don't offer privacy as a market differentiating fea-
ceptions. DuckDuckGo is a search engine whose
ves around *not* tracking its users. Wickr offers
Ello is a social network that doesn't track its users.
r as big as their established competitors, but they're
l new ones are opening up shop all the time.
rising importance of customer and user privacy in
r of corporations with chief privacy officers: senior
for managing the legal and reputational risk of the
rporation holds. These executives have their own
rnational Association of Privacy Professionals, and
and regulations even in the absence of government
g this because it's good for business.

10

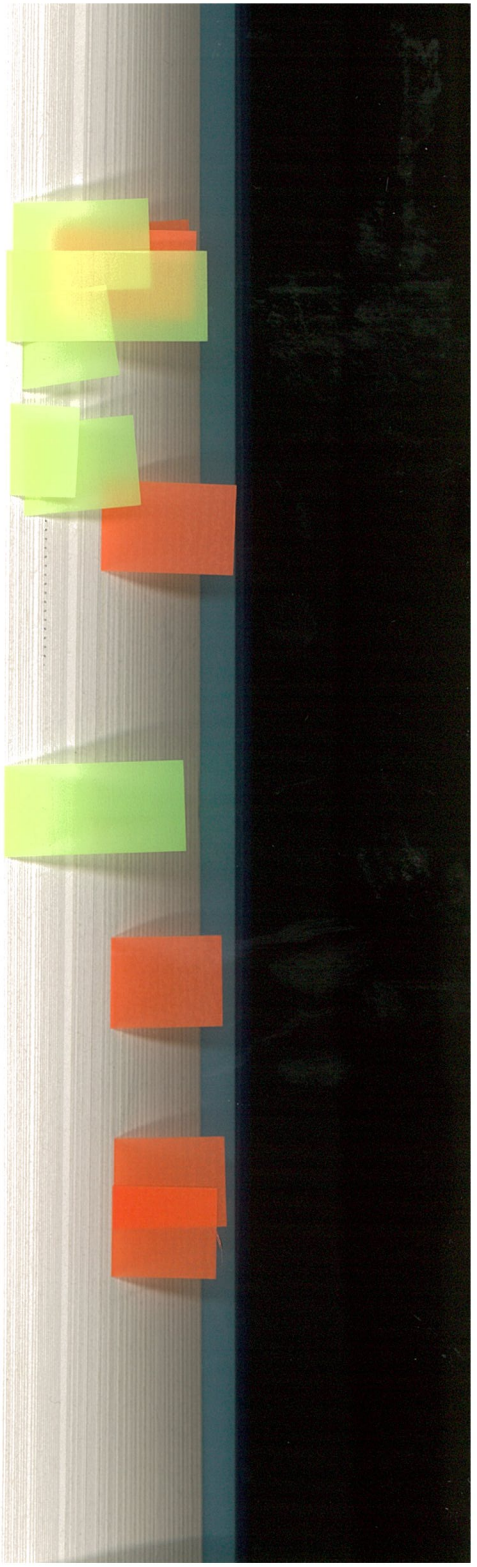
Privacy

The most common misconception about privacy is that it's about having something to hide. "If you aren't doing anything wrong, then you have nothing to hide," the saying goes, with the obvious implication that privacy only aids wrongdoers.

If you think about it, though, this makes no sense. We do nothing wrong when we make love, go to the bathroom, or sing in the shower. We do nothing wrong when we search for a job without telling our current employer. We do nothing wrong when we seek out private places for reflection or conversation, when we choose not to talk about something emotional or personal, when we use envelopes for our mail, or when we confide in a friend and no one else.

Moreover, even those who say that don't really believe it. In a 2009 interview, Google CEO Eric Schmidt put it this way: "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place." But in 2005, Schmidt banned employees from talking to reporters at CNET because a reporter disclosed personal details about Schmidt in an article. Facebook's Mark Zuckerberg declared in 2010 that privacy is no longer a "social norm," but bought the four houses abutting his Palo Alto home to help ensure his own privacy.

There are few secrets we don't tell *someone*, and we continue to believe



something is private even after we've told that person. We write intimate letters to lovers and friends, talk to our doctors about things we wouldn't tell anyone else, and say things in business meetings we wouldn't say in public. We use pseudonyms to separate our professional selves from our personal selves, or to safely try out something new.

Facebook's CEO Mark Zuckerberg showed a remarkable naïveté when he stated, "You have one identity. The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly. Having two identities for yourself is an example of a lack of integrity."

