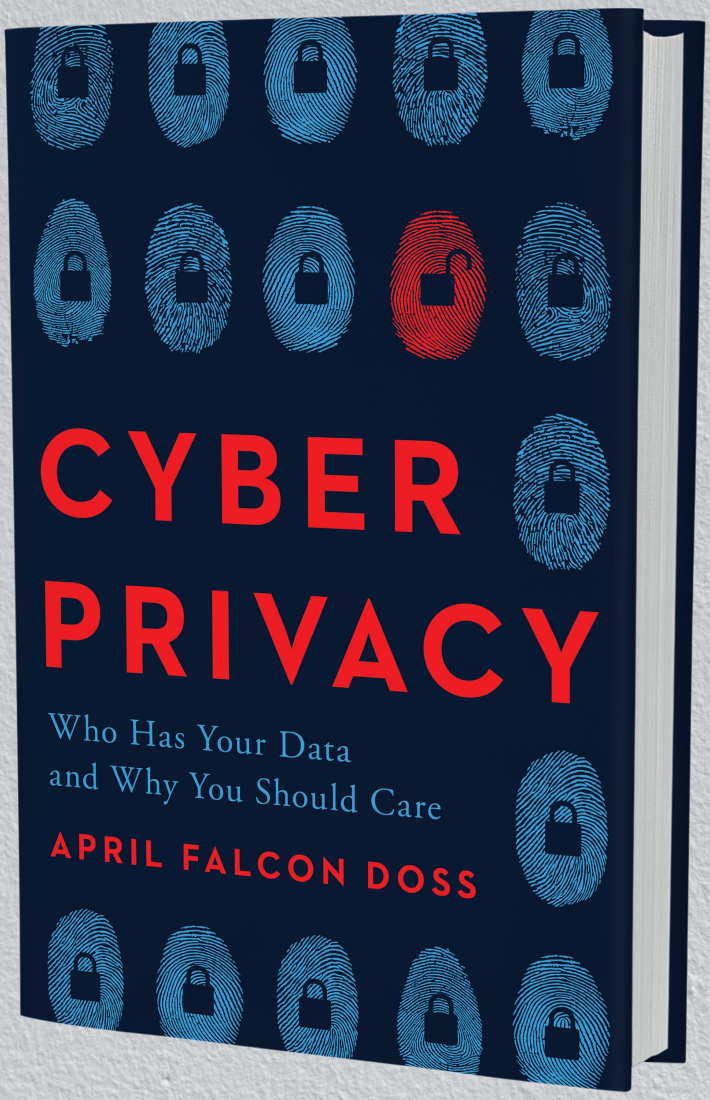
SPECIAL EXCERPT



This excerpt is exclusively for attendees of the Privacy + Security Forum Fall Academy

SECTION II

If You're Not Paying for the Product, You Are the Product

By the summer of 2019, an estimated 2.4 billion people around the world had Facebook profiles, and nearly 1.6 billion people were logging into the social media platform every day. Years before, academic researchers and advertisers had discovered that analyzing a person's Facebook "likes" revealed a surprisingly detailed and accurate personal portrait.[1] In 2015, an academic journal published research showing that computer models equipped with a person's history of Facebook likes outperformed friends and family members in gauging the individual's personality traits and inclinations.[2] They studied 17,000 computer assessments and 14,000 human assessments, and concluded that with somewhere between 10 and 300 likes, the algorithms become more accurate than the humans in assessing someone else's personality. With 10 likes, Facebook knows you better than your coworkers do. With 70 likes, Facebook knows you better than your roommates and real-life friends do. With 150 likes, Facebook knows you better than your own family does. And with 300 likes, Facebook knows you better than your spouse.[3]

If companies can make individually tailored, and highly educated, guesses about us based on innocuous social media scrolling behavior, how should that shape our thinking on what limits should be in place for said companies to collect and use this information? Indeed, there is no clear, uniform, or clearly articulated legal framework for consumer data privacy in the United States. Companies that operate in the United States, from massive social media platforms to the tiniest home-basement-developed apps, have almost completely unfettered ability to collect, retain, compile, cross-reference, enhance, sell, share, buy, and use information about the consumer. In most cases, for most kinds of information, the only requirement is that they provide the user with a privacy notice explaining their data practices. The consumer may or may not have read or understood the notice, but they typically are deemed to have given consent if they keep using the service or if they've clicked on a box somewhere that says something like, "I accept." This notice-and-consent-based approach dates back nearly a half-century, and has taken root not only in the United States but in other countries around the world as well.

The 1970s and 1980s proved to be formative years in privacy-related technology, policy, and law. With the expansion of computerized record-keeping, policymakers in the United States were searching for the right balance of principles and practices that would facilitate expanding the potential societal benefit of online databases, while mitigating harm to individuals. Policymakers and thought leaders, such as the authors of the US government report that established the Fair Information Practices, believed that a framework of privacy notices coupled with consumer consent would allow the public as a whole to engage in reasoned review of government information practices.[4] National and international bodies looked to privacy notices and consumer consent as important tools to support individuals' ability to make decisions about the use of their data. By extension, this notice-and-consent model could provide some check on unsavory corporate behavior.[5] Now, decades on, it isn't hard to see that the effectiveness of the notice-and-consent model has been eroded by a combination of factors: the proliferation of wordy, unintelligible privacy policies; consumers' recognition that they're powerless to negotiate any better or different privacy terms; and the fact that new types of data tracking, collection, and analysis are being developed faster than consumers can become aware of them.[6]

In today's complex data privacy environment, it's more difficult for consumers to understand what they're consenting to: how their data might be collected and used by the owner of a free product or service they've signed up for, how it might be sold to others, what the impacts of cross-platform data aggregation are, and how artificial intelligence algorithms are creating behavioral prediction models about them. The reality is, those prediction models can be used for purposes as ordinary as direct marketing and targeted commercial advertising, as well as for purposes as consequential as political advertising or as sinister as political viewpoint manipulation by hostile foreign governments looking to sway public opinion in Western democracies.

When it comes to private-sector use of data, there is no equivalent to the Fourth Amendment protections that restrict government data collection. There is no federal data privacy law. And courts haven't decided how they feel about common law claims for invasion of privacy when a company takes, uses, shares, or loses an individual's data.

When it comes to private-sector use of data, there is no equivalent to the Fourth Amendment protections that restrict government data collection. There is no federal data privacy law. And courts haven't decided how they feel about common law claims for invasion of privacy when a company takes, uses, shares, or loses an individual's data.

Faced with these challenges, state legislatures have been considering ways to fill the privacy protection gap. Illinois' Biometric Information Protection Act, passed more than a decade ago,[7] became a wellspring of litigation in 2019 when the Illinois Supreme Court held that individuals could file lawsuits against companies that collected handprints and other biometric information without their consent.[8] The lawsuits, ranging from the precedent-setting complaint against an amusement park's use of handprints for its ride lines to the now-routine lawsuits against companies who direct their employees to use fingerprints or handprints for logging in and out of biometrics-based company time clocks, have prompted other states to adopt similar laws restricting

the collection and use of biometric data and have also prompted companies to lobby against these new proposals and urge changes in the Illinois law to take away the right of individuals to sue.[9]

In what has been the most significant change to US privacy law so far, California passed a sweeping new law, the California Consumer Privacy Act (CCPA), that took effect on January 1, 2020. The CCPA started as a grassroots voter referendum intended to give individuals more insights into who is collecting "personal information" about them and why, and what they're doing with it. One of the hallmarks of this new law is that it vastly expands the kinds of data governed by the new privacy protections. Most state data privacy laws focus on breaches of narrow categories of information that typically center around Social Security numbers, payment card and financial account information, and sometimes medical records. Under CCPA, however, "personal information" is defined to include almost every imaginable fact or inference about California-based individuals or households: the definition includes traditional items like Social Security numbers and credit card and banking information, as well as IP addresses, online shopping and other internet activity, location data, biometrics, education and employment information, "audio, electronic, visual, thermal, olfactory, or similar information," and "inferences" that are drawn from personal data to create consumer profiles "reflecting the consumer's preferences, characteristics, psychological trends, preferences, predispositions, behavior, attitudes, intelligence, abilities, and aptitudes." Under CCPA, consumers can ask for access to, correction of, or deletion of their data, and can instruct companies that they don't want their data to be sold. Data collection about children under thirteen requires parent or guardian consent, and the law presumes that no one under the age of sixteen wants their data to be sold. The law only applies to companies doing business in California that meet certain revenue or data-handling thresholds, and the consumer protections only apply to residents of California.[10] By July 2020, enforcement had only just begun, and a long list of amendments were part of a ballot referendum in the November 2020 election. With so much uncertainty, it isn't clear yet what the law's long-term impact will be.[11] Nonetheless, given California's status as the world's fifth-largest economy, the law has a broad sweep, and privacy advocates in California are already working on

changes to further strengthen the law. Meanwhile, legislators in other states as well as in Congress are watching CCPA developments to see what provisions might be adopted elsewhere and which ones are more trouble than they're worth.

Whether it's Facebook, Google, or the countless free apps and web-based services we use, the companies providing these services are getting something from us in return. Since we aren't paying them, they aren't getting money directly from us. Instead, they're monetizing our time on-screen. In some cases, they profit by offering us products from partners who pay them a commission when we make a purchase. In other cases, they profit by selling information about us to other companies who are interested in buying it. In still other cases, they create vast ecosystems of personal data about us and invite other companies to advertise to us in exchange for a fee. No matter what the specific business model is, the outcome is the same. When we're not paying for the product, we *are* the product. That doesn't mean that we shouldn't use these "free" services. But it helps, when we do, to know exactly what we're doing and to think about how much our privacy is worth.

CHAPTER 5
THE BIG 4

Apple, Google, Facebook, Amazon

On April 21, 2016, a friend's post made it onto my Facebook feed:

> I may delete FB from my phone. Yesterday I used a gas station I'd never used before. An hour later an ad for that gas station appeared in my feed. My phone has become a totalitarian state and is spying on me.

Facebook—along with Apple, Google, and Amazon—knows you better than your own mother does. At least that's the conclusion of academic researchers and advertisers who pay to leverage the platforms' insights about you. It's even what these platforms' own marketing departments will say if you catch them in an unguarded moment. These four companies have unprecedented and unparalleled access to data about our preferences in news and entertainment, our online searches and purchases, our religious and political affiliations, our health, our education and employment, our hobbies and interests, our social connections, and our "psychosocial profiles." Their power and influence has grown so great

that it's hard to recall their near-total dominance of entire data ecosystems is a recent phenomenon.

To understand the scope and reach of Apple, Google, Facebook, and Amazon, we should recall how much data has expanded in recent years and consider the ways that data matters to these companies. According to one report from December 2018:

> Ninety percent of the world's data was created in the last two years, and over 2.5 quintillion bytes of data are produced every day . . . And this data is then used to market products to us. In 2018, almost half of all advertising spend will be online, rising to over 50 percent by 2020. And two digital giants—Facebook and Google—now control 84 percent of the market. The companies are hugely reliant on ad revenue, with Facebook collecting 97 percent of their overall revenue from ad spending while at Google it accounts for 88 percent.[1]

For every minute in 2017, YouTube users watched over 4 million videos, Google responded to 3.6 million search requests, and Amazon earned over $250,000 in sales.[2] That's *every minute*—for an *entire year.* And every year, these numbers go up. Now think about the volume of personal information stored by just one of those companies, let alone all four. Every minute, Google collects and stores data about those millions of YouTube (they share a parent company, Alphabet) videos watched: which individuals are tuned in to which videos, who clicks on those videos served up in search results versus auto-load or recommender analytics pushing videos to the top of an individual's feed, and so on. Because of Google's market dominance in search and heavy market share in services like webmail and products like mobile phones, it can correlate the information about users' video-viewing habits with information about the identity of the contacts in their digital address book, what they're talking about in the emails they exchange, and where they're traveling using Waze or Google Maps—including the difference between their occasional trips and their daily routines. All of this data and more can be combined, correlated, and crunched with the data that Google receives and

stores from the 3.6 million web searches it executes every minute. This is a staggering granularity of detail about individuals that has never been available to anyone—governments or corporations—in the past.

Perhaps it's because the rise of these companies is so recent that their privacy policies and practices have varied widely and changed frequently over the years. Government regulators in the United States and abroad have vacillated between wanting to encourage corporate growth and innovation and wondering when to rein them in. As governments around the world are beginning to grapple with the consequences of these vast data pools, most of the regulatory effort has been concentrated in regions like Europe and countries like Australia, Canada, and the United States. The greatest scrutiny has landed on the corporations—Facebook, Google, Amazon, and Apple—that have grown so large they function as extra-national fiefdoms, courting governments and challenging government mandates in every country where they operate. It isn't just privacy that's at stake; competition is suffering, too. Smaller competitors can't enter the market because of the dominant position of the digital behemoths. Startup companies wanting to introduce a new search engine point to Google's market dominance as a barrier to entry; new social media platforms point to Facebook and raise the same concerns. Within the United States, the most important curbs on corporate exercise of data-related power have come from the Federal Trade Commission and from private litigation, with a handful of states starting to step into the mix. As the nation's antitrust regulator and consumer protection watchdog, the FTC may be uniquely positioned to leverage existing legal tools to bear in considering whether these giants may have grown too powerful.

The first antitrust law in the United States, the Sherman Act, was passed in 1890 to break up the railroad, steel, oil, and sugar monopolies that were dominating the economy and politics of the late nineteenth century. By the time that Teddy Roosevelt took office, the federal government was aggressively pursuing antitrust litigation against a bevy of companies that were operating in "restraint of trade."[3] Early regulators focused primarily on the anti-competitive impact of vertically integrated companies, a trend that would continue for a century as regulators asked, "Are all the steps in a supply

chain controlled by a single corporation? Are large companies preventing smaller ones from entering the market? Are consumers paying higher prices as a result?" Although FTC regulators investigating Standard Oil might not have anticipated a Google, Amazon, or Facebook, the principles captured in that last question will likely prove to be key to the ways that today's FTC thinks about the role that antitrust laws can or should play in data privacy: the idea that competition benefits consumers is what prompted the United States, and many other countries, to create a single regulatory agency charged with both anti-competition regulation and consumer protection.

Over a century of experience, however, indicates that merely measuring the price of products doesn't provide a complete picture of whether or how corporate mergers and growth are impacting individuals for the worse. As noted in the previous chapter, pricing alone doesn't capture the full spectrum of benefits that users enjoy in "free" apps and services. It also doesn't account for the nonmonetary costs, such as the ways in which technology intrudes on our sense of security, self, and autonomy; sets us up for viewpoint manipulation; and interferes with our right to be left alone. Perhaps it's no surprise, then, that progressive politicians are starting to join privacy advocates in arguing that big tech needs to be broken up.[4]

WHAT KINDS OF DATA ARE WE TALKING ABOUT?

Although many of us have a general sense that digital platforms collect a lot of data from us, it's often harder than one might expect to get a clear picture of precisely what data is being scooped up. It's even harder to understand how that information is used in drawing inferences about us and in attempts to influence us. The Australian Consumer and Competition Commission (ACCC) tried to tackle these questions in its 2019 report, explaining that platform providers go to great lengths to capture their users' attention: a longer attention span translates into more user data. To meet that goal, data is collected in three ways: actively (e.g., when a user enters their contact information in an online form, watches a video, clicks on a link, or navigates to a

new page); passively (e.g., background collection of location data from Wi-Fi networks); and by inference (e.g., by analyzing active and passive user data to draw inferences about a user's age, gender, health, sexual orientation or identity, political affiliations, hobbies, interests, and so forth).[5]

The ACCC's research underscored the gap between the kinds of data that individuals consider to be "personal information" and the data that's covered by most privacy policies and data protection and data breach laws. According to the ACCC report, when Australian consumers were asked what kinds of data they viewed as "personal information," they included the following items in the list: date of birth (86 percent), a person's name (84 percent), photographs (79 percent), telephone and device information (79 percent), and location information (78 percent).[6] Under most US state data breach laws, name and date of birth are only considered personal information if they're combined with other, more sensitive information, like Social Security numbers or payment card information. Privacy laws generally don't protect location information, except with respect to certain kinds of government uses. Telephone and device information is largely unregulated by these laws. And under most state data breach laws, photos aren't protected at all.

We didn't need the ACCC report to see that gap; the platform providers' privacy policies are proof enough. By and large, privacy notices are written with an eye toward complying with whatever laws govern the data collection practices of the service, app, platform, or device. The risks and requirements under those laws are proliferating—back in 2012, researchers estimated that if a person were to read all of the privacy notices that accompanied every service they use, it would take seventy-six straight days to complete the reading.[7] That number is almost certainly higher now. Most of us largely accept the fiction that these privacy policies might impact the decisions we make. Even courts acknowledge that these privacy policies offer little more than a fig leaf of user notice and consent, since they are cumbersome to read, difficult to understand, and individuals have few alternatives when it comes to using the major digital platforms.

One stark example of this emerging view among courts came in a decision involving the class-action litigation filed against Facebook in 2018, alleging that

the platform violated users' privacy rights when it shared personal information with the behavioral research and political marketing company Cambridge Analytica. The District Court for the Northern District of California had to assess, among other things, whether users consented to having their Facebook profile information, posts, photographs, and contacts' information shared with Cambridge Analytica so that it could target political messaging campaigns. Facebook's defense rested in part on the position that at least some of the provisions in the various versions of its privacy policy that were in effect at different times should have put users on notice that their profiles, and their friends' profiles, would be shared with third parties. As a result, Facebook argued, there was no harm and therefore no legally cognizable foul when Facebook allowed the consulting company to export information of some 87 million users. In addressing this issue, the court noted that, "The parties agree that California law requires the Court to pretend that users actually read Facebook's contractual language before clicking their acceptance, even though we all know virtually none of them did."[8] This creates a difficult conundrum for plaintiffs and judges. As Judge Vince Chhabria wrote:

> To be sure, for the rare person who actually read the contractual language, it would have been difficult to isolate and understand the pertinent language among all of Facebook's complicated disclosures. Thus, in reality, virtually no one "consented" in a layperson's sense to Facebook's dissemination of this information to app developers. But under California law, users must be deemed to have agreed to the language quoted [in the privacy policy].[9]

This particular opinion was written at an early stage in the litigation, and at the time this book was going to press, there had yet to be a final resolution. It also isn't clear yet whether other courts will give a sympathetic hearing to future claims from users who say, in effect, "Sure, there was a privacy policy, but I didn't read it because I knew I didn't really have a choice," or "I read it, but I didn't understand it," or "It changed so often that I couldn't keep up." On some of these points, Alvira Roberts, if she were alive today, might

sympathize. Future courts may find a parallel among plaintiffs who were presented with a privacy policy whose language seemed to disclose what was being done with their data (the cyber privacy equivalent of, "My friend is here to assist you") while failing to draw attention to other crucial facts ("He's not a doctor or medical staff of any kind"). Until that happens, Judge Chhabria's opinion is noteworthy for its blunt assessment of the practical utility that privacy policies have for most users.

Platform privacy notices are frequently updated and changed, but a snapshot shows some typical definitions of personal information. Google's privacy policy, for example, defines personal information as "information that you provide to us which personally identifies you, such as your name, email address, or billing information, or other data that can reasonably be linked to such information by Google, such as information we associate with your Google Account." Facebook's policy doesn't include an explicit definition of personal information, but does refer to "information that personally identifies you" as "information such as your name or email address that by itself can be used to contact you or identifies who you are." Apple's privacy policy states that "personal information is data that can be used to identify or contact a single person."[10]

Many of the kinds of data that seem most "personal" to the average user—photos, videos, interests, and membership in closed groups—fall outside these definitions. Even worse, none of the definitions even hint at passive collection or inferences. These definitions don't tell their users that providers are tracking their location from their internet connection and picking up all manner of digital detritus that spills out of other, leaky apps on the user's device. Or that the provider's cookies are tracking the user's web browsing, online shopping, and more, even after the user logs out of the provider's platform or app. Or that the providers are running complex analytics across all of this actively and passively collected data in order to analyze the user's personality, derive inferences about their interests, or influence their future behavior. None of these definitions explain that many platforms share personal data with corporate partners like data brokers who may sell the information to still other corporate partners (see sidebar on page 76), or that the platforms

USER ENGAGEMENT TACTICS: COMPANIES KEEPING YOUR EYES ON THE SCREEN

This is why YouTube and Netflix automatically cue up the next video as soon as the one you're watching is almost done, or why they make you wait until the next video has fully commenced before you can pause or stop the automated queue. It's because studies have shown that when a video simply ends, people are less likely to start a new one. Auto-queuing the next episode in a series lures people into binge-watching.

Just like willingness-to-accept and willingness-to-pay are often driven by the default privacy setting, how long people stay online—and provide more data for apps and platforms—is often driven by whether the default is an infinite scroll or an infinite video queue. Proactive steps always take more effort. So video games require half a dozen separate clicks to exit the game, and platforms make the default setting one that allows passive users to keep watching or scrolling into eternity.[11] According to a growing body of insider accounts, research, and reports, "Features such as app notifications, autoplay—even 'likes' and messages that self-destruct—are scientifically proved to compel us to watch/check in/respond *right now* or feel that we're missing something new or important."[12]

invite corporate partners into their ecosystems to support paid, microtargeted advertising. And none of the blandly worded policy notices advise users that, when platforms say they may share user data with their "partners," the scope of that sharing could include hundreds or thousands of entities and individuals, including app developers who pull in detailed personal information through application programming interfaces (APIs).

When Facebook, Google, and other companies allow third-party developers to provide new products and services on their platforms, those new apps encourage users to spend more time on the platform. This increases the advertising value of the platform and gives both the platform and those app developers more data about the user, which can be used to further target those users

and shape their behavior. This vicious cycle is so opaque that many users don't know it exists. But the cycle is that way by design—it begins with teams of psychologists and social scientists who educate Silicon Valley on how to change user interfaces and services in order to entice people to stay online longer.

Lest individuals should think they're immune to this kind of data gathering so long as they don't use the major services that carry out these practices, the reality of the situation is heavily lopsided, and not in individuals' favor. Google, for example, collects information not only from individuals' own use of Google services, such as Gmail, Google's Chrome browser, and Google search, but also via the visiting of websites that use Google's analytics or advertising services. Similarly, Facebook tracks data of users who are logged into their Facebook accounts, users who are logged *out* of their Facebook accounts, and individuals who don't even have a Facebook account but who visit pages that have Facebook "like" buttons or other Facebook plug-ins that let the website owner boost traffic to their page or take advantage in other ways of the analytics and advertising opportunities that are made possible by Facebook's global reach.[13]

Although consumers and regulators alike are becoming more aware of the kinds of information being collected through digital means, rapid changes mean there are always new surprises. In February 2019, reports surfaced that Facebook was harvesting sensitive health information from people who were using completely independent apps.[14] The mechanism was a simple one: Facebook created a tool called App Events. When non-Facebook apps built App Events into their design, App Events allowed Facebook to gather data from those apps. Some data was relatively benign, like how many times a day the app was opened, or how long it was used during a particular session. But App Events could also gather data that the user manually input. In the case of health and fitness apps, this included things like blood pressure, weight, medications, exercise, heart rate, blood sugar, and the like. What grabbed the headlines, however, was the fact that, based on this feature, apps like Flo Health were sending Facebook data about the timing and frequency of women's menstrual cycles and sexual activity, and the apps' users had no idea—and no intention—of giving that very personal information to Facebook.[15]

WHAT KINDS OF CONSENT ARE WE GIVING?

By and large, the consents that we're giving are blanket, wide-reaching, and uninformed. That's a stark, bleak description of the situation, but it also provides a fair thumbnail sketch of the ways in which privacy notices and consent operate for most of the major digital platform providers. The advantage lies entirely with the platform. As one pair of researchers noted, "When a company can design an environment from scratch, track consumer behavior in that environment, and change the conditions throughout that environment based on what the firm observes, the possibilities to manipulate are legion." With that kind of reach and influence, the platform is able to "reach consumers at their most vulnerable, nudge them into overconsumption, and charge each consumer the maximum amount that he or she may be willing to pay."[16]

WHAT'S BEING DONE WITH OUR INFORMATION?

Once it's been harvested, our data takes on a life of its own.

Sometimes, our data sits right where it started. In 2019, Apple famously rolled out an advertising campaign with billboards proclaiming that "What happens on your iPhone, stays on your iPhone." It was a catchy slogan, and well designed to capitalize on Apple's carefully cultivated reputation for privacy. The campaign was timed perfectly to coincide with a major electronics industry conference in Las Vegas, where the ad's theme echoed the old and slightly scandal-suggesting trope that "What happens in Vegas, stays in Vegas."

Within the community of information security and privacy researchers, the billboard campaign was met almost immediately with a sentiment of, "Challenge accepted." Sure enough, it didn't take long for researchers to discover that the marketing hype was, to a large degree, merely hype.

In May 2019, the *Washington Post* reported that it had carried out tests to see just how much data really did stay on the iPhone. *Post* technology columnist Geoffrey A. Fowler teamed up with a security research lab to assess his iPhone's activity and found that, within seven days, the phone had exported data via 5,400 hidden app trackers.[17] Fowler's location information, IP address, phone number, and other device-identifying information were being exported off the device to thirsty apps that slurped the data in. The security research lab estimated that the trackers would have exported 1.5 GB of data over the span of a month—an amount that could easily chew up half of the monthly allotment for someone subscribed to a basic-level phone plan.[18]

Although Apple might be the emperor of privacy-based marketing campaigns, it wouldn't be fair to say the emperor has *no* clothes. After all, Apple devices leak less data than equivalent Android-based phones that leverage Google's vast data empire. But Apple might only be wearing a tank top and shorts, and not the full three-piece suit they've led us to believe.

Facebook and Google, by contrast, don't make sweeping promises about protecting user data. But they also don't generally sell it outright, and they don't sell it to third parties as often as people might assume. Instead, Facebook and Google operate "walled gardens of data."[19] That is, they create an ecosystem within which personal data is generated by users who intentionally volunteer information such as name, email address, birthday, searches, interests, photos, videos, contacts, and the like. And users add to that growing garden of data with information they provide unintentionally. Indeed, every time a user logs into Facebook or carries out a search on Chrome or sends or receives mail via their Gmail accounts or uses their Android phone, their information is collected by the gardener, or ecosystem host, and added to the already-rich profile of information on hand. Data about activity inside the garden is supplemented with data about activity outside the garden—activity that the gardener can track by means of the footprints that we leave as we traipse around the internet and stumble across the piles of mud that the gardeners have left to trip us up outside of their walled ecosystems.

DATA BROKERS: THE INVISIBLE BUSINESSES THAT SEE US ALL

Sometimes, our data takes flight, traveling in nanoseconds around the globe and through the servers and algorithms of companies whose names we've never heard. Our data gets shared with data brokers and aggregators who further compile, collate, collect, massage, integrate, interpret, analyze, sell, and re-sell our information to a range of bidders in a marketplace that we have little insight into and virtually no control over.

Broadly speaking, "data brokers" are companies, or business units within companies, that earn their primary revenue by supplying data or inferences about people that are gathered mainly from sources other than the data subjects themselves.[20] Because brokers generally get their data from third parties, consumers are largely unaware of their activities or of the profiles that data brokers are maintaining on them. In order to monetize these profiles, data brokers frequently create lists of people with shared attributes. In some cases, the lists seem tied to relatively benign, nonsensitive information: dog owners, winter activity enthusiasts, or "mail order responders."[21] In other cases, the lists relate to medical conditions, like wheelchair users, people with cancer, insulin-dependent diabetics, or people with depression. They may be sorted to identify people with breast cancer, impotence, vaginal infections, or HIV. Other lists are tied to religion, ethnicity, immigration status, or national origin, while others still are linked to economic circumstances, such as "Pay Day Loan Central—Hispanic," "One Hour Cash," or "Help Needed—I Am 90 Days Behind with Bills." Lists may also reflect family status, such as "expectant parent"; a combination of socioeconomic and family status factors, such as "upper-middle class with no children"; or shorthand categories created by the data brokers themselves, such as "rural everlasting" to refer to single men and women over the age of sixty-six with "low educational attainment and low net worth." In some cases, data characteristics are aggregated by neighborhoods, buildings, or households; in other instances, the data is specifically tied to identifiable individuals,

with each person in a household uniquely identified for purposes of the data profiles and their membership on various lists.[22]

How do data brokers make money off of all this information? Generally speaking, they sell (or rent) personal data to other companies who want to use it for three main purposes: marketing, risk mitigation, and people search.[23] In the marketing context, the lists and individual profiles are used to analyze, segment, and sort prospective customers for targeted advertising campaigns based on particular characteristics, behavior, profitability, and projected lifetime value as a consumer for the company.[24] Companies interested in risk mitigation products are usually looking to confirm individuals' identities or detect fraud. "People search" services are often available online to any user, with limited information returned at no direct cost to the searcher, and more comprehensive profiles available for a fee.[25]

In 2014, the US Federal Trade Commission renewed its calls for Congress to pass legislation governing the activities of data brokers. The FTC acknowledged that the data broker business can provide benefits to consumers by increasing the likelihood that consumers will be presented with ads for products and services they're interested in. However, the FTC also pointed out the risks to privacy and data security of having all of this consumer information held in just a few hands, and recommended that new laws be passed that would, among other things, give consumers the right to find out what information data brokers have about them and to opt out of having their data sold for marketing purposes. Thus far, Congress hasn't made any progress in this area. But states are beginning to take action, with Vermont passing a data broker registration law in 2018 that requires companies whose primary business is the aggregation and sale of personal data to register with the Vermont attorney general.[26] And, as noted, the California Consumer Privacy Act of 2018 restricts the sale of data about children, gives consumers the right to prevent companies from selling their data, and requires that data brokers release detailed statistics about the types and volume of the data they collect.[27]

For example, Facebook tracks users outside their platform by embedding Facebook tracking pixels on participating websites. These pixels are invisible to users, and allow Facebook to track their activity on those sites *regardless of whether the site visitor has a Facebook account*. Facebook employs other external tracking tools as well, such as the Facebook "like" button that appears on many websites, the "login with Facebook" function available for many other platforms, and Facebook analytics, which many websites use for measuring traffic to and through their sites. As of April 2018, the Facebook "like" button appeared on 8.4 million websites, the Facebook "Share" button appeared on 93,000 websites covering 275 million web pages, and there were 2.2 million Facebook pixels installed on websites around the world.[28]

Although major platforms such as Facebook and Google don't generally sell user data to advertisers, they do make that information directly available to third-party app developers. Between February and April 2018, there were approximately 1.8 million apps on Facebook and 1.5 million app developers active on Facebook. Although the apps and developers are supposed to be operating within the confines of Facebook's privacy policy, that policy allows apps to create their own privacy notices that users seldom read and never consider objecting to. Facebook insists that it has remedied the practices that allowed third-party app developer Cambridge Analytica to siphon off the detailed personal information of some 87 million users who had never given consent to sharing their information. In July 2019, the US Federal Trade Commission levied a fine of $5 billion on Facebook because of its data handling practices, included the broken promises that were demonstrated by the Cambridge Analytica scandal.[29] However, many commentators pointed out that $5 billion, although precedent-setting in its size as it dwarfed the previous-largest privacy-related fine levied by a US regulator, was still only a fraction of Facebook's quarterly profits and cash reserves, amounting to little more than a slap on the wrist.[30]

In addition to that wide world of data sharing, the fact that companies of all sizes and across all market sectors and industries have access to so much information about us makes it possible for corporate and government entities to make a whole range of data-driven assumptions, conclusions, and decisions about individuals, including decisions that are inaccurate, arbitrary, or biased.[31]

DO PRIVACY AND COMPETITION INTERESTS ALIGN?

By 2017, antitrust regulators in the European Union were considering action against tech giants like The Big 4. The EU Competition Commissioner expressed concern about the ways in which the platform providers' data advantage served as a barrier to entry for other businesses. Since then, some countries in Europe have stepped out in front. For example, in 2019, Germany's Federal Cartel Office ruled that when Facebook harvests and processes data from third-party sites, it violates European data protection law, because users had no meaningful opportunity to object to the third-party data collection; their only choice was between widespread data collection or not using Facebook at all.[32]

Meanwhile, in the United States, Senator and 2020 presidential candidate Elizabeth Warren (D-MA) proposed a plan to "break up big tech," emphasizing that undoing some of the massive mergers of recent years—Amazon's purchase of Whole Foods and Zappos, Facebook's purchase of WhatsApp and Instagram, Google's purchase of Waze, Nest, and DoubleClick—could have the benefit of making tech companies more responsive to users' concerns, including those about data privacy.

The US Federal Trade Commission has been soliciting input on whether the biggest tech platforms are engaging in unlawful anti-competitive practices, and whether antitrust enforcement action could have ancillary benefits for privacy. In addition to the potential for federal-level action, investigation, regulation, or enforcement from the FTC, all fifty US states have consumer protection laws, many of which are closely modeled on the national FTC Act.[33] As the FTC has been exploring antitrust implications of big tech, it has invited input from states, where new litigation and legislation sometimes move more quickly than at the federal level.

As part of that federal-state interaction, in October 2018 a dozen attorneys general, representing eleven states plus the District of Columbia, wrote to the FTC about their concerns over data privacy and competition.[34] In their letter, the state AGs point to the central trade-off that these companies rely on: users' willingness to "make certain of their personal data available for monetization in return for the often 'free' services they receive."[35] The AGs

were concerned that so much data is concentrated in the hands of just a few companies: nearly all searches use just one search engine, "over 90 percent of young people have a profile on one social media platform," and 99 percent of smartphones use either Apple's iOS or Google's Android operating system.[36]

According to the AGs, large-scale data aggregation by a small number of platforms can lead to a number of anti-competitive harms.[37] First, consumers suffer: the "immense" power imbalance between market-dominating platforms and consumers results in lengthy and opaque user agreements and few realistic alternatives for consumers, setting up a cycle in which consumers believe they have no choice but to agree to the platforms' collection of their data. Second, the big platforms' data advantage chokes out competition: without deep and detailed data about individual consumers, rivals are unable to serve up equally targeted advertisements, and therefore unable to attract the advertising dollars they need to stay afloat. To illustrate the problem, the letter noted that having access to historical search data improves the quality of new search results by up to 31 percent. "In effect," the AGs wrote, "today's search engines cannot reach high-quality results without this historical user behavior."[38] (Ironically, one counterbalance to this problem might be to allow more companies to have access to consumers' historical information—which could result in better search results but could also undermine privacy by making the data available to a wider base of companies.)

The AGs also urged caution with respect to big data algorithms, noting that in some cases algorithms could lead to price-discrimination or price-targeting that disadvantaged certain groups—a risk made more acute by the fact that there's so little transparency around how algorithms reach their conclusions.[39] (Issues surrounding algorithmic analysis and decision-making are addressed in greater depth in chapter 7.)

Perhaps the most important point was the AGs' position that "focusing on price to consumers is too narrow an interpretation of the principles of antitrust law."[40] With this statement, the AGs opened the door to considering privacy, discrimination, and other impacts from data-driven technologies—effectively undermining the ability of platforms, apps, and services to defend themselves solely on the basis that their products are "free."

The scope of the 2019 FTC fine against Facebook illustrates the ways in which a handful of major tech platforms have become so large that they're virtually ungovernable. When privacy advocates protested that $5 billion was far too little for Facebook to pay for privacy violations spanning nearly a decade, part of their rationale was that, in the same quarter that the fine was announced, Facebook earned $15 billion in revenue, and the company was sitting on $40 billion in cash reserves. Facebook's global userbase and deep pockets reinforce the power of its walled garden of data, further cementing its monopoly position and continuing to create incentives for millions of unaffiliated websites to embed Facebook tracking mechanisms.

Where the tech sector had enjoyed decades as the darlings of American economic innovation, federal and state regulators are now starting to question whether it's wise to allow so much of that growth to take place without some degree of oversight or regulation. In October 2019, a bipartisan group of forty-seven attorneys general launched an antitrust investigation into Facebook. Their individual press releases offered statements as varied as their constituencies: couched in different terms, with different areas of emphasis. Nonetheless, they agreed on a joint statement indicating that they "all are concerned that Facebook may have put consumer data at risk, reduced the quality of consumers' choices, and increased the price of advertising." All committed to "use every investigative tool at our disposal" to investigate the social media behemoth.[41] A bipartisan, multi-state investigation of this scope and scale would have been virtually unimaginable five years ago. That it's gained such widespread national traction is evidence of the growing realization that consumers may suffer as much or more from the ways that anti-competitive practices undermine their privacy as they do from practices that increase their out-of-pocket costs for participating in modern society and in the digital economy.

CHAPTER 8
DIFFERENTIATING THE REAL FROM THE FALSE

Social media platforms have become the front lines of viewpoint manipulation and propaganda. Media-conscious influencers can hire "black PR" firms to develop content, create fake follower accounts, and manipulate social media engagement in order to create carefully curated public images for politicians, entertainers, or anyone else willing to pay for their services.[1] Savvy individuals, campaigns, and governments are using the widespread reach of social media and the built-in thought bubbles created by our decisions about whom to friend and whom to follow as a means of leveraging our personal data to change the ways we think about the world.

To be clear: this kind of influence operation is different from traditional news outlets that post perspectives, information, and sometimes even lies. Traditional content publishers either "push," or broadcast, their information to receptive users, or let users "pull" content by visiting the publishers' websites, clicking on links, and the like. Those publishers and sites rely on a marketplace of ideas: they provide information, and leave it up to their readers, viewers, or listeners to form an opinion about it. When it comes to social media manipulation and microtargeting, however, platforms, advertisers, and public relations firms use an array of tools to subtly influence how we receive the content that is being pushed

to us. They use detailed, individualized profile information to identify whom to target—based on our demographics, shopping habits, online friends, and more—and then assess how to *most effectively* target us, based on the platforms' assessments of our individual personality traits, such as how much we sympathize with people who are different from us and how drawn we are to autocratic tendencies.

This kind of individually targeted viewpoint manipulation has been used to shape public opinion about the protests in Ferguson, Missouri, and Baltimore, Maryland; to influence Britain's Brexit vote and the 2016 US presidential election; and to try to sway the elections in Germany and France in 2017 and 2018. These influence campaigns rely on a mix of tools: automated bots and human trolls create propaganda and fake news; they leverage users' personal data to identify specific individuals likely to be susceptible to their messaging; and, at very little cost, they flood users' feeds with a nearly endless supply of posts that make it appear as though there's a groundswell of opinion in favor of a certain candidate or idea.

Social media platforms are making it possible for manipulators to use our own personal data as the most formidable weapon against us, and to great effect. National security experts fully expect that foreign adversaries will try to interfere with the 2020 US election cycle. The risk isn't limited to foreign governments. In 2019, Facebook's founder, Mark Zuckerberg, claimed that an Elizabeth Warren presidency—with a renewed focus on antitrust regulation, privacy protection, and wealth taxation—would be the social media platform's worst nightmare.[2] Zuckerberg's comments raise a whole new set of concerns about the ways in which the owners of these social media empires are not only profiting from others' use of the platforms to manipulate opinion, but also how the platforms themselves might target us with individually tailored messages designed to shape our opinions to suit their own political goals.

Social media platforms are making it possible for manipulators to use our own personal data as the most formidable weapon against us, and to great effect.

In March 2019, Robert Mueller's Office of Special Counsel released a two-volume report on Russia's interference with the 2016 US presidential election.[3] Based on nearly two years of investigation, Mueller and his team concluded that the Russian government interfered in a "sweeping and systematic fashion." One prong of that interference was an "active measures" campaign that included using fake social media accounts to sway opinion in the United States. The Russian government worked through a St. Petersburg–based troll farm known as the Internet Research Agency (IRA), run by a close associate of Russian president Vladimir Putin. Beginning in 2014, IRA employees created fake accounts on Facebook, Instagram, Twitter, and other social media platforms. The accounts, which varied widely, were designed to look like authentic accounts associated with real people living in the United States.

One of the most widely followed IRA trolls was a Twitter account with the handle @TEN_GOP—an account that fooled hundreds of thousands of Twitter users into believing that it was the "unofficial account" of the Tennessee Republican party. Another account, @Jenn_Abrams, famously duped her 70,000 Twitter followers. The fake account posted tweets about pop culture, ballistic missiles, the Confederate flag, and Rachel Dolezal, a white woman who's known for self-identifying as black. "Jenn Abrams's" persona was so persuasive that her tweets were featured in articles in *USA Today*, the *New York Times*, *The Daily Caller*, *BuzzFeed*, and a host of other US and international news outlets. @Jenn_Abrams followed a modus operandi common to many of these troll accounts: build up a following with an entertaining and engaging online persona that posted content about nonpartisan issues—celebrity gossip, general news—and then, after attracting a substantial following, start pushing out deeply divisive content on wedge issues in American politics, like immigration, race, gay rights, and, closer to the 2016 election, content deeply critical of, or spreading conspiracy theories about, Democratic nominee Hillary Clinton. Abrams was particularly successful at creating a total package of a person; in addition to her Twitter account, she also had a personal website, a Medium page, a Gmail address, and a GoFundMe page.[4]

In 2019, the *Columbia Journalism Review* published research showing that most major news outlets had at some point in time unknowingly quoted

Russian hoax accounts, usually in stories reporting on public reaction to recent events. News outlets with a more left- or right-leaning bent, such as *The Daily Caller* and *Huffington Post*, meanwhile, were more likely to report on the often incendiary social media posts. But the challenge of differentiating authentic accounts from fake ones—the problem of differentiating the true from the false—hit major mainstream news outlets in the center of the political spectrum as well, including such highly regarded ones as the *New York Times*, NPR, and the *Washington Post*.[5]

The Russian active measures campaign was carried out on every major social media platform. On Facebook, for example, the IRA controlled 470 accounts that made over 80,000 posts between 2015 and 2017, reaching some 126 million people.[6] These accounts began creating and sharing openly pro-Trump and anti-Clinton posts and reaching out to pro-Trump groups. They even contacted Facebook followers via private direct messages to encourage them to show up for live, in-person rallies in the United States—rallies that were staged by the IRA, acting through its Russian-government-backed trolls from St. Petersburg, thousands of miles away.[7]

Neither the criminal indictments against the IRA nor the widespread news coverage about Russia's fake social media profiles and influence campaign seem to have had any deterrent effect. Russia's troll farms were accused of attempting to influence the 2017 French presidential election and the 2017 German federal election, and of stoking France's "yellow vest" protests in 2018. Russian troll farms were also implicated in using Facebook, Twitter, and YouTube accounts in their attempts to sway European Union elections in 2019.[8] The playbook was the same as it had been in the United States in 2016: suppress voter turnout, deepen political divides, and advance a far-right policy agenda through fake accounts, disinformation, and the exploitation of local political divisions that already existed. Adding insult to injury, in 2019, a Russian troll farm that was a close cousin to the IRA, the Federal Agency of News, filed a lawsuit in US federal court, charging Facebook with violating its First Amendment rights for kicking it off the platform.[9]

The problem is increasingly well documented. Reports from the UK Parliament have laid out in excruciating detail the ways in which social media disinformation is fueling social discord and political divisiveness, as well as

uninformed or misinformed political thinking. The Australian Competition and Consumer Commission has issued a report describing the ways in which social media platforms are not only spreading misinformation but also driving legitimate news outlets out of business.

It's important to be clear what the risk is here: the pernicious fact of social media's role in this shifting information landscape is that social media makes it so easy to feed the trolls. Because social media platforms know so much about our individual behavior, interests, personality traits, and inclinations, and because of the way that their advertising and engagement models work, they are the perfect mechanism for people with a message—including distorted, damaging, false, or divisive messages—to individually target their content toward each one of us.

Meanwhile, the platforms are eschewing any responsibility for their role in fomenting social divisions or political unrest. Facebook has mounted an aggressive defense to the class action lawsuits filed against it by users who allege that the platform has violated their privacy by providing their personal data to the political advertising firm Cambridge Analytica, which then used it for microtargeted messaging without the users' knowledge or consent.[10] As described by the court overseeing the litigation, "Facebook argues that people have no legitimate privacy interest in any information they make available to their friends on social media." Further, according to the plaintiffs' allegations, although Facebook had a nominal policy restricting third-party access to Facebook user data, in effect no such policy existed at all, because "with the tens of thousands of app developers who interacted with users on the Facebook platform . . . Facebook was intent solely on generating revenue from the access it was providing."[11]

It's hard to know exactly how much money Facebook earns from political ads compared with other kinds of advertising, but when confronted with concerns, Mark Zuckerberg has consistently taken the position that, unlike some other platforms, Facebook will continue to sell advertising to political campaigns.[12] It will also continue to benefit from the indirect revenue generated by increased user engagement and time-on-platform that's created by unpaid troll accounts. And in a twist that underscores the dangers that Zuckerberg's personal views could be used to drive Facebook's content moderation,

it turns out that some of the biggest political advertising dollars spent on Facebook are spent by Zuckerberg himself.[13]

HOPE ON THE HORIZON?

Think tanks like the Alliance for Securing Democracy at the German Marshall Fund have pointed to the destabilizing effect across all of the Western democracies of authoritarian and nationalist movements that are being fueled in part by information operations that include these individually directed and microtargeted political messaging campaigns.[14]

Academics like Kathleen Hall Jamieson at the University of Pennsylvania and Briony Swire-Thompson at Northeastern University have done extensive research on the phenomena of fake news, Russian troll farms, and targeted social media political advertising. Jamieson, a prominent expert in public policy and political communications strategies, was one of the first to do a comprehensive social science analysis of the impact that Russia's information operations had on the 2016 US elections. Her conclusion: the Russian active measures campaign was powerful and effective in shaping US public opinion, and likely changed the outcome of the election in favor of Donald Trump.[15] Swire-Thompson's research has a somewhat different focus, examining the cognitive mechanisms behind our susceptibility to "fake news." In one of her recent papers, she notes that when people are attempting to assess the trustworthiness of information, we too often default to cognitive biases like, "They might be a liar, but they're my liar."[16]

In looking for solutions, some policymakers have suggested new legislation like the Honest Ads Act, which would require political advertising to carry funding disclosure statements much like those that are required for radio or television ads. Others have correctly pointed out that, while microtargeting happens through paid advertising, there should also be more content moderation, flagging bots and trolls, taking down inauthentic accounts, and removing offensive content.[17] Other policymakers focus on the need to emphasize critical thinking skills in school but tend to overlook the research showing that younger people are generally more savvy about the need to be

skeptical of internet content. It's older voters who are more likely to be swayed by slanted content or deceived by outright falsehoods that are posted online.[18] Given those demographic distinctions, it's clear that K–12 or college education alone isn't nearly enough to combat a cultural receptivity to misinformation, especially when it has been microtargeted to individually exploit us.

Finland might have an answer. The country's population of 5.5 million has been wary of Russian threats to its national security ever since declaring its independence from Russia over a century ago.[19] As part of an initiative launched in 2014 to counter Russian information warfare that attempted to stoke divisions within Finland over issues such as immigration, the European Union, and NATO, Finland has taken a multi-pronged approach to citizen education that includes classroom projects at all age levels as well as adult education programs in local community centers. The lessons address how to identify bots on Twitter, how to spot deep fake videos, how to spot slanted news coverage and identify an outlet's or an article's biases, and more. These measures alone may not be enough to counter the ways in which our individual information is being used to manipulate our views. But it appears to be having positive effect, both in maintaining civil discourse in Finland and in preserving free and fair elections there. Countries across Europe and from other parts of the world are looking to Finland's experience to create critical thinking and public education programs of their own. As part of a larger toolkit, Finland's lessons could contribute in important ways to countering the negative effects of platforms that know the details about us entirely too well.

CLICK TO ORDER *CYBER PRIVACY* FROM YOUR FAVORITE VENDOR

amazon.com

BAM! BOOKS·A·MILLION

BARNES & NOBLE BOOKSELLERS

Indigo Books & Music Inc.