch on the wall. There was a sharp snap. The voice had stopped.
>
> Julia uttered a tiny sound, a sort of squeak of surprise. Even in the midst of his panic, Winston was too much taken aback to be able to hold his tongue.
>
> "You can turn it off!" he said.
>
> "Yes," said O'Brien, "we can turn it off. We have that privilege. ...Yes, everything is turned off. We are alone."

Sometimes we can still turn it off today, and should. But mostly we don't want to. We don't want to be alone; we want to be connected. We find it convenient to leave it on, to leave our footprints and fingerprints everywhere, so we will be recognized when we come back. We don't want to have to keep retyping our name and address when we return to a web site. We like it when the restaurant remembers our name, perhaps because our phone number showed up on caller ID and was linked to our record in their database. We appreciate buying grapes for $1.95/lb instead of $3.49, just by letting the store know that we bought them. We may want to leave it on for ourselves because we know it is on for criminals. Being watched reminds us that they are watched as well. Being watched also means we are being watched over.

And perhaps we don't care that so much is known about us because that is the way human society used to be—kinship groups and small settlements, where knowing everything about everyone else was a matter of survival. Having it on all the time may resonate with inborn preferences we acquired millennia ago, before urban life made anonymity possible. Still today, privacy means something very different in a small rural town than it does on the Upper East Side of Manhattan.

We cannot know what the cost will be of having it on all the time. Just as troubling as the threat of authoritarian measures to restrict personal liberty is the threat of voluntary conformity. As Fano astutely observed, privacy allows limited social experimentation—the deviations from social norms that are much riskier to the individual in the glare of public exposure, but which can be, and often have been in the past, the leading edges of progressive social changes. With it always on, we may prefer not to try anything unconventional, and stagnate socially by collective inaction.

For the most part, it is too late, realistically, ever to turn it off. We may once have had the privilege of turning it off, but we have that privilege no more. We have to solve our privacy problems another way.

The digital explosion is shattering old assumptions about who knows what. Bits move quickly, cheaply, and in multiple perfect copies. Information that used to be public in principle—for example, records in a courthouse, the price you paid for your house, or stories in a small-town newspaper—is now available to everyone in the world. Information that used to be private and available to almost no one—medical records and personal snapshots, for example—can become equally widespread through carelessness or malice. The norms and business practices and laws of society have not caught up to the change.

The oldest durable communication medium is the written document. Paper documents have largely given way to electronic analogs, from which paper copies are produced. But are electronic documents really like paper documents? Yes and no, and misunderstanding the document metaphor can be costly. That is the story to which we now turn.