We're not the same to everyone we know and meet. We act differently when we're with our families, our friends, our work colleagues, and so on. We have different table manners at home and at a restaurant. We tell different stories to our children than to our drinking buddies. It's not necessarily that we're lying, although sometimes we do; it's that we reveal different facets of ourselves to different people. This is something innately human. Privacy is what allows us to act appropriately in whatever setting we find ourselves. In the privacy of our home or bedroom, we can relax in a way that we can't when someone else is around.

Privacy is an inherent human right, and a requirement for maintaining the human condition with dignity and respect. It is about choice, and having the power to control how you present yourself to the world. Internet ethnographer danah boyd puts it this way: "Privacy doesn't just depend on agency; being able to achieve privacy is an expression *of* agency."

When we lose privacy, we lose control of how we present ourselves. We lose control when something we say on Facebook to one group of people gets accidentally shared with another, and we lose complete control when our data is collected by the government. "How did he know that?" we ask. How did I lose control of who knows about my traumatic childhood, my penchant for tasteless humor, or my vacation to the Dominican Republic? You may know this feeling: you felt it when your mother friended you on Facebook, or on any other social networking site that used to be just you and your friends. Privacy violations are intrusions.

There's a strong physiological basis for privacy. Biologist Peter Watts makes the point that a desire for privacy is innate: mammals in particular don't respond well to surveillance. We consider it a physical threat, because animals in the natural world are surveilled by predators. Surveillance makes us feel like prey, just as it makes the surveillors act like predators.

Psychologists, sociologists, philosophers, novelists, and technologists have all written about the effects of constant surveillance, or even just the perception of constant surveillance. Studies show that we are less healthy, both physically and emotionally. We have feelings of low self-esteem, depression, and anxiety. Surveillance strips us of our dignity. It threatens our very selves as individuals. It's a dehumanizing tactic employed in prisons and detention camps around the world.

Violations of privacy are not all equal. Context matters. There's a difference between a Transportation Security Administration (TSA) officer finding porn in your suitcase and your spouse finding it. There's a difference between the police learning about your drug use and your friends learning about it. And violations of privacy aren't all equally damaging. Those of us in marginal socioeconomic situations—and marginalized racial, political, ethnic, and religious groups—are affected more. Those of us in powerful positions who are subject to people's continued approval are affected more. The lives of some of us depend on privacy.

Our privacy is under assault from constant surveillance. Understanding how this occurs is critical to understanding what's at stake.

THE EPHEMERAL

Through most of history, our interactions and conversations have been ephemeral. It's the way we naturally think about conversation. Exceptions were rare enough to be noteworthy: a preserved diary, a stenographer transcribing a courtroom proceeding, a political candidate making a recorded speech.

This has changed. Companies have fewer face-to-face meetings. Friends socialize online. My wife and I have intimate conversations by text message. We all behave as if these conversations were ephemeral, but they're not. They're saved in ways we have no control over.

On-the-record conversations are hard to delete. Oliver North learned this way back in 1987, when messages he thought he had deleted turned out to have been saved by the White House PROFS Notes system, an early

form of e-mail. Bill Gates learned this a decade later, when his conversational e-mails were provided to opposing counsel as part of Microsoft's antitrust litigation discovery process. And over 100 female celebrities learned it in 2014, when intimate self-portraits—some supposedly deleted—were stolen from their iCloud accounts and shared further and wider than they had ever intended.

It's harder and harder to be ephemeral. Voice conversation is largely still unrecorded, but how long will that last? Retail store surveillance systems register our presence, even if we are doing nothing but browsing and even if we pay for everything in cash. Some bars record the IDs of everyone who enters. I can't even buy a glass of wine on an airplane with cash anymore. Pervasive life recorders will make this much worse.

Science fiction writer Charles Stross described this as the end of prehistory. We won't forget anything, because we'll always be able to retrieve it from some computer's memory. This is new to our species, and will be a boon to both future historians and those of us in the present who want better data for self-assessment and reflection.

Having everything recorded and permanently available will change us both individually and as a society. Our perceptions and memories aren't nearly as sharp as we think they are. We fail to notice things, even important things. We misremember, even things we are sure we recall correctly. We forget important things we were certain we never would. People who keep diaries know this; old entries can read as if they were written by someone else. I have already noticed how having a record of all of my e-mail going back two decades makes a difference in how I think about my personal past.

One-fourth of American adults have criminal records. Even minor infractions can follow people forever and have a huge impact on their lives—this is why many governments have a process for expunging criminal records after some time has passed. Losing the ephemeral means that everything you say and do will be associated with you forever.

Having conversations that disappear as soon as they occur is a social norm that allows us to be more relaxed and comfortable, and to say things we might not say if a tape recorder were running. Over the longer term, forgetting—and misremembering—is how we process our history. Forgetting is an important enabler of forgiving. Individual and social memory fades, and past hurts become less sharp; this helps us forgive past wrongs. I'm not convinced that my marriage would be improved by the ability to produce transcripts of old arguments. Losing the ephemeral will be an enormous social and psychological change, and not one that I think our society is prepared for.

ALGORITHMIC SURVEILLANCE

One of the common defenses of mass surveillance is that it's being done by algorithms and not people, so it doesn't compromise our privacy. That's just plain wrong.

The distinction between human and computer surveillance is politically important. Ever since Snowden provided reporters with a trove of top-secret documents, we've learned about all sorts of NSA word games. The word "collect" has a very special definition, according to the Department of Defense. It doesn't mean collect; it means that a person looks at, or analyzes, the data. In 2013, Director of National Intelligence James Clapper likened the NSA's trove of accumulated data to a library. All those books are stored on the shelves, but very few are actually read. "So the task for us in the interest of preserving security and preserving civil liberties and privacy is to be as precise as we possibly can be when we go in that library and look for the books that we need to open up and actually read."

Think of that friend of yours who has thousands of books in his house. According to this ridiculous definition, the only books he can claim to have collected are the ones he's read.

This is why Clapper asserts he didn't lie in a Senate hearing when he replied "no" to the question "Does the NSA collect any type of data at all on millions or hundreds of millions of Americans?" From the military's perspective, it's not surveillance until a human being looks at the data, even if algorithms developed and implemented by defense personnel or contractors have analyzed it many times over.

This isn't the first time we've heard this argument. It was central to Google's defense of its context-sensitive advertising in the early days of Gmail. Google's computers examine each individual e-mail and insert a content-related advertisement in the footer. But no human reads those Gmail messages, only a computer. As one Google executive told me pri-

vately in the early days of Gmail, "Worrying about a computer reading your e-mail is like worrying about your dog seeing you naked."

But it's not, and the dog example demonstrates why. When you're watched by a dog, you're not overly concerned, for three reasons. The dog can't understand or process what he's seeing in the same way another person can. The dog won't remember or base future decisions on what he's seeing in the same way another person can. And the dog isn't able to tell anyone—not a person or another dog—what he's seeing.

When you're watched by a computer, none of that dog analogy applies. The computer is processing what it sees, and basing actions on it. You might be told that the computer isn't saving the data, but you have no assurance that that's true. You might be told that the computer won't alert a person if it perceives something of interest, but you can't know whether that's true. You have no way of confirming that no person will perceive whatever decision the computer makes, and that you won't be judged or discriminated against on the basis of what the computer sees.

Moreover, when a computer stores your data, there's always a risk of exposure. Privacy policies could change tomorrow, permitting new use of old data without your express consent. Some hacker or criminal could break in and steal your data. The organization that has your data could use it in some new and public way, or sell it to another organization. The FBI could serve a National Security Letter on the data owner. On the other hand, there isn't a court in the world that can get a description of you naked from your dog.

The primary difference between a computer and a dog is that the computer communicates with other people and the dog does not—at least, not well enough to matter. Computer algorithms are written by people, and their output is used by people. And when we think of computer algorithms surveilling us or analyzing our personal data, we need to think about the people behind those algorithms. Whether or not anyone actually looks at our data, the very facts that (1) they could, and (2) they guide the algorithms that do, make it surveillance.

You know this is true. If you believed what Clapper said, then you wouldn't object to a camera in your bedroom—as long as there were rules governing when the police could look at the footage. You wouldn't object to being compelled to wear a government-issued listening device 24/7, as long as your bureaucratic monitors followed those same rules. If you do object, it's because you realize that the privacy harm comes from the automatic collection and algorithmic analysis, regardless of whether or not a person is directly involved in the process.

IDENTIFICATION AND ANONYMITY

We all have experience with identifying ourselves on the Internet. Some websites tie your online identity to your real identity: banks, websites for some government services, and so on. Some tie your online identity to a payment system—generally credit cards—and others to your bank account or cell phone. Some websites don't care about your real identity, and allow you to maintain a unique username just for that site. Many more sites could work that way. Apple's iTunes, for example, could be so designed that it doesn't know who you really are, just that you're authorized to access a particular set of audio and video files.

The means to perform identification and authentication include passwords, biometrics, and tokens. Many people, myself included, have written extensively about the various systems and their relative strengths and weaknesses. I'll spare you the details; the takeaway is that none of these systems is perfect, but all are generally good enough for their applications. Authentication basically works.

It works because the people involved want to be identified. You want to convince Hotmail that it's your account; you want to convince your bank that it's your money. And while you might not want AT&T to be able to tie all the Internet browsing you do on your smartphone to your identity, you do want the phone network to transmit your calls to you. All of these systems are trying to answer the following question: "Is this the person she claims to be?" That is why it's so easy to gather data about us online; most of it comes from sources where we've intentionally identified ourselves.

Attribution of anonymous activity to a particular person is a much harder problem. In this case, the person doesn't necessarily want to be identified. He is making an anonymous comment on a website. Or he's

launching a cyberattack against your network. In such a case, the systems have to answer the harder question: "Who is this?"

At a very basic level, we are unable to identify individual pieces of hardware and software when a malicious adversary is trying to evade detection. We can't attach identifying information to data packets zipping around the Internet. We can't verify the identity of a person sitting in front of a random keyboard somewhere on the planet. Solving this problem isn't a matter of overcoming some engineering challenges; this inability is inherent in how the Internet works.

This means that we can't conclusively figure out who left an anonymous comment on a blog. (It could have been posted using a public computer, or a shared IP address.) We can't conclusively identify the sender of an e-mail. (Those headers can be spoofed; spammers do it all the time.) We can't conclusively determine who was behind a series of failed log-ins to your bank account, or a cyberattack against our nation's infrastructure.

We can't even be sure whether a particular attack was criminal or military in origin, or which government was behind it. The 2007 cyberattack against Estonia, often talked about as the first cyberwar, was either conducted by a group associated with the Russian government or by a disaffected 22-year-old.

When we do manage to attribute an attack—be it to a mischievous high schooler, a bank robber, or a team of state-sanctioned cyberwarriors—we usually do so after extensive forensic analysis or because the attacker gave himself away in some other manner. It took analysts months to identify China as the definitive source of the *New York Times* attacks in 2012, and we didn't know for sure who was behind Stuxnet until the US admitted it. This is a very difficult problem, and one we're not likely to solve anytime soon.

Over the years, there have been many proposals to eliminate anonymity on the Internet. The idea is that if everything anyone did was attributable—if all actions could be traced to their source—then it would be easy to identify criminals, spammers, stalkers, and Internet trolls. Basically, everyone would get the Internet equivalent of a driver's license.

This is an impossible goal. First of all, we don't have the real-world infrastructure to provide Internet user credentials based on other identification systems—passports, national identity cards, driver's licenses, whatever—which is what would be needed. We certainly don't have the infrastructure to do that globally.

Even if we did, it would be impossible to make it secure. Every one of our existing identity systems is already subverted by teenagers trying to buy alcohol—and that's a face-to-face transaction. A new one isn't going to be any better. And even if it were, it still wouldn't work. It is always possible to set up an anonymity service on top of an identity system. This fact already annoys countries like China that want to identify everyone using the Internet on their territory.

This might seem to contradict what I wrote in Chapter 3—that it is easy to identify people on the Internet who are trying to stay anonymous. This can be done if you have captured enough data streams to correlate and are willing to put in the investigative time. The only way to effectively reduce anonymity on the Internet is through massive surveillance. The examples from Chapter 3 all relied on piecing together different clues, and all took time. It's much harder to trace a single Internet connection back to its source: a single e-mail, a single web connection, a single attack.

The open question is whether the process of identification through correlation and analysis can be automated. Can we build computer systems smart enough to analyze surveillance information to identify individual people, as in the examples we saw in Chapter 3, on a large-scale basis? Not yet, but maybe soon.

It's being worked on. Countries like China and Russia want automatic systems to ferret out dissident voices on the Internet. The entertainment industry wants similar systems to identify movie and music pirates. And the US government wants the same systems to identify people and organizations it feels are threats, ranging from lone individuals to foreign governments.

In 2012, US Secretary of Defense Leon Panetta said publicly that the US has "made significant advances in . . . identifying the origins" of cyberattacks. My guess is that we have not developed some new science or engineering that fundamentally alters the balance between Internet identifiability and anonymity. Instead, it's more likely that we have pene-

trated our adversaries' networks so deeply that we can spy on and understand their planning processes.

Of course, anonymity cuts both ways, since it can also protect hate speech and criminal activity. But while identification can be important, anonymity is valuable for all the reasons I've discussed in this chapter. It protects privacy, it empowers individuals, and it's fundamental to liberty.