Excerpts from:

Why Fear Data

Omri Ben-Shahar
University of Chicago Law School

(Under contract, Harvard University Press)

Note to NYU's Innovation Policy Colloquium:

Why Fear Data is a book that challenges how we think about the technologies that process personal data. Let me briefly bring you up to speed—a background to the chapters enclosed.

Our data protection laws, I argue, are barking up the wrong tree. They diagnose the central problem of the digital society as one of *data privacy* and seek solutions in the form of privacy protections. People are given legal rights to determine how their data are tracked (and shown countless daily reminders of these rights in every website.) The cardinal flaw in this scheme is one of misdiagnosis. By focusing on private rather than social harms – on the potential intrusions suffered by individuals whose personal data are collected, instead of the effect on public institutions – it misses the bigger picture.

This misdiagnosis is responsible for two striking distortions. First, big data's most troubling societal harms – things like algorithmic discrimination, dissemination of fake content, political polarization, erosion of communal norms, social media's impact on youth, or harmful meddling by foreign governments – remain beyond the reach of data privacy laws with their focus on the well-being of individuals rather than environments. Second, many of data analytics' promising benefits, like reduced auto accidents, improved health treatments, and prevention of heinous crimes, known to save thousands of lives, millions of injuries, and trillions of dollars, are tragically restricted by heavy-handed data privacy laws.

In a part of the book not distributed here, I develop a theoretical framework of what I call "data pollution" – an account that shifts the lens from private injuries to social harms. I apply that framework to suggest various interventions, of the types deployed to control industrial pollution and other externalities—a "data environmental law." (I'll briefly demonstrate this argument in my colloquium opening remarks.)

The two chapters here consist of part III of the book, where I examine another blind zone of data privacy law – the indifference to data's social benefits. I show how privacy-protective regulation of data technologies restricts data applications with enormous social value. Surprisingly, this question – privacy protection, at what cost? – is rarely studied. Exposing the overlooked costs of privacy law could reshape the design of data protection law.

Part III

At What Cost?

Part II shifted the lens from private to social harms. Many serious harms arise from the mass use of data to train and employ the algorithms running our daily lives, and we saw that these harms are primarily social, rather than private. I introduced the data pollution framework to help us look beyond data privacy and address this broader, public scope.

If the effects on public goods and social interests ought to take center stage, the discussion cannot be restricted to data's harms and must be expanded to include data's social benefits. There is data pollution, but there are also "data greens." This is what Part III will now do: examine how the regulation of data technologies affects society's opportunities to capture these technologies' potential for social value.

These social benefits take many forms: millions saved in life and limb, billions in dollars and resources, and a substantial reduction in misallocation, discrimination, and litigation. Realizing this value is especially appealing because it requires minimal investment; the technologies rely on data processing rather than costly infrastructure. And yet regulatory roadblocks slow and sometimes block their adoption.

Chapter 7 presents prominent data technologies that have benefits so undeniably significant that they outweigh even the most pessimistic assessment of their effect on data privacy, and yet they face crushing resistance from lawmakers and privacy advocates. The examples in this chapter come from important parts of life: tracking tools that reduce highway crashes and fatalities, medical databases that improve health outcomes, biometric recognition technologies that rescue victims of heinous crimes, and AI models that reduce discrimination in many areas. As diverse as these data technologies are, the grounds for their resistance and the magnitude of the stated concerns are surprisingly uniform, rarely scaled to the specific threats, and always entirely invariant to the upsides. What Chapter 7 shows is a pattern of misalignment that is rarely if ever discussed: a dramatic gap between the private harms that privacy regulations hope to forestall and the social benefits they are ready to surrender.

Chapter 8 then examines the underpinnings of this resistance to data technology, deeply bound in what I call "data precautionism," a philosophy whose application is not only invariant to technology cost and benefits but triggers a heavy regulatory hand even when least justified. Data precautionism applies even when accumulated experience has reduced the outcome uncertainty, showing that the foreseeable benefits easily outweigh foreseeable harms. t More strongly than in earlier chapters, readers will notice—and, I hope, share—my exasperation as I illustrate data precautionism's tragic costs.

Chapter 7

Data's Benefits

> "Some of you may die, but it's a sacrifice I am willing to make."
>
> — **Lord Farquaad, Shrek**

Data technologies have social benefits.

This is a well-kept secret, hidden in plain sight, scarcely nodded towards in the massive legal and ethical commentary on data technology. It is simple but electric, and it ought to change how we think about data regulation, and specifically about data privacy laws.

You must be looking at me with bewilderment. A "secret"? Isn't it obvious that data technologies have benefits which we all richly enjoy every moment of every day? GPS apps are more useful than paper maps, Uber is more sufferable than taxi, and internet search—the most convenient index of information—blows away the card catalog. Yes, numerous data services bring us massive *private* benefits, but what is often overlooked—what eludes data privacy law, policy discussions, and academic writings—is any serious discussion of data's *social* benefits. Social benefits, like social harms, accrue to society, creating synergies greater than the sum of private functional conveniences. They go beyond the simple utility of data services embedded in countless apps and gadgets.

Whether we choose to call these positive externalities, public goods, aggregate improvements, lives saved, equity, or inclusivity, data technologies generate such outcomes abundantly, frequently, and dependably. And yet, entire aisles in law libraries are now devoted to discussions of nebulous privacy harms, an echo chamber that is largely disinterested in, and frequently dismissive of, data's benefits.

You remain skeptical. I get it—this sounds as an unfair overstatement. Let me then present several striking illustrations. Here are three data technologies that utilize some of the most sensitive personal information: people's driving habits, faces, and health. These technologies are subject to some of the stiffest regulations, enacted through laws whose only goal is to protect data privacy. When confronted with the forceful benefits of the technologies they restrict, these regulations do not budge, instead receiving reinforcement from advocates who uniformly applaud the restrictions and demand intensification. But these technologies do have dramatic upsides. Endless lives can be saved, victims of horrific crimes can be rescued, and progress in health care can be vastly accelerated. Harms of true importance can be avoided.

Let me, then, turn to my first illustration, the site of much unnecessary death: the roads. From a perspective you've probably never considered, I offer the story of a data technology that costs nothing, offers enormous lifesaving benefits, but succumbs to the data privacy project. It sits at the forefront of an auto insurance revolution, and while "insurance" may feel like anesthesia, fear not. There's no boredom in this tale. Heartache? Perhaps.

TRACKING HOW YOU DRIVE (BUT NOT IN CALIFORNIA)

> “If you know that [insurers are] monitoring the way you’re braking or how you’re driving ... you better believe that’s going to change your driving behavior.’’
>
> - California Insurance Commissioner (2019)

> “We won't bend on protecting consumer data, privacy, and fair rates. … Since 2009 we allowed vehicle data only to determine actual miles driven, and only in a way that protects the driver’s privacy.”
>
> - California Insurance Commissioner (2012)

Road accidents are a major cause of injuries and death. More than 40,000 people die every year in U.S. Reckless drivers kill more than 15,000 pedestrians and cyclists. Five million people are injured in crashes, many of them disabled, and the overall cost of car accidents is half trillion dollars.[1] Making roads safer is possibly the last remaining bipartisan aspiration in America, and while cars are increasingly built safer, casualties on the road have not shrunk. Seatbelts and airbags reduce the severity of injuries but not their incidence, and—in a canonical illustration of the phenomenon of *unintended consequences*—they may impart a sense of invincibility among some drivers that causes them to be more reckless.

The New Technology

Against this grim backdrop, a new data technology emerged and has gradually become the most cost-effective auto-safety technology ever. I’m not kidding. First introduced by auto insurers for an entirely different purpose—for more accurate pricing—this technology involves no driving restraints, engineering feats, or protective gear. Nothing physical. It is just a formula, an algorithm processing data. Everyone knows how lenders use borrowers’ financial data to predict lending risks, establish credit scores, and personalize the loans accordingly. Well, insurers adopted the same approach. Based on billions of miles of driving data, they trained algorithms to identify accident-causing behaviors. Then, they began collecting information on how policyholders drive their insured cars in order to ascribe them personalized “safety scores.” Better scores mean lower insurance premium.

Not everyone is subject to this program. Policyholders must opt into “usage-based insurance” (UBI) and agree to install tracking devices in their cars or activate tracking functionality in their apps. If they do not enroll in the program, they’ll continue to be rated in accordance with the standard demographic factors of auto insurance. But if enrolled, the insurance algorithm adjusts premiums based on uncontroversially relevant factors like sharp acceleration, aggressive turns, hard braking, unsafe following, and smartphone distractions. It also provides policyholders feedback while they drive and digital dashboards to understand their scores and the reasons they periodically change.

[1] National Safety Council, *Injury Facts Motor Vehicle Overview*.

By definition, this is an invasive data technology: it records how, where, and when people drive. Seeing only this side of things, intrusion into the private sanctuary of the driver's seat, various states' data privacy laws limit, and even prohibit, the use of UBI technology. But this represents only part of reality. On the other hand, driving is a very dangerous public behavior conducted on busy public roads. Immediately, you see the tension that made me choose this example. If there are safety benefits to UBI technology, squandered as result of such privacy-based regulatory prohibitions, what are they?

The Safety Effect

The key to understanding the social benefit of UBI lies in how drivers are affected by this new pricing algorithm. Premiums are personalized with real time data about policyholders' driving. As drivers learn to anticipate how their premiums would adjust, they pursue less costly, safer driving. This change is the critical benefit driving my interest in UBI technology. However, it doesn't stop there; in a minute, I will also succumb to the temptation to thoroughly outline UBI's co-benefits. I will explain why I think it fair to charge reckless drivers higher premiums ("carry their own weight", in insurance lingo). And even more so, why it is distributively just to replace the traditional social-demographic rating factors like income, gender, credit scores, or education—which have long disfavored the poor—with a less discriminatory pricing scheme. But this sermon must wait. Far and away the most important impact of UBI is the reduction of accidents and road fatalities. Plain and simple: when people are tracked, they drive more safely.

Why? The obvious reason is financial—when scores go up, premiums go irresistibly down. People care about this effect because the cost of auto insurance is a meaningful component of household budgets. This is unlike other incentives. Traffic fines are incurred only probabilistically, the risk of crash injury is a motivator only when immediately salient. Insurance premiums, however, are paid continuously with reminders upon each monthly payment. It is a cost that people notice and—in contrast to other types of insurance—people also sense control over these costs.

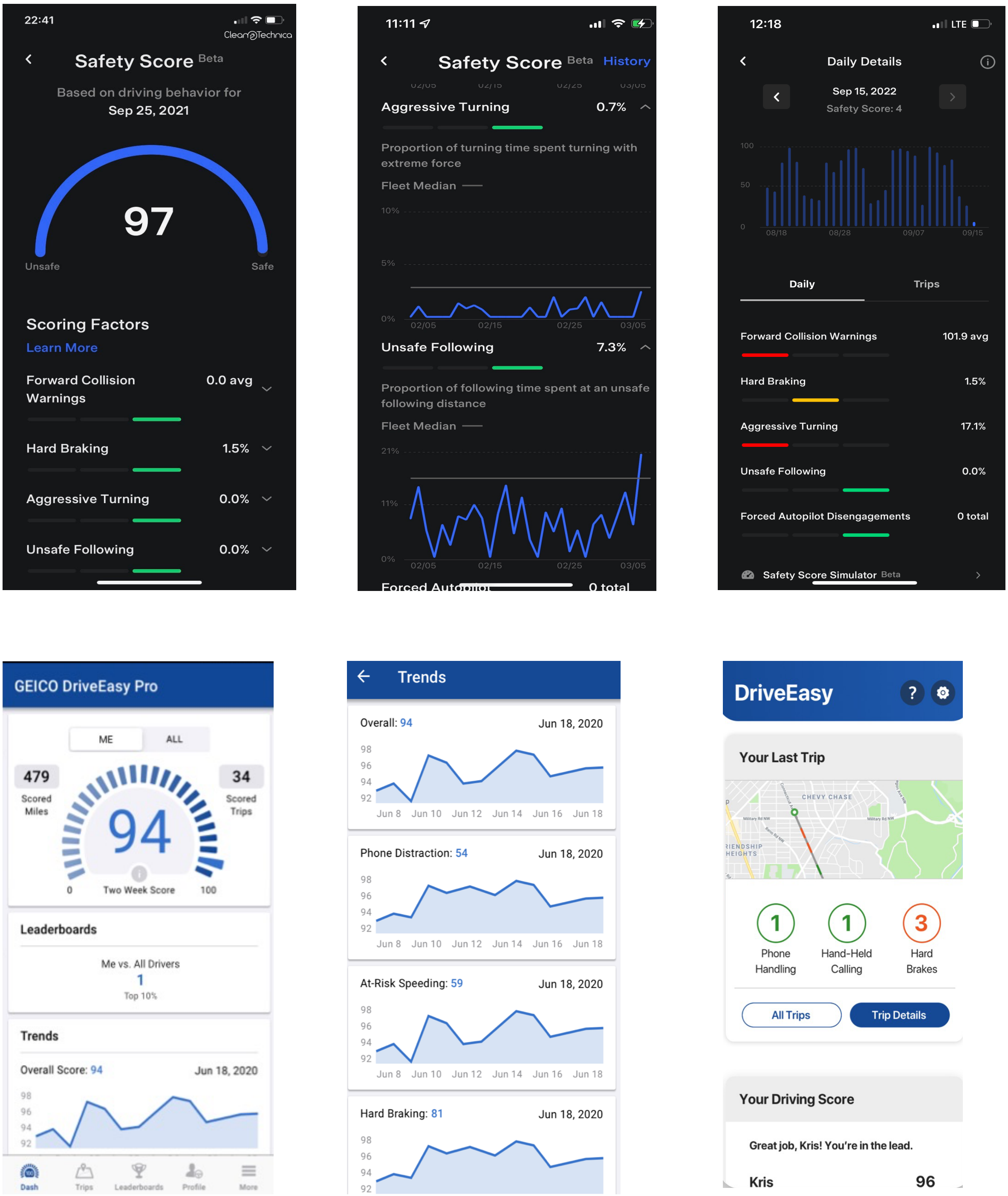
The second reason is pedagogical: they are coached about safe driving. Their dashboards beep every time they engage in risky maneuvers or get too close to other cars, and their apps visualize and explain these factors (see Figure). This helps overcome "illusory superiority," a classical social psychology concept, whereby people tend to overestimate their skills. Nowhere is this better-than-average bias more pronounced, and more repeatedly documented, than in driving, where over 90 percent of people say that their driving skills are above average.[2] In other words, people think they are better drivers than they truly are. Resist the hunch that this is caused by the infrequency of accidents—the absence of opportunity for drivers to reevaluate their immodesty. In fact, self-serving assumptions about causality resist evidence: even drivers adjudged responsible for accidents deflect any reevaluation of their driving competence.[3] Undoubtedly, this overconfidence, this sense of invincibility in confronting the dangers of the road, leads to risky behaviors. A major step towards accident reduction, then, is to emancipate

[2] Ola Svenson, Are *We All Less Risky and More Skillful than Our Fellow Drivers?*, ACTA PSYCHOLOGICA 47, no. 2 (1981): 143–148.

[3] Caroline E. Preston & Stanley Harris, *Psychology of Drivers in Traffic Accidents*, 49 J. APPLIED PSYCH. 284, 286 (1965).

people from that misjudgment. The educational tools of UBI, its live feedback, may overcome people's reluctance to learn. It's like an on-the-fly driver's-ed for the stubbornly delusional.

FIGURE HERE



22:41

CleanTechnica

Safety Score Beta

Based on driving behavior for Sep 25, 2021

97

Unsafe Safe

Scoring Factors

Learn More

Forward Collision Warnings	0.0 avg
Hard Braking	1.5%
Aggressive Turning	0.0%
Unsafe Following	0.0%

11:11

Safety Score Beta History

Aggressive Turning 0.7%

Proportion of turning time spent turning with extreme force

Fleet Median

Unsafe Following 7.3%

Proportion of following time spent at an unsafe following distance

Fleet Median

Forced Autopilot 0 total

12:18 LTE

Daily Details

Sep 15, 2022

Safety Score: 4

Daily Trips

Forward Collision Warnings	101.9 avg
Hard Braking	1.5%
Aggressive Turning	17.1%
Unsafe Following	0.0%
Forced Autopilot Disengagements	0 total

Safety Score Simulator Beta

GEICO DriveEasy Pro

ME ALL

479 Scored Miles

94 Two Week Score

34 Scored Trips

Leaderboards

Me vs. All Drivers

1

Top 10%

Trends

Overall Score: 94 Jun 18, 2020

Trends

Overall: 94 Jun 18, 2020

Phone Distraction: 54 Jun 18, 2020

At-Risk Speeding: 59 Jun 18, 2020

Hard Braking: 81 Jun 18, 2020

DriveEasy

Your Last Trip

1 Phone Handling

1 Hand-Held Calling

3 Hard Brakes

All Trips Trip Details

Your Driving Score

Great job, Kris! You're in the lead.

Kris 96

A third reason, and this is my own speculation, is psychological: when people know they are "being watched" certain cognitive faculties activate, making decisions more deliberate and less impulsive. One mechanism is fear—that the spectator, the scoring algorithm, will inflict a monetary sanction.[4] Its complement is reward—where the satisfaction from achievement, intensified by the concrete manifestation of a score increase, encourages driving that recaptures this satisfaction.[5] A behavior that is otherwise thoughtless, aggressive, and temperamental becomes more reasoned, judicious, and safe.

How Many Lives Saved?

Combined, these mechanisms are forceful. Studies measuring their effect show that, within a month of UBI enrollment, personalized risk scores improve and accidents decline. I surveyed the entire social science literature that used measured this effect. Without exception, they all report significant effects on road safety. But the real surprise lies in the sheer scale of the impact. Looking at the empirical findings altogether, the estimated decline in fatal accidents is in the range of 30 percent! If every car in America were required to adopt the tracking technology, 12,000 lives would be saved annually!

One study, for example, found that when of 9% of policyholders switched to UBI the total fatal accidents fell by 4.61%. For 9% of drivers to explain a 4.61% aggregate reduction in fatalities they must have experienced a 51% reduction in fatal accidents.[6] Of course, participants may be a non-representative group, but if anything, they are likely to be the safest drivers, eager to separate themselves from the insurance pool and enjoy the high-score discounts. Another study examined how the safety score of one million drivers changed after they enrolled in UBI. It concludes that "consumers who opt in to monitoring become 30% safer, on average, while they are being monitored."[7] A third study compared UBI participants to non-participants and found that within a couple of months participants decreased their hard-brake frequency by 21%.[8] Hard braking reflects risk factors like speeding and unsafe following, and long term estimates suggest that a 5% reduction of speeding may lead to as much as 10% reduction in injury accidents and a 20% reduction in fatalities.[9] Thus, if only half of the 21% decrease in hard brakes is due to lower speeds, the decline in fatalities resemble the 30 percent range found in other studies.[10]

What does all of this tell us? If road safety is a social concern, every car in America should be required to adopt the tracking technology. UBI should be mandatory, just like seatbelts and airbags. In the two decades since the technology emerged, hundreds of thousands of lives could have been saved. This, without budgets, delay, or new taxes. Just the sweet dividend of sizable

[4] Patricia A. Ellison et al., "Anonymity and Aggressive Driving Behavior: A Field Study," *Journal of Social Behavior and Personality* 10, no. 1 (1995): 265, 270–71; Lior J. Strahilevitz, "'How's My Driving?' for Everyone (and Everything?)," *NYU Law Review* 81 (2006): 1699–1765, 1744.

[5] Dimitris Karapiperis et al., *Usage-Based Insurance and Vehicle Telematics: Insurance Market and Regulatory Implications*, Center for Insurance Policy and Research, March 2015, 24–25.

[6] Imke Reimers & Benjamin R. Shiller, *The Impacts of Telematics on Competition and Consumer Behavior in Insurance*, *Journal of Law and Economics* 62, no. 3 (2019): 613, 628-29.

[7] Yizhou Jin & Shoshana Vasserman *Buying Data from Consumers: The Impact of Monitoring Programs in U.S. Auto Insurance*, National Bureau of Economic Research, Working Paper Series No. 29096 (2021).

[8] Soleymanian et al., *Sensor Data and Behavioral Tracking: Does Usage-Based Auto Insurance Benefit Drivers?*, *Marketing Science* 38, no. 1 (2019): 21–43, 40.

[9] OECD/ECMT Working Group, *Speed Management* report (2006): 39.

[10] There is also a long list of experimental findings that support this estimate of accident reduction. *See* Omri Ben-Shahar, Privacy Protection, At What Cost? Exploring the Regulatory Resistance to Data Technology in Auto Insurance, *Journal of Legal Analysis* 15, no. 1 (2023): 129, 138–39.

insurance discounts flowing to poor communities' good drivers, who've suffered most under demographic rating's tyranny. More safety, more fairness, more savings—lights out.

Regulatory Restrictions

This is a book on "why fear data," not "why fear roads." Where, then, is this story headed? It turns out that despite such phenomenal and *undisputed* life-saving benefits, fairer premium allocation, and enrollment being entirely optional, UBI fights an uphill battle against data privacy regulation. Some states, in the name of consumer protection, have a "discount only" rule, meaning that insurers are not allowed to "downtier" bad drivers and raise their premiums. (I admit that I'm struggling to understand how this counts as "consumer protection.") These states also say that data on texting-while-driving may not be used in computing a driver's overall safety score. And that scoring algorithms must have incredibly short-term memory: distracted driving scores must be "refreshed at each policy renewal."[11] Why are lawmakers so ticklish about texting-while-driving, seeking extra protection for the most reckless of all traffic violations? It's because we've constructed a privacy halo around personal devices. If smartphones are private, then likewise is tracking the timing of smartphone use. Sure, if a driver is observed texting by police, it's ok to penalize them, as there is no digital tracking involved. Collecting data about such violations through electronic "surveillance" is whole different ballgame because it bumps against what privacy advocates regard as the fundamental right to be let alone, free of "intimate invasion" and "deep body periscopes."

These petty regulatory restrictions on how the use of the tracking data are eclipsed by the most extreme opposition to UBI, found in California's explicit and unconditional prohibition on real time tracking of drivers. The state has been firm that no data besides miles driven may be tracked for insurance purpose, even if drivers agree. For a fleeting second, the state's Insurance Commissioner did recognize that UBI "can save lives" and admitted that "breathing new life" into the "antiquated" prohibition is warranted.[12] But his lightbulb moment would soon dim, after drawing strong condemnation from privacy advocates. What's more, as one might expect of an elected, ambitious politician, he'd later glow with the fashionable pathos of privacy protection. "We won't bend on protecting consumer data, privacy, and fair rates" he announced.[13] Autonomy for drivers, freedom of contract, road safety, affordable insurance—all these other good things? Not in California.

UBI is where I chose to start this Chapter because it is a bookend of sorts. It easily shows the massive gap between data' benefits and their possible harms, and how the absence of social lens leads to false alarm. It is a case of false alarm—resistance to a technology with enormous

[11] See New York Department of Financial Services, *Updated Guideline for New York UBI Programs (Plug-in Telematics Devices and Smartphone Apps)*, § 10, declaring, "The data collected for the UBI program will not be used to affect policyholders in a negative way (e.g., increasing premiums (including application of surcharges), non-renewing policies, preventing downtiering, etc.)."; see also *Ibid.*, Additional Guideline for Smartphone-based UBI Programs, § 6a, specifying, "A company may collect distracted driving statistics; however, such statistics may not be used in the algorithm to determine the final UBI score/factor."; see also *Ibid.*, § 6ab, providing that, "A company may establish a separate distracted driving discount … provided the score/factor is refreshed at each policy renewal."

[12] Carla Marinucci & Jeremy B. White, "Lara Tells Insurers He's 'Receptive' to Their Ideas, Including Vehicle Data Use," *Politico*, July 29, 2019.

[13] Ricardo Lara (@ICRicardoLara), Twitter, January 27, 2022.

value and only weak non-consequentialist downsides. We have an opportunity to embrace a technology that removes much of the ignorance about how insurance is priced and teaches people to accomplish the impossible—driver better. A regulatory paradigm that impulsively forfeits such advances is bad.

UBI exposes our regulatory regime. What kind of paradigm reflexively forfeits social advance this profound? How could a state with 4000 annual highway fatalities prohibit—rather than permit (or, perish the thought, mandate)—a technology that could save 1200 lives annually? We'll explore the answers at the next chapter, what lies underneath the data privacy dogma. In preparation, though, let's look at a few more slain innovations.

FACIAL RECOGNITION IN HUMAN TRAFFICKING

> "It's a technology that has such a profound potential to erode the very fabric of human society that any potential benefits are outweighed by the inevitable harms."
> - Kashmir Hill, Your Face Belongs to Us

> "Over the past three years, law enforcement have reported using [facial recognition] in more than 21,000 investigations and to identify more than 18,000 trafficking victims and more than 6,000 traffickers."
> - Letter by twenty-one state Attorneys General

The second demonstration of data's underappreciated social benefits come from an area loaden with the strictest limitations: data-powered facial recognition technology.

Human trafficking is the world's second largest and fastest-growing criminal activity. Up to 28 million people worldwide are subject to forced labor, much of which involves sex trafficking. Tragically, 20% of these victims are children, exploited by highly profitable and sophisticated interstate criminal rings.[14] A modern day slavery, this is one of the hardest crimes to combat because victims are rarely allowed to appear in public. As state Attorneys General acknowledge, "with over 150,000 escort ads posted in the country every day, law enforcement cannot manually keep up with the volume of ads as they work to identify and track potential victims."[15]

The New Technology

Thankfully, there are still chinks in its armor. First, human trafficking has digital footprints, leaving a trail of images showing its victims and perpetrators on the dark web and even on social media. Second, trafficking is a cross-border phenomenon, meaning victims and their

[14] International Labour Organization, *Profits and Poverty: The Economics of Forced Labour*, March 19, 2024, reporting the number of victims and estimating the profits of criminal enterprises at $236 billion.

[15] See Letter from Attorneys General to Congressional Appropriations Subcommittees regarding Spotlight Sex Trafficking Investigation Tool, 2018, at https://www.scag.gov/wp-content/uploads/2018/05/Final-Spotlight.pdf.

kidnappers are scanned by cameras as they pass through airports.[16] In the old day, spotting such trafficking was difficult, slow, and inaccurate; law enforcement and border officials had to sniff out behavioral indicators of suspected trafficking. Today's facial recognition technology changes the game. Law enforcement agencies may use facial recognition software in analyzing online images, match them with profiles of missing persons (including their AI-generate age-progression images), and use the data to identify trafficking operations. They can share these data with authorities in other countries, increasing the chances of identification and rescue during victims' rare public appearances. These rescue efforts depend on the use of biometric search tools and particularly on access to private databases of facial images.

Facial recognition tools rely on large databases of facial images and on AI technology trained on these images to identify individuals. At the development stage, the software is taught to detect, recognize, and classify features of people and their surroundings. The training set includes individual identifiers, which allows the algorithm to then perform a screening function and, when shown a new image, recognize the person in it. Advances in AI technology enable great accuracy in identification even when a facial image is obscured, outdated, clipped, or otherwise unidentifiable to the human eye.

Facial recognition software is widely used for one-to-one verification ("are you who you say you are?")—a popular function in unlocking smartphones, entering buildings, or passing airport security. It can also be used much more controversially in surveillance—the one-to-many function—to identify an otherwise unknown person. The most common applications of this capacity are in law enforcement. Police may receive images from suspected crime scenes—taken by street cameras or buildings' CCTVs, posted on social media or in escort ads—and find matches for these images within various databases, providing leads on the identifies of the recorded individuals. The relevant databases include missing persons, past offenders, and—increasingly—everyone in society. In much the same way that other biometrics such as fingerprints or DNA traces have long been used to identify crime suspects, facial recognition technology is the new frontier in detection and identification.

The technology has additional facial-screening capabilities, such as classification of sex and race, emotional and fatigue recognition, fitness assessment, and attention monitoring in operation of trucks and self-driving cars. Indeed, it is a tool that has many potential uses stretching well beyond criminal investigative leads—uses that could be socially valuable but that also ring the strongest privacy alarms. I focus in this chapter on the value in law enforcement, which I argue must be a key component in evaluating the privacy restrictions.

Victims Rescued

What concrete evidence is there of the benefits of facial recognition technology in law enforcement? I would love to have found precise estimates of the magnitude—of the trafficking victims rescued through facial recognition methods, but all I have is a collage of reports. Still, these will do. These snapshots give ample indication of the technology's benefits and its vast potential. Here are some examples:

[16] United Nations Office on Drugs and Crime, *Global Report on Trafficking in Persons*, 2020, p. 10.

- Twenty-one state Attorney Generals wrote to Congress that a facial recognition tool was used to identify more than 18,000 trafficking victims and more than 6,000 traffickers, while reducing investigative time by up to 65%.[17]
- A nonprofit company named Thorn developed a tool that uses Amazon's facial recognition software and reported that it has been used by law enforcement on almost 40,000 cases in North America, in which investigators found more than 9,000 children, and over 10,000 traffickers.[18] While the exact numbers Thorn reports have not been audited, the use of the tool is prevalent.
- In a 2023 international operation, the Department of Homeland Security used a private facial recognition tool to identify 311 children who appeared in sexually graphic materials and succeeded to rescue some of them from active abuse.[19]
- In India, over 3,000 missing children were rescued in 2020, many of them while held as laborers, in a month-long operation that relied on the Darpan facial recognition app. Follow-up operations with similar success are taking place each year.[20]
- A coordinated project in the Philippines rescued 23 victims (22 minors).[21]

Numerous specific reports provide detail on individual cases in which sex traffickers and violent offenders have been arrested and their victims rescued after initial identification by facial recognition tools. There are credible first-hand testimonies from crime investigators who work with these tools, noting the impact of the software in providing investigative leads. One veteran human trafficking DHS officer stated that "no single effort like [facial recognition] has resulted in that amount of identifications in such a short period of time."[22]

Curiously, this type of evidence rarely makes headlines. The vast literature on facial recognition technology—government reports, academic commentary, and especially the popular media—chooses to focuse exclusively on data privacy and problems of misidentification. It rarely if ever mentions any benefits, viewing them as negligible and speculative anecdotes. Here, look at a recent report from the Government Accountability Office, solicited by the Senate in response to facial recognition privacy alarms. It surveys enforcement agencies' use of the technology, briefly noting that it was used in 63,000 crime searches. The social upsides from these successful leads are an afterthought, and the report does not bother to discover the enforcement outcomes in these cases. Without any evidence of actual violations, its 70 pages dissect the endless metrics of what it regards as enforcement agencies' insufficient privacy training in the use of the tools and the various categories of missing procedural guardrails.[23]

[17] See Letter from Attorneys General to Congressional Appropriations Subcommittees regarding Spotlight Sex Trafficking Investigation Tool, 2018 (https://www.scag.gov/wp-content/uploads/2018/05/Final-Spotlight.pdf.)

[18] See Tom Simonite, "How Facial Recognition Is Fighting Child Sex Trafficking," *Wired*, June 19, 2019.

[19] U.S. Department of Immigration and Customs Enforcement, HSI, Partners Launch First US-Based International Victim Identification Surge, article, August 9, 2023.

[20] See Anuradha Nagaraj, "Indian Police Use Facial Recognition App to Reunite Families with Lost Children," *Reuters*, February 14, 2020; see also Swethavimala M, "Cops Rescue Kids from Child Labour Across Telangana," *New Indian Express*, January 11, 2024.

[21] Science and Technology Directorate, "S&T Tech Leads to Children Rescued and Traffickers Arrested," May 9, 2022.

[22] Thomas Brewster, Exclusive: DHS Used Clearview AI Facial Recognition in Thousands of Child Exploitation Cold Cases, *Forbes*, August 7, 2023.

[23] U.S. Government Accountability Office, "Facial Recognition Services: Federal Law Enforcement Agencies Should Take Actions to Implement Training, and Policies for Civil Liberties," report, September 5, 2023.

What’s going on? Why is the technology’s proven success in solving the toughest and most elusive crimes, directed at the most vulnerable of victims, overlooked? The story has two parts, one going to the core of data privacy, and the other to concerns with racial disparities.

“Unacceptable” Privacy Risk

> “The use of AI systems that create or expand facial recognition databases [… is] prohibited because that practice adds to the feeling of mass surveillance and can lead to gross violations of fundamental rights, including the right to privacy.”
>
> - EU Artificial Intelligence Act, Recital 43

It is not an exaggeration to say that data privacy law’s number one priority is to restrain the use of biometric technologies, which identify people via records of physical characteristics like facial image, fingerprint, or voice. These technologies evoke dystopic trepidation, summoning images of Big Brother and autocratic surveillance, ominously depicted in Sci-Fi films like *Minority Report*, *Terminator Judgment Day*, and *Demolition Man*. While the technology is no longer fiction—more than 7 out of 10 people use facial recognition to unlock their smartphones—its mythic legacy lingers, prompting data privacy law’s heavy artillery: limits on the collection and use of such biometric data. In this framework, the social upsides are not regarded as a factor that should shape the regulation.

The most commanding regulations come from the EU. Its 2024 Artificial Intelligence Act outright bans “real-time remote biometric identification” that relies on “untargeted scraping of facial images from the internet,” classifying the function of identifying people as “unacceptable risk.” The act categorizes other biometric recognition systems, like those that manage access into phones and spaces by one-to-one verification, as “high risk,” and slapping on strict restraints before market introduction and throughout their lifecycle.[24] In the U.S., federal law does not (yet) have a specific biometric privacy act, but it is headed in that direction. The proposed National Biometric Information Privacy Act regards this information as most sensitive and would subject it to heightened protections.[25] At the state level, Illinois’ highly litigated Biometric Information Privacy Act (BIPA) regulates the collection, storage, and use of biometric identifiers, including “scan[s] of hand or face geometry.”[26] In its distinctly non-European approach, BIPA stops short of outright prohibition, instead requiring robust consent rituals. However, through national class action litigation, its chilling effect is no small matter; many a company have been caught off guard by its heavy civil penalties.

Privacy law critically restricts the supply of images needed to train an algorithm to identify images. There are image-rich databases like Flickr, for example, which are readily downloadable and may be easily accessed and used by others. But not without the consent from each of the millions of people who posted their photos to Flickr. When IBM, Microsoft, and Google trained facial recognition algorithms using these resources, they were hit with privacy class actions under the Illinois BIPA—all, by the way, filed by the same plaintiff—for

[24] Artificial Intelligence Act, Article 5 (2024).
[25] National Biometric Information Privacy Act of 2020, S. 4400, 116th Cong., 2d Sess. (August 3, 2020).
[26] 40 ILCS 14/15 §§ 10, 15(b).

failing to secure proper consent, a harm for which lawsuit sought $5000 per image.[27] Think what you may about the merits of the consent requirement—a meaningful protection of one's right to self-determination or yadda-yadda boilerplate—it is a feature that obstructs the path to more accurate (and, we'll see, more equitable) facial recognition. Because of the privacy lawsuits, the tech giants simply abandoned that straightforward path to building the facial recognition software.

Privacy advocates celebrate this effect. If consent obstacles to training algorithms end up shutting down the technology, *tant mieux*—data privacy has done its job! Instead of helping the technology become more effective in law enforcement by feeding it more data and by focusing on the necessary safeguards, it is squelched. While their concern is presently centered around its use in policing, it stems from deeper anxiety evoked by other potential applications which, they say, would dispense with anonymity and "end privacy as we know it."

Why is facial recognition so threatening? Biometrics are biologically unique and permanent, and their exposure evokes a deeper sense of psychological impact, one of bodily penetration.[28] What if governments use it to identify protesters and dissidents? Or businesses refuse entry to folks who stick them with negative reviews or run afoul of their owners' politics? What if, in the same way that they can presently identify birds or flowers, people will be able to take photos of strangers, instantly know who they are, and use the information in invasive or offensive ways? These would gravely diminish our anonymity enjoyment in the public.

Indeed, the technology could be taken into dystopian directions. The prevalent use of surveillance cameras in countries where they serve to suppress a whole host of basic human and civil rights is spine-chilling. In China, facial recognition technology fuels a surveillance state that reinforces the totalitarian policies of the ruling Communist Party. It is a tyranny so salient that the growing use of facial recognition in U.S. law enforcement has led advocates to warn—without distinguishing the contrasting goals—that we are approaching "the day when America becomes more like China."[29] "America Under Watch," declares a report by an academic research center, describing the surveillance blanket in a Chinese city and warning that it is not "a remote, future concern for the United States" but rather "an imminent reality" for the "millions of people living in Detroit and Chicago."[30] The rhetorical hyperboles of dystopia and the Chinese analogy are used so indiscriminately that the ACLU is now accusing TSA security checks in airports for going in China's footsteps and "threatening a dystopian future." For what? For allowing people (who choose to opt in) to breeze through security check with nothing more

[27] *Vance v. International Business Machiines*, 2020 WL 5530134, at *2 (N.D. Ill., 2020); *see also Vance v. Microsoft Corp.*, 2022 WL 9983979 (W.D. Wash., 2022); *Vance v. Google LLC*, 2024 WL 1141007 (N.D. Cal., 2024).

[28] See, e.g., Sara H. Katsanis et al., "U.S. Adult Perspectives on Facial Images, DNA, and Other Biometrics," *IEEE Transactions on Technology and Society* 3, no. 1 (March 2022): 9–15, showing that users' comfort with biometric data capture decreases progressively from fingerprint, voice sample, hand geometry, and eyescan to DNA, revealing how discomfort is proportional to the physical depth of the feature measured—the extent to which data collection must penetrate..

[29] *See, e.g.*, Adam Schwartz, "Chicago's Video Surveillance Cameras: A Pervasive and Poorly Regulated Threat to Our Privacy," *Northwestern Journal of Technology and Intellectual Property* 11, no. 2 (2013): article no 4.

[30] Clare Garvie and Laura M. Moy, "America Under Watch," *Center on Privacy & Technology,* Georgetown University, 2019, at https://www.americaunderwatch.com/.

than an instantaneous facial scan (while smugly sizing up the curling procession of the biometric recalcitrant awaiting their ID checks).[31]

These apocalyptic portrayals trigger bans. Some cities have been quick prohibit their police from using facial recognition, even in the immediate aftermath of highly successful deployments of the technology in these localities. Here is the irony: New Jersey's Attorney General barred the use by state police of a facial recognition tool, while simultaneously celebrating in a press conference the arrest of nineteen child predators—an arrest that he knew was made possible by the technology.

As we already saw in the tracking-of-drivers context, data privacy cloaks its concerns in such transcendental gravity that even routine conveniences, such as skip-the-line airport entry, incite paramount concerns, and end up falling by the wayside. On this view, nothing outweighs the urgency of annihilating the risks of misidentification and predatory use. The daily expediencies seem petty and negligible in comparison to incidents of misidentification and the potential weaponization. As one critic put it, "[i]t's a technology that has such a profound potential to erode the very fabric of human society that any potential benefits are outweighed by the inevitable harms."[32]

Racial Discrimination

> "One of the most pressing threats to human rights and racial justice is the proliferation of racist facial recognition technology."
>
> - Amnesty International (2023)

Over the past decade, facial recognition technology has risen to the forefront of the algorithmic equity conversation, due to perceptions of racial disparities in error rates. Popular media has reported several troubling cases in which Black men were falsely arrested due to erroneous facial recognition matches. These reports have fueled some of the most prominent "algorithms-are-biased" complaints.

In one case, Nijeer Parks, a Black man from Woodbridge, New Jersey, was arrested when his state ID photo erroneously matched that of an offender's fake Tennessee ID, only to be released after 10 days in jail. In another case, Robert Williams, a Black man from Detroit, was traumatically arrested in front of his family and spent a night in jail when, again, his ID photo was matched by facial recognition software with the video image of a Black person committing theft of luxury watches. Parks' case triggered a scathing reaction from the ACLU, claiming that "this flawed and privacy-invading surveillance technology ... disproportionately harms the Black community."[33] And Williams' case was widely reported in the national press and became the subject of an entire chapter in *Your Face Belongs to Us*, a book by the New York Times privacy

[31] Jan Stanley, "The Government's Nightmare Vision for Face Recognition at Airports and Beyond," ACLU (February 6, 2020).

[32] Hill, *Your Face Belongs to Us*, 238.

[33] Kashmir Hill, "Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match," *New York Times*, January 6, 2021.

reporter Kashmir Hill, on the risks of facial recognition technology.[34] After recounting three cases in which police made such mistaken matches, Hill's verdict underscores the inequity aspect: "In every case, the man wrongfully arrested was Black."[35]

Given these issues, privacy laws restricting biometric collection are hailed as antidiscrimination efforts, sterilizing a tool that performs unevenly across racial groups. These concerns are grounded in the now-conventional dread that algorithms fueled by personal information are perpetuating and even magnifying historical biases against underprivileged groups in society. However, one only arrives to the conclusion that biometric tools are discriminatory by overlooking the sources for the disparate accuracy, and one can only justify the suspension of the technology by ignoring its significant equity benefits.

To fully understand how facial recognition serves this point, we must begin by challenging the widespread belief it is inherently biased against racial minorities. The truth? The technology performs this way, less effectively for non-Caucasians, only in its infancy and in large part when handcuffed by privacy rules. This might be hard to swallow, but a few basic questions let us address it piecemeal. Why do facial recognition algorithms seem to perform in ways seemingly discriminatory, and what causes the racial disparity in error rates?

Unlike risk assessment algorithms used in criminal law, facial recognition is not trained by data that may reflect discriminatory past practices. It is merely trained on images of faces. So when outputs "discriminate," something else must be going on—something far more covert, intervening as images of minority group members *enter* the training set. Turns out, the mystery phenomenon is a new-age version of the old-school "other-race effect," rehashed in the digital era and aggravated by data privacy restrictions.

Long before algorithms, humans worldwide have demonstrated a heightened proficiency in recognizing members of their own racial groups. This stronger ability is largely driven by greater interaction with members of one's own group. Greater familiarity enhances processing fluency.[36] Notably, the same is true for algorithms. If they are trained primarily on images of a majority race, they will make more errors when recognizing images of a minority race—and this apply not only to race but also age and gender. For instance, algorithms developed in Western countries are less accurate with respect to East Asian individuals compared to Caucasians. By contrast, algorithms trained in China have the reverse tendency, with comparatively lower identification errors for East Asian faces.[37]

AI ethicist Alice Xiang explains that limits on images available to train the algorithms affect their screening accuracy, with a disproportionate adverse impact on racial minorities.[38] In any society, there are fewer images of minority group members, resulting in lower accuracy,

[34] Hill, *Your Face Belongs to Us*, 169-184.
[35] Ibid., 183.
[36] *See, e.g.*, Jessica L. Yaros et al., A Memory Computational Basis for the Other-Race Effect, Sci. Rep., Dec. 18, 2019, at 1.
[37] P. Jonathon Phillips et al., "An Other-Race Effect for Face Recognition Algorithms," *ACM Transactions on Applied Perception (TAP)* 8, no. 2 (2011): article no. 14.
[38] Alice Xiang, "Being 'Seen' Versus 'Mis-Seen': Tension Between Privacy and Fairness in Computer Vision," *Harvard Journal of Law & Technology* 36, no. 1 (2022): 1–60, 16.

particularly in the early stages of the technology. Addressing this problem is complicated by privacy and antidiscrimination norms—of the data minimization genre we discussed earlier—which restrict the ability to tag images by race or ethnicity. These tags are important because algorithms are slower to recognize group membership of an individual in an image unless demographic markers are used.[39] As a result, we often lack data to train the technology for equal accuracy across races. In other words, privacy and data protection have the unintended consequence of making it harder for algorithms to avoid mistakes.

Notice that imprecision is not itself bias, but it leads to disparate outcomes. There will be more mistaken identifications among members of a minority group, as well as more mistaken failures to identify. But whether they are false positives or false negatives the errors disproportionately fall upon minority group members. False positives could lead to wrongful arrests, as seen in the cases mentioned above. They could also lead to compromised security by granting impostors access to biometrically protected accounts. False negatives, in turn, may deny individuals access to privileges like entry, electronic payments, or many touchless conveniences of digital recognition. Because these issues stem less from overt programming bias and more from constraints placed on the data and training process itself, what's needed is more, not less, personal information—access to a sufficiently large, diverse, universal set of training data.[40] But to even discern whether a training dataset is diverse, we need tags for demographic categories, especially for protected attributes.

Ultimately, privacy law's data minimization norm did not cause lasting harm. In the early days of the technology, it had the effect of diminishing the accuracy of facial recognition tools and therefore contributed to the appearance of racial bias. In the end, though, it only delayed the solution. As more images were used for training and stronger AI tools deployed, the disparate accuracy problem subsided. Revealingly, in a moment of candid lament a privacy activist conceded that "the advocacy community had 'led with its chin' by focusing so much attention on a fixable problem with the technology." Critics now admit that "the window of time for that criticism to be effective is closing as top developers have focused on addressing the problem of biased algorithms." Lest my ears deceive me, am I hearing them say "fixable problem"? Is it possible that advocates are *agonizing* that such fix arrived so fast, worried—as they are quoted to say—that the "greater accuracy across diverse groups" would be used "as a justification to deploy the technology more widely"?[41]

Facial Recognition's Sensible Future

Facial recognition technology is a tough nut for policymakers to crack. It triggers real privacy anxiety, and while I doubt that the dystopian applications are indeed "inevitable," I may be proven wrong. My main contention targets the dogmatic refusal to weigh the proven benefits when enacting precautionary rules. We need a sober assessment of the full spectrum of social benefits and harms. A data protection law that rushes to shut down the technology and squander its enormous law enforcement benefits is unacceptable. Sadly, it is the prevailing approach in many jurisdictions.

[39] Ibid., 9.

[40] Patrick Grother et al., "Face Recognition Vendor Test (FRVT) Part 3: Demographic Effects," report, U.S. Department of Commerce, 2019.

[41] Hill, *Your Face Belongs to Us*, 239.

The benefits I mentioned, specifically in the fight against human trafficking, are not unknown to privacy advocates and lawmakers. They are simply not a priority. Jurisdictions prohibiting facial recognition tools say that they are ready, in principle, to soften these bans when dealing with human trafficking and child sexual exploitation. Such exception has been carved, for example, from the EU's strict ban on facial recognition.[42] However, with so many safeguards, the principle rarely translates into practice and allowances are seldom made. For example, when the EU's Interpol launched an intense biometric enforcement program to combat human trafficking, the EU's data protection czar was not supportive. It perfunctorily acknowledged the enforcement effort's "important objective of general interest" before warning that "given the nature of the personal data at stake—sensitive biometric data—and that vulnerable people may be involved—migrants" more attention needs to be paid to the "impact on the fundamental rights to privacy and data protection" to make sure that such social sacrifice is "actually justified".[43]

What a revealing statement. To the data privacy credo, there is a clear hierarchy. At the top are the "fundamental rights" where privacy belongs. Then there are second-order values of "general interest" that, in case of conflict with privacy, must yield. Rescuing human trafficking victims, it appears, is a lower priority.

You've seen this ranking before. We watched as California's insurance regulator acknowledged the general interest in safe driving and reduced accidents but abandoned its pursuit saying it would threaten a more sacred mandate: to protect drivers' data privacy. This same hierarchy is also found in the media's coverage of the trade-off, which is overloaded on anecdotes of erroneous facial recognition and the potential violations of privacy rights. The New York Times repeatedly warns that facial recognition tools "might end privacy as we know it."[44] It rarely if ever mentions the human rights of thousands of trafficking victims, many of them minors, who have been or could be saved by the same technology. One time it did, but then quickly pivoted to privacy risks, stating, "the exchange of freedom and privacy for some early anecdotal evidence that it might help some people is wholly insufficient to trade away our civil liberties."[45] The Chutzpah to complain about "anecdotal evidence".

What would it mean for data protection law to shift from a privacy-centered ideology to a rational approach that strikes a balance between social costs and benefits? For one, the sweeping moratoria on the technology would give way to proportionate restraints. More severe crimes provide a stronger justification for deployment of the technology; conversely, not every

[42] Artificial Intelligence Act, art. 5; European Parliament, "Artificial Intelligence Act: MEPs Adopt Landmark Law," press release, March 13, 2024.

[43] European Data Protection Supervisor, *Opinion on the Proposal for a Regulation on Enhancing Police Cooperation in Relation to the Prevention, Detection and Investigation of Migrant Smuggling and Trafficking in Human Beings, and on Enhancing Europol's Support to Preventing and Combating Such Crimes and Amending Regulation (EU)*, 23 January 2024.

[44] See, e.g., Kashmir Hill, "The Secretive Company That Might End Privacy as We Know It," *New York Times*, January 18, 2020; *see also* Kashmir Hill, "Unmasking a Company That Wants to Unmask Us All," *New York Times*, October 28, 2021; see also Kashmir Hill, "Your Face Is Not Your Own," *New York Times*, March 18, 2021.

[45] See Kashmir Hill et al., "Clearview's Facial Recognition App Is Identifying Child Victims of Abuse," *New York Times*, February 7, 2020.

park littering incident should initiate widespread search of cellphone geolocation data or of bystanders' facial images.

Second, data protection law must recognize that being identified is often an advantage, not a threat. Great social benefits accumulate from the use of biometrics in authentication. They reduce fraud and are conveniently deployed in banking and even voting. These benefits—the billions of daily micro advantages—should not be frowned upon. Those whose features are less frequently recognized find themselves left behind: more often denied entry, rejected by the payment system, and suffering the indignation of being "treated as a second-class citizen, living in a world that cannot detect or recognize you."[46]

Third, when applied in law enforcement, facial recognition supplies *leads*, not dispositive proof. A photo-match, like witness identification lineups, should be the beginning of an investigation, not its end. Each of the troubling false arrests discussed above resulted not from proper use of the technology but from misapplications that could have been easily avoided had police followed standard investigative protocols. In fact, when used prudently, biometric identification can *cure* erroneous arrests, correcting the false positives so-often introduced by humans' identification errors.[47]

These safeguards are necessary components of the data technology, and they must be calibrated with an eye to the technology's benefits. Presently, our political advocacy is dedicated to constructing privacy-protective bureaucracies that restrict the technology, without properly considering the harms and the benefits. Anecdotes surrounding misapplications are treated as systematic data, whereas reliable reports on the impact in relation to human trafficking are relegated to footnotes. Our data protection regime cannot operate as if the rights to control how one's images are used—images that people thoughtlessly splash onto social media—are more significant, more essential to fundamental human rights, than the freedom and safety of sex-trafficked minors.

ELECTRONIC HEALTH RECORDS

> "Data privacy is a major challenge for the application of machine learning in health care because it restricts the potential for pooling together sensitive data such as the electronic health record (EHR) from multiple sites"
>
> — *Surat Rajendran et al., Cornell University*[48]

Who does not believe that medical privacy is vital, that it is crucial to personal safety and dignity, and central to the ethical delivery of health care? America and Europe have thick webs of privacy laws that shield people's medical information and health records—one of those rare agendas embraced by all political ideologies. And yet, it is a regime that lacks

[46] Xiang, "Being 'Seen' Versus 'Mis-Seen,'" 19.

[47] Kashmir Hill, "Clearview AI, Used by Police to Find Criminals, Is Now in Public Defenders' Hands," *New York Times*, September 18, 2022.

[48] Surat Rajendran et al., *Learning across diverse biomedical data modalities and cohorts: Challenges and opportunities for innovation*, 5(2) Patterns 1 (2024).

clear limiting principles. In both theory and practice, it has over the past generation ratcheted up restrictions on the use of personal health information. As new data science tools offer vast potential in medicine, privacy law responds with tougher constraints on the data these tools may use, often without pausing to consider the cost. Have medical privacy rules stretched so wide, to realms that interfere with the delivery of beneficial health outcomes or handcuff important biomedical research, all without yielding meaningful added protection?

Sadly, the answer is yes. Medical privacy protections have expanded such that they are interfering with effective health care, they are slowing down research, adding very little in terms of data privacy and individual control. To understand the distortion created by this regime, I will focus on a central—but by no means the only—battleground of medical information privacy: the restrictions on the use of electronic health records (EHRs). People's health records are essential for accurate diagnosis and effective treatment. Stored electronically, they are easier to preserve, access, share, combine, analyze, and audit. As data, they not only help identify patterns in the health and treatment of individuals, but within heterogeneous populations at-large. In expanding the eyeballs that may analyze them, they trigger data privacy and security alarm.

Medical Privacy Regulation

The law's primary safeguard for EHRs is the "minimum necessary" standard, which limits disclosure of personal health data across providers or for the purpose of medical research and requires the patient's consent for any such transfer. In the U.S., the HIPAA Privacy Rule (HIPAA is the cornerstone of federal medical privacy law) does this work.[49] The effect of the rule is a prohibition on medical providers to share health information through which patients may be identified, unless patients grant specific *informed and written* consent.

HIPAA governs medical staff, instructing them to not disclose patients' information to anyone absent consent from the patients and proper data management certifications by the recipients.[50] A different regulation governs scenarios when EHRs are sought for a purpose beyond further treatment, of biomedical research. This regulation, the Federal Policy for the Protection of Human Subjects (known as the "Common Rule"), applies and is enforced at the researchers' end, mandating that every proposed study get ethically pre-approved by institutional review boards (IRBs). It specifies that any retrospective study of preexisting EHRs obtain fresh consent from the data subjects, even if the study involves no patient interaction and only aims to analyze the data in their charts. It is not enough for subjects consent to the original recording of their data; a new specific consent must be procured for any further record examination. IRBs require such consent to be based on comprehensive disclosures given to each participant about the specific secondary research, including its purpose, procedures, risks, benefits, and data confidentiality.[51]

[49] 45 C.F.R. §§ 164.502(b), 164.514(d).

[50] 45 C.F.R. §§ 164.508(c)(2) and (a)(1); See also GDPR art. 9(2)(a), allowing consent to lift the general prohibition on the processing of health data; Additionally, the "broad consent" seemingly permitted by GDPR recital 33 cannot be "relied on for processing health data for 'any kind of—unspecified—future research." Instead, new consent is required for processing what doesn't foreseeably flow from the broad consent.

[51] 45 C.F.R. § 46.116.

The Privacy Rule and the Common Rule operate (often cumulatively) in two separate spheres: medical treatments and medical research. While their restrictions, and especially the privacy-centered informed consent, are widely supported by privacy advocates and bioethicists and implemented by a devoted medical privacy bureaucracy, they have a significant social cost. They have bred a culture of the non-disclosure of EHRs, where medical providers who treat patients "have become reluctant to share information for fear of over-disclosure, leading ironically, to under-disclosure."[52] And when patients' consents are too impractical to procure, medical research, medical progress, and treatment of these patients suffer.

Social scientists and epidemiologists who have studied the effects of EHRs see a very clear pattern that emerged already in the early days of digitized records, when paper files where being gradually transferred into computers. At that time, laws in some states explicitly intended to block EHR technology and did so successfully, so much that they had become "[t]he largest single predictor of [EHR] adoption."[53] Recognizing the trade-off between EHR deployment and privacy protection, one study calculated that health data privacy laws reduce the adoption of EHR technology by 24%.[54] In the present era of AI-driven precision medicine, data privacy is again a major impediment in building and sharing patient datasets that would yield medical breakthroughs.

Electronic Health Records in Clinical Care

What are the consequences of limited availability of EHRs in the ability to effectively treat patients? In an important paper, economists Miller and Tucker asked, "can health care information technology save babies?" and their answer was a ringing "yes."[55] Recognizing that the US lags well behind the EU in neonatal mortality, they searched for causes and found a surprising one in data privacy law. It turns out that a significant fraction of preventable neonatal mortality was attributed to the absence of digitized patient records—a deficiency that denied physicians critical instantaneous information about pregnancy complications. The magnitude of the effect was startling. A 10% increase in EHR adoption would have reduced neonatal mortality rates by an average of 16 to 26 (and possibly as many as 80) baby deaths per 100,000 live births. Stated plainly, "a complete national transition from paper to computer records could save as many as 6,400 infants per year."[56]

These benefits are particularly critical for high-risk pregnancies concentrated among underprivileged communities. For example, black babies are twice as likely to die in the first four weeks of their lives and suffer three times as many neonatal deaths from pregnancy complications of the kind that EHR-induced monitoring could address.[57] Because clinical

[52] "Strengthening Health Data Privacy for Americans: Addressing the Challenges of the Modern Era," Statement of Senator Bill Cassidy, Senate Committee on Health, Education, Labor, 2024.
[53] Amanda R. Miller and Catherine Tucker, "Can Health Care Information Technology Save Babies?" *Journal of Political Economy* 119, no. 2 (2011): 289–324, 309.
[54] Amanda R. Miller and Catherine Tucker, "Privacy Protection and Technology Diffusion: The Case of EMRs," *Management Science* 55, no. 7 (2009): 1077–1093.
[55] Miller and Tucker, "Can Health Care Information Technology Save Babies?," 310.
[56] Ibid., 319.
[57] Ibid., 315, 318.

uncertainty and physician discretion is a primary source for variation in treatments, improved EHR information would weaken these factors and have a particularly pronounced effect in treating minority populations.

Another illustration of the tension between EHRs and patient privacy—with equally fateful consequences—comes from the treatment of AIDS patients in Africa, where more than 25 million people suffer from the epidemic. When patients are diagnosed with the disease, having good health records that allow clinical staff to effectively follow up is critical to limiting new infections and saving lives. A study of EHR adoption in Malawi, titled "Privacy at What Cost?," demonstrated that switching patient records from paper to electronic had a large impact. EHRs allowed providers to monitor patients' compliance with treatment, track patients at risk, and reduce the incidence of lapses in care. Within five years, the use of EHRs led to an estimated 34% increase in patients receiving care and a 28% decrease in deaths. The effect was acute for children under 10 who are particularly vulnerable to lapses in treatment—their mortality decreased by 44%.[58] Over 5000 AIDS deaths were prevented in Malawi by the adoption of EHRs. The suppression of the AIDS epidemic in that country is in large part due to the EHR system.

It is important to recognize that, in this context, the privacy interests were of the utmost importance. If information about a patient's AIDS diagnosis were to circulate within their community, they would endure severe social stigma. Consequently, privacy law's requirement of patients' consent to the participation in the EHR system has greater urgency. Indeed, at the time of the Malawi study, patients possessed the privacy rights to withhold consent to EHR participation, and a substantial number of them—often those who would benefit most—did so. The problem with this privacy protection privilege, as the authors of the study state, is that "[f]or many patients, the cost of privacy is death."[59]

So here you have it, the privacy/health trade-off in its most acute manifestation. Unlike the more subtle privacy rights that we will soon discuss in the context of medical research, the EHR opt-out rights in Malawi protect against concrete privacy and dignity harms. Still, the law grants these powers to patients under social pressures hindering their prudent decision-making, and in a context where their decisions could inflict severe harm on others. In the name of data self-determination, parents may deny their young, AIDS-stricken children the participation in an EHR system—often the most promising path to survival.[60] By reducing EHR adoption, this approach forgoes essential care, weakening efforts to protect community members from the paths of AIDS contagion.

Yes, there are contexts in which privacy risks are heightened. For example, the disclosure of prohibited procedures (like abortions) to the government, genetic information to life insurers, or sexually transmitted diseases, demand significant data protections. The Malawi case is certainly one of those contexts. However, data protection law should respond with finesse—for example, by restricting some channels of redisclosure of information. The need

[58] Laura Derksen, Anita McGahan, and Leandro Pongeluppe, "Privacy at What Cost? Using Electronic Medical Records to Recover Lapsed Patients Into HIV Care," Mimeo, 2022.

[59] Ibid*.*, at 2.

[60] Ibid., at 4. ("honoring patient requests for privacy can significantly hamper the effectiveness of an EMR system and comes at the cost of disruptions in care and even deaths.")

to prioritize targeted protections for the gravest privacy concerns is critical to guarantee that people seek medical care and counseling.[61] However, heightened privacy needs in certain contexts should eliminate the privacy-health balancing altogether. Unfortunately, this often happens, resulting in a blanket regulatory preference to prioritize privacy over health.

Electronic Health Records in Biomedical Research

> "As currently implemented, the [HIPAA] Privacy Rule impedes important health research."
> - *Institute of Medicine*[62]
>
> "[L]egislative efforts to protect individual privacy have reduced the flow of health care data for research purposes and increased costs and delays, affecting the quality of analysis."
> - *Lane and Schur, National Science Foundation*[63]

In medical treatments, the EHR ship has sailed. Digital records are now everywhere, and the privacy/health tradeoff that was so tragically present in the early days of the data technology or in places where digitization is still undergoing has, in the US, largely diffused. But in medical research, the privacy-health trade-off not only remains, but has intensified.

Datasets containing comprehensive repositories of patients' health records store vast troves of information that, when analyzed in the aggregate, can uncover patterns within the heterogeneous patient population, yielding critical medical discoveries. The potential is now unlimited, with AI methods identifying new ways to diagnose patients, prescribe personalized treatments, improve health outcomes, and reduce health care costs. But they need data, and the large volumes of data contained in EHRs—data that were originally assembled for other purposes, such as for clinical care or for some specifically authorized research purpose—are a primary resource for this enterprise. The value of this research is further bolstered when medical records are co-mingled or combined with the endless medically relevant information humans release to every possible tracking device, as well as with non-health data sources of different modalities, different cohorts, and different categories of information—what is technically referred to as cross-cohort cross-category learning (c^4). This is the kind of information that, with cutting edge data science, yield new discoveries in medicine and public health at low cost.[64]

[61] Amalia R. Miller et al., "Privacy Protection, Personalized Medicine, and Genetic Testing," *Management Science* 64, no. 10 (2018): 4648–4668; Joseph Buckman et al., "Privacy Regulation and Barriers to Public Health," *Management Science* 69, no. 1 (2023): 342–350.

[62] Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research* (Washington, DC: National Academies Press, 2009).

[63] Julia Lane and Claudia Schur, "Balancing Access to Health Data and Privacy: A Review of the Issues and Approaches for the Future," *Health Services Research* 45, no. 5 (2010): 1456–1467.

[64] Surat Rajendran et al., "Learning across Diverse Biomedical Data Modalities and Cohorts: Challenges and Opportunities for Innovation," *Patterns* 5, no. 2 (2024): article no. 100913.

Research Impeded

As mentioned, data privacy law impedes this enterprise through two channels. Clinical facilities who hold the EHR charts are restricted by HIPAA from sharing them, absent specific patient consent.[65] And research facilities seeking these data must navigate through IRBs' interpretations of the Common Rule, which generally establish that any retrospective study of preexisting health records must be freshly consented to by the data subjects, and then again for any additional secondary use of the charts.[66]

These hurdles are difficult to surpass, and good studies are therefore delayed or abandoned. Obtaining the secondary consents is often impractical, and although the Common Rule permits in principle the waivers of consent or the procurement of a priori broad consent, IRBs rarely approve them. Even when granted, the waivers prove insufficient because clinicians may still not share EHRs with researchers, requiring conformity with HIPAA's extra-stringent privacy safeguards. In other words, "HIPAA is so often used as a smokescreen to preclude sharing of data."[67] As a result, much beneficial charts-based research, which pose no clinical risk, never gets off the ground.[68] Studies found that due to HIPAA's Privacy Rule, 77% of the study applications were abandoned and patient accrual (i.e., the size the studies) declined in some areas by almost three quarters.[69] In a national survey of clinical scientists only a quarter thought the HIPAA Privacy Rule has enhanced participants' privacy while it was widely perceived to have had a substantial negative effect on research, often adding uncertainty, cost, and delay.

Now, you are thinking, how could that be? What's the big deal in getting consents? Don't we all accept or sign mandated disclosures numerous times a day? Couldn't researchers simply take boilerplate consent forms, tweak a paragraph or two to describe the secondary use of the data, and have individuals sign them, just as they did in the initial consent to treatment? Sure, a bit more bureaucracy, but nothing to get all riled up about, right? Unfortunately, the barriers are far more than printing costs. First, boilerplate consents for unspecified future uses are not approved by IRBs because they are not considered "informed." Even if the consent forms are summarily approved, clinical staff in custody of the EHRs, fearing the wrath of medical privacy officers, are often reluctant to share them without specific and robust patients' authorizations. At the very least, this requirement for detailed explanations are required for each secondary use imposes significant time and financial costs on the approval process.

[65] 45 C.F.R. §§ 164.508(c)(2), (a)(1); GDPR art. 9(2)(a), allowing consent to lift the general prohibition on the processing of health data.

[66] 45 C.F.R. § 46.116.

[67] Institute of Medicine, *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary* 25 (Washington, DC: National Academies Press, 2010)**.**

[68] Sebastian Porsdam Mann et al., Facilitating the Ethical Use of Health Data for the Benefit of Society: Electronic Health Records, Consent and the Duty of Easy Rescue," Philosophical Transactions of the Royal Society A 374, no. 2083 (2016): article no. 20160130.

[69] Institute of Medicine, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, 2009; Michael S. Wolf and Charles L. Bennett, "Local Perspective of the Impact of the HIPAA Privacy Rule on Research"**,** *Cancer* 106, no. 2 (2006): 474–479.

A bigger difficulty is obtaining a response from the people. Entire cohorts of past patients must be reached. These new requests for additional consent are sent well after the original records were compiled, and some patients may have changed their contact information or passed away. If reached, these ex-patients are no longer in need of treatment and may not bother to respond. One study found that the consent rate to follow-up data analysis was 96% in the pre-HIPAA period, when it could be given orally, and only to 34% post-HIPAA when it had to be more formalized. It also noted that the post-HIPAA consenting patients were older and with lower mortality rates.[70]

Pause quickly to think about this last sentence, "older and with lower mortality rates." When consents to secondary research are harder to get and are given selectively, the resulting sample is likely to be non-representative of the population. It could be older or younger, healthier or sicker, disproportionally lacking in members of some race or ethnicity, or biased in a host of unpredictable ways.[71] For example, old studies that used retrospective data famously concluded that induced abortions were associated with a higher subsequent incidence of breast cancer. However, these studies had a fatal methodological problem: they needed consent from cancer patients to obtain their abortion histories, and many women declined to grant this consent. Surely, those who consented are not a random sample. Indeed, follow up "prospective" research—where abortion information was recorded before the diagnosis of cancer, demonstrated no such abortion/cancer association.[72]

It is hard to get ex-post consents, but the via dolorosa to reach the secondary chart reviews only gets testier. The difficulties multiply when researchers seek to compile and combine data from separate sources. Bear in mind: algorithmic methods of statistical analysis require data sorted along multiple variables. For findings to be robust, algorithms must be trained on different biomedical datasets consisting of different sample cohorts with different informational content and classifications. In some biomedical areas, especially in cancer research, data must be sought from multiple sites because the disease varies across patients and treatments. This requires larger sample sizes than any single clinical site could provide.

As a result, scientific hypotheses that could be tested within hours or days instead must be written into formal protocols and await their turn to appear in front of the IRB (and, when rejected, as often happens, return to the queue and reappear after a few months, and so on until the consent forms are satisfactory to the ethicists on the boards). Then, they hit the wall of patients' habits of non-responsiveness. An entire channel of medical discovery—one that is particularly ripe for AI methods—is disrupted. All, in the name of protecting patients against privacy harm that "in a vast majority of retrospective clinical studies . . . is highly unlikely."[73]

[70] David Armstrong et al., "Potential Impact of the HIPAA Privacy Rule on Data Collection in a Registry of Patients with Acute Coronary Syndrome," *Archives of Internal Medicine* 165, no. 10 (2005): 1125–1129.

[71] See, e.g., Yvonne de Man et al., "Opt-In and Opt-Out Consent Procedures for the Reuse of Routinely Recorded Health Data in Scientific Research and Their Consequences for Consent Rate and Consent Bias: Systematic Review," *Journal of Medical Internet Research* 25 (2023): article no. e42131.

[72] Collaborative Group on Hormonal Factors in Breast Cancer, "Breast Cancer and Abortion: Collaborative Reanalysis of Data From 53 Epidemiological Studies, Including 83,000 Women With Breast Cancer From 16 Countries," *Lancet* 363 (2004): 1007–1016.

[73] E. Stefánsson et al., "Are Ethics Rules Too Strict in Retrospective Clinical Studies?" *Acta Ophthalmologica* 86 (2008): 588–590.

Here, too, there is heightened impact on medical research relating to sub-populations like ethnic minorities or people with rare diseases, where to data are already limited and where medical discoveries under traditional research methods have lagged. EHRs have the potential to invigorate these areas. They enlarge the databases for analysis, provide relevant information from multiple sources and sites that enrich the findings, and shed light on factors responsible for disparate health outcomes.[74] But the research community must be permitted to acquire the data, and the charts themselves must overcome data minimization rules that impoverish the demographic classifications so needed to fuel this exploration. Presently, the difficulty to recruit minority participants leads to poor estimates how they will experience new treatments. These imprecise estimates increased healthcare expenditure by $1.2 trillion in 2003-2006.[75]

Futile Protections

Unlike rules relating to facial recognition or to tracking of drivers, medical privacy rules do not resort to absolute prohibitions. The main regulatory technique—the privacy-preserving panacea—is the informed consent rule. I, along with many others, have shown how it chills research, yet healthcare's privacy ethos rarely asks itself "what is the true value of informed consent rituals?" They are hailed as serving the autonomy of patients, and in cases of invasive and unpleasant clinical research, consent might truly alert subjects to risks. But when research is performed en masse on pre-existing records, this justification is hollow. The typical risks are "not greater in and of themselves than those ordinarily encountered in daily life" (this is the regulatory definition of "minimal risk"), and all that is left is an abstract notion of dignitary harm. Does informed consent to protect even this?

We discussed the futility of privacy consent rules in chapter 5, and I'll say more about the false hopes feeding them in the next chapter, but for now please bear in mind that *by design* the consent-to-research forms are long—very long—typically exceeding ten pages of fine print, written at a textual literacy level that many recipients cannot follow, and containing unfamiliar and complex information that novices cannot master.[76] So people don't read them, period. Is the signature at the bottom truly "informed"? Are people any more dignified for having the opportunity to sign a form that everyone knows they will not read? If one of the driving concerns is the hacking of EHRs, where medical data reach the wrong hands, surely informed consent does nothing to reduce that risk. And so, when asked, only a quarter of epidemiologists thought that HIPAA's Privacy Rule enhanced patients' data protection interests.[77]

[74] Kelly Devers et al., *The Feasibility of Using Electronic Health Records (EHRs) and Other Electronic Health Data for Research on Small Populations***,** 2013.

[75] Manuel A. Ma et al., "Minority Representation in Clinical Trials in the United States: Trends Over the Past 25 Years," *Mayo Clin. Proc. 96* no. 1 (2015): 264–266.

[76] Omri Ben-Shahar and Carl E. Schneider, *More Than You Wanted to Know: The Failure of Mandated Disclosure* (Princeton, NJ: Princeton University Press, 2014).

[77] Roberta B. Ness, "Influence of the HIPAA Privacy Rule on Health Research," *JAMA* 298, no. 18 (2007): 2164–2170.

Much of what is described here is well understood and often discussed in some public health circles. The tension between data privacy and research progress is a central theme, especially in the research community. It is well understood that in addition to HIPAA and the Common Rule's direct effects, the surrounding data privacy culture, the tendency of the research bureaucracy to err on the side of caution, the variability in application and misinterpretation of the privacy rules, all "may result in increased difficulty in gaining permission from individuals or organizations to release data. This may lead to fewer research studies or studies that are less scientifically robust."[78] Unfortunately, the centrality of the data privacy paradigm in the law of big data and the orthodoxy of data protection that it yields continue to disrupt a rational response to the privacy-health tradeoff.

Making Society Worse

I would've liked to end this discussion with a quantitative measure of the social cost from privacy restrictions on EHRs. Unfortunately, despite the certainty that this cost is significant, systematic estimates of counterfactuals—as in, what advances in health and medicine could have been made—are hard to come by. Perhaps this explains the endurance of the privacy restrictions and the largely useless consent rules. Regardless, while we can't estimate the benefits of research that never happened, we can look instead at research that, under the same burdens, was not blocked but merely delayed. This gives some idea of the cost paid by society for this regime's effect on biomedical research, in terms of statistical lives lost.

A striking illustration emerges from a case in which the IRB delayed a 1980's experiment in the use of blood thinners (thrombolytics) for heart attack patients. The treatment was ultimately shown to be highly efficacious, reducing vascular mortality among patients by 5.6%, and indeed it had thereafter become the standard of care. But recruitment of subjects to the experiment took much longer in the U.S. than in 15 other countries because of the Common Rule's far stricter consent requirements, where subjects had to sign a 1750-word consent form describing risks, side effects, and their rights as subjects. The slower enrollment caused an eight-month delay in the experiment and in the subsequent approval of the treatment. What did eight months cost us? With close to 20,000 patients ultimately eligible for the treatment each month, there are 160,000 people who could've received the treatment but for the delay. A 5.6% increased survival rate among that group means that close to 9000 people, whose lives could have been saved by the treatment, died.[79]

CONCLUSION

In this chapter, I assembled evidence that some of the most heavily regulated data technologies can produce enormous social value. I specifically focused on *social* rather than *private*, benefits of the technologies—how they advance public goods like road safety, the fight against brutal crimes, and medicine. We may easily note how breathtakingly large this social value is.

[78] Julia Lane and Claudia Schur., "Balancing Access to Health Data and Privacy: A Review of the Issues and Approaches for the Future," *Health Services Research* 45 (2010): 1456–1467, 1459.

[79] See Simon. N. Whitney and Carl. E. Schneider, "Viewpoint: A Method to Estimate the Cost in Lives of Ethics Board Review of Biomedical Research," *Journal of Internal Medicine* 269, no. 4 (2011): 396–402.

When such benefits of data technologies are willingly—even fervently—sacrificed, we have to ask "For What?" Who is this god demanding their surrendered? In writing this book, I have spoken with many privacy disciples about this evidence, about the benefits of tracking dangerous driving, dangerous criminals, or dangerous diseases. People are stunned by the evidence and especially by the sheer magnitude of lives saved by technologies they have grown accustomed to suspect. In response, their first reflex often the same: "surely these systems would harm racial minorities." They are then even more surprised when shown why the opposite is true.

But then I hit a wall. The aha moment I had been expecting does not arrive. The evidence of lives saved, the potential for greater equity, the sheer magnitude of benefits sacrificed—none of it changes privacy advocates' mind. Why are data critics so hard to seduce? Why such strong support for privacy laws that impose disproportionate costs relative to how little they further privacy? Why feat data?

The next chapter examines these questions. It explores core justifications for data privacy fundamentalism, finding them in other protective values and ethical commitments. We discover a conviction of moral absolutism too pure for pragmatic messiness, "without its being compromised in any way by the happenings of the world."[80]

[80] Ernest Becker, *The Denial of Death* (New York: The Free Press, 1973), 166.

Chapter 8

Origins of Data Anxiety

> "It has become appallingly obvious that our technology has exceeded our humanity."
>
> — Albert Einstein.

> "My life has been full of terrible misfortunes most of which never happened."
>
> — Michel de Montaigne

We have a puzzling phenomenon: here are technologies that offer immense social benefits, protecting people's lives, health, and freedom, and yet lawmakers charged with protecting people's lives, health, and freedom are hostile to them. With such clear benefits, how has the resistance persisted for so long? Why not embrace the opposite approach and mandate the technologies?

This chapter explores the privacy-first angst, finding its roots in various ethical and economic precommitments. These principles align to propell data privacy protection the exclusive status of a basic human right, overlooking many other social concerns. The chapter examines a variety of goals—equal protection, personal autonomy, the taming of corporate power, accuracy—which data privacy protection purports to serve.

The restrictive approach embodies an intensified version of the "precautionary principle," where we'd rather be safe than sorry when confronting intimidating novelty. Data technologies often involve pivotal transformations of existing private and social practices. They render obsolete ways of lives and routines that rely on human expertise, situational knowledge, and intuition. These technologies evolve rapidly, bluntly, and often autonomously, blending the artificial with the natural in ways not seen before. They establish new norms of surplus distribution, concentrating enormous and potentially unchecked power in the hands of the data-wealthy. And all of this happens at quick, accelerating, and (unless you're Ray Kurzweil) unpredictable pace. While the private benefits of these technologies are immediately visible, irresistible, and even addictive, their potential downsides are uncertain and may take time to manifest—but if they do, the consequences could be significant.

The cloud of uncertainty that accompanies these "subversive" breakthroughs is met with an alarmist instinct. The skeptics say: "some things *could* go catastrophically wrong with this new scheme, and although such disasters have not happened before nor are they likely to happen, we should put in place a political and bureaucratic order to safeguard against the potential upheaval, and in the meantime slow down the introduction of the technology, no matter how flawed the status quo and how big the forgone benefit."

So seductive is this precautionary instinct—so often does it seem to be a good approach to the uncertainty brought upon by an unfamiliar technology—that many of its advocates do not pause to ask, "At what cost"? And those who do ask don't stick around for true answers, assuming—often without analysis—that the sacrifice is worth making.

As discussed earlier, this ultra-precautionary approach is often anchored in concrete dystopian illustrations, with the most disturbing being the Chinese surveillance state. There, data collected from street cameras, internet activity, cellphone locations, biometric databases, and countless other sources are harnessed through algorithmic systems to suppress political dissent, target minority groups, and enforce strict loyalty to the Chinese Communist Party's objectives. The chilling reality of such oppression stands in stark reminder of how data-driven surveillance can spiral into authoritarian misuse, fueling worst-case thinking. We are thus urged to treat the *improbable* like *imminent*.

This is obviously backwards. For technologies like usage-based insurance, facial recognition in law enforcement, and electronic health records, experience proves the benefits undeniably *imminent*, while the apocalypse purely anecdotal. The technologies matured, their benefits multiplied, and nuanced harm-prevention measures are available. In these areas, one would expect the flame of precautionary principle to dim, to succumb to a rational scheme that embraces the benefits of the technology with more narrowly tailored safeguards. But such adaptation never arrives. Instead, the data privacy restrictions are heightened, the accompanying rhetoric is emboldened, and the public grows more anxious.

The next few pages explore the foundations of this dogged resistance. I begin with a short reference to what I call *privacy fundamentalism*—the view that data privacy *is* the ultimate value, that it need not be justified as instrumental for more basic goals. I then pursue other justifications commonly offered by privacy advocates, including anti-domination, anti-discrimination, promotion of autonomy, the limits of informed consent, and error-reduction.

A. Data Privacy Fundamentalism

> "Privacy is a foundational good."
> — Anita Allen (2011)
>
> "Privacy is a means, not an end."
> — Lior Strahilevitz (2006)

At its genesis, privacy was thought to be a means, not an end. A means to good things, like the cultivation of one's personality, relationships, and health. And sometimes a means to bad things, as Richard Posner famously pointed out, warning that privacy could foster deceptive withholding or concealment of information, such as "information concerning […] moral conduct at variance with a person's professed moral standards."[81] Posner's skepticism never won many followers, and present-day discourse views privacy as an

[81] Richard A. Posner, "The Right of Privacy," *Georgia Law Review* 12 (1978): 393–422, 399.

essential fountain of good. The list of goods promised by data is ambitious and includes, "personality development and moral autonomy, personal honor, dignity, identity, creativity, and innovation; psychological well-being, intimacy, and family; civic association, religious expression, and ideals of a limited, tolerant government"[82] as well as the values of love, friendship, and trust.[83]

When we called privacy a means, we intended its protection to serve other, more fundamental values, insofar as these underlying benefits are not outweighed by countervailing costs. On this view, data privacy is desirable so long as it advances human flourishing at both the individual and community levels. Its protection must be carefully balanced against tradeoffs, especially when the benefits to some come at the expense to others.

At some point, we saw earlier in the book, when data privacy became a central organizing paradigm of lawmaking in the digital era, the intellectual origins of privacy-as-a-means faded. Detached from its original raison d'être, having joined the ranks of fundamental human rights, data privacy took stage as an end in-itself. It no longer mattered that people might live in communities with varying norms of data privacy, exhibit diverse preferences towards to data privacy rights, or that the protection of those rights could backfire and diminish other basic goods. Data privacy fundamentalism is a normative, not a pragmatic-descriptive account. Its idealism is calibrated to lofty values of personhood and human agency, and it therefore breeds non-waivable rights, invariant to diminished benefits or unintended harms.

If I've told the story of data privacy fundamentalism correctly (as I tried earlier, in chapter 2), two developments show that our thinking about data protection has lost its way. First, privacy commentary has not succeeded in convincingly arguing how exactly the "surveillance monster" is harming people. It is increasingly sufficient to say that people shouldn't have big tech looking over their shoulders – "that's Big Brother, that's wrong." Any data technology that collects personal data is, by definition, said to raise privacy concerns, especially when sensitive attributes are processed.

Second, the data privacy conversation no longer requires participants to examine the extent, if at all, that an objected-to "surveillance" technology impedes human flourishing. Nor it is warranted to double check whether any such diminution is outweighed by advancement of central hallmarks of private lives. Curiously, the literature devoted to the "problem of privacy" pays strikingly little attention to the benefits of data technology.

Often, such privacy-per-se concerns emerge from the core of constitutional privacy law. The dilemma arises, for example, when driving location data are sought by police or subpoenaed by parties in civil proceedings. Privacy advocates view the tracking technology as the culprit, cautioning that you are bringing "the spy along for the ride to "build a compelling case against you".[84] While I don't quite share the anguish that "tracking data have convicted

[82] Anita L. Allen, *Unpopular Privacy: What Must We Hide* (Oxford: Oxford University Press, 2011), xii, 19.

[83] Charles Fried, *An Anatomy of Values: Problems of Personal and Social Choice* (Cambridge, MA: Harvard University Press, 1970), 142.

[84] See, e.g., Ed Leefeldt and Amy Danise, "The Witness Against You: Your Car," *Forbes,* 2021.

murderers, hit-and-run drivers and thieves of their crimes"—what's wrong with that?—I do understand the anxiety behind it. However, this specific anxiety, I leave alone; there's just so much to say about that major battleground surrounding government search and seizure powers, and my goal here is more general.

Still, the constitutional battle surrounding the Fourth Amendment data privacy limits is a useful illustration of a balanced approach to data protection. It reflects a conviction that law enforcement benefits could sometimes outweigh privacy costs—that, for example, when offending in public places, offenders do not have expectations of privacy worth legal protection[85] Bypassing the constitutional balance by prohibiting the technology that enables the initial collection of such data would result in overprotection. A nuanced approach that carves data privacy safe harbors—for instance, by shielding the data from divorce proceedings or employment disputes—seems most desirable.

The remainder of this chapter explores various theoretical attempts to anchor data privacy rights to normative foundations that are specific, compelling, and susceptible to empirical evaluation.

B. Domination

Information does not exist in a social vacuum. Sociologically-alert writers are richly portraying how the accumulation of personal data and their computational renditions in the hands of already strong entities—platforms, financial institutions, sellers, employers—redefine market interactions, redistribute welfare, and disrupt traditional channels of self-advancement.[86] Personal information enables businesses to exert control over the "data subjects" (what a revealing term!), allowing them to micromanage and even manipulate people's behavior, while extracting commercial value from their databases. These businesses put their customers in what some have analogized to a "Sophie's choice"—between their privacy and their ability to obtain affordable services.[87]

For example, the commentary is filled with conjecture about the power imbalance between auto insurers and the drivers they track. In her brilliantly researched book *Data Driven*, Karen Levy describes the disempowerment professional drivers experience when tracking devices are installed in their freight trucks.[88] Unlike in personal auto insurance, the law mandates this technology to ensure compliance with traffic safety rules. At the most abstract level, power comes from hierarchy, and this mandate reinforces that hierarchy by imposing a constant reminder: an electronic eye monitoring the cabin 24/7. This surveillance undermines drivers' sense of "captainship" of the vehicles. Just as yesteryear's ship officers asserted their rank by keeping ordinary seamen out of their private quarters, today's truckers are subordinated by the surveillance of their intimate work environment.

[85] *Smith v. Maryland*, 442 U.S. 735 (1979).

[86] See, generally, Jannis Kallinikos, *The Consequences of Information: Institutional Implications of Technological Change* (Cheltenham, UK: Edward Elgar, 2007).

[87] Karl Bode, "Consumer Groups Slam Comcast's Plan to Charge Users for Privacy," *DSL Reports,* August 5, 2016.

[88] Karen Levy, *Data Driven: Truckers, Technology, and the New Workplace Surveillance* (Princeton, NJ: Princeton University Press, 2022).

Truckers, to put it mildly, do not like these data collection devices. For them, the micro-actions that were traditionally "self-contained and immune from immediate oversight" in a manner that retained "a degree of autonomy unmatched in other blue-collar jobs" are now, in the era of "organizational surveillance," measurable and quantifiable. The electronic monitoring creates new pathways of domination and control over daily practices, bolstering "the entrenchment of power in modern organizations."[89]

What are these channels of control? Seeing what people buy, how they drive, who they meet, or what entertainment they prefer enables sellers, insurers, media companies, and platforms to induce certain choices. Call these choices free, but they can be cleverly framed by the design of a menu, the layering of information, or the persistent reminders. So, who has the power here? The gullible client or the tactical business?

Critics assert that the data collected by auto insurers, for instance, put policyholders "under the domination of insurance company algorithms, either because they are not sure about the consequences of their travel behavior for future premiums, or they cannot control them. In the paradigm of usage-based policies, any small event (e.g. friend visit, mood change in the case of driving assist, weather change) comes with a possibility of a change of premium."[90] Insurers have the data and determine the resulting premiums; policyholders are unable to decipher the algorithm and remain in the dark. The knowledge gap leaves policyholders powerless.

Similar domination of consumers is seen in how data's economic value is apportioned. The digital economy is a joint production enterprise: individuals provide data for which platforms devise implementations, and their cooperation creates value and profits. People get some payoff, typically in free access to content, but this is a crumb of the profit pie at best. The rest of the value generated by the data, all but the crumb, is appropriated by the platforms. To address this inequity, data privacy law often restricts the flow of data from individuals to businesses. Now nobody profits! Recognizing this lost value, Eric Posner and Glen Weil proposed a different solution—pay for the data. They analogized data to labor, for which people ought to be compensated, and more so if they would be able to form "data labor unions" to bargain for a larger share of the pie.[91]

Power imbalance is also due to the non-portability of data. Businesses create data files on their customers, and can sell or combine them, but individuals are often unable to take their data files and move from one business to another. They are locked in. The question whether personal data are private property—on which rivers of ink have been spilled in law reviews—has not crystalized in American law to include general portability rights. A driver insured by Geico cannot take the digital record containing their tracked driving history and move to Travelers. If they switch, they must start over and build a new record to qualify for

[89] Ibid.

[90] Martim Brandão, "Discrimination Issues in Usage-Based Insurance for Traditional and Autonomous Vehicles," in *Culturally Sustainable Social Robotics—Proceedings of Robophilosophy*, 395–406 (Amsterdam: IOS Press, 2020).

[91] Eric A. Posner and Glen Weil, *Radical Markets: Uprooting Capitalism and Democracy for a Just Society* (Princeton, NJ: Princeton University Press, 2018).

the safety score discounts. When people cannot shop around for discounts or for better service, their bargaining power diminishes. They are captives to their present provider, who keeps small these people' share of the data enterprise's profits.

An additional power imbalance manifests in the one-sided uses of the data, especially when there are conflicting interests. Let's say that there is a dispute between an insurer and a policyholder as to whether the accident was caused by an event that is excluded from coverage. The insurer could use the driving data to establish proof in support of its position, but not vice versa; when the data vindicate the position of the policyholder, the insurer is unlikely to volunteer such evidence. Similarly, facial recognition software helps police prosecute a suspect, but how often does it help arrested suspects prove their innocence?[92] Sure, defense attorneys could gain access to image databases, but it is costly and at least at present it remains disproportionally in service of the police. Not just facial images, any data that are relevant for a criminal case. In building the case against a defendant, the prosecution can readily sift through data on recorded communications and actions; defense attorneys do not have such legal access.[93]

The information asymmetry allows sellers to exercise another form of domination—to identify cognitive biases and weaknesses among their clientele, to shape preferences and priorities, personalize the offerings, and engage in a multitude of legally questionable tactics which ultimately enable them to charge higher prices. The pipelines businesses build into which people's personal data are fed, spew significant profits on the output end.

These are various data-driven dominations often cited to justify stronger data privacy rights. While there is truth in many of these power imbalance depictions, is data privacy the appropriate intervention? In two critical ways, this prescriptive conclusion falls short. First, an absence of data privacy restrictions is not the actual cause to any of the channels of power imbalance. Second, these depictions of domination reveal only half of the power story; the other half lies in an account of meaningful enablement.

The harm that society confronts in dealing with domination-though-data is not to privacy, and often is not even a private injury. When data create lock-in effects that increase prices or reduce quality, bestow market power and entrench incumbents, or shape market demand in ways that favor sellers, the problem is fundamentally social. The fear is that overall consumer welfare and innovation would decline. True, many consumers could then be privately harmed if, for example, they face higher personalized prices; but others may gain with lower prices. The popular press thinks the overall effect is bad for consumers, but economists disagree. Studies have found that data-fueled price personalization could be a win-win for consumers and businesses and that it mostly benefits poorer consumers.[94] More generally, the accumulation of market power through the collection of consumer data

[92] It sometimes does. e.g., Kashmir Hill, "Clearview AI, Used by Police to Find Criminals, Is Now in Public Defenders' Hands," *New York Times*, September 18, 2022

[93] Rebecca Wexler, "Privacy Asymmetries: Access to Data in Criminal Defense Investigations," *UCLA Law Review* 68 (2021): 212–287.

[94] Jean-Pierre Dubé et al., "The Intended and Unintended Consequences of Privacy Regulation for Consumer Marketing: A Marketing Science Institute Report," Marketing Science Institute, 2024.

reshapes the structure of the marketplace. For purpose of designing solutions, the resulting harms must be conceptualized as occurring to *environments*, not individuals.

Consider again the lock-in problem. Could privacy law solve it? Lock-ins take different forms, but they all originate from high switching costs. Sometimes these are the termination fees agreed upon in long-term contractual commitments, but these are now less common (consumers have learned to notice them).[95] Frequently, switching costs flow from rich data businesses have on users, allowing them to personalize the service—a benefit users are reluctant to squander by switching. The generic solution to this problem is in the form of portability rights that allow your data to travel with you. Portability successfully addressed lock-ins in cellphone and pensions markets: when people switch carriers or employers, they have the right to take their devices or their pensions with them. The same could apply to data. If people could easily switch providers by bringing their data files along, a major source of power imbalances would be competed away.

Privacy rights are the wrong approach, and not only because advocates argue that porting data could expose other people's privacy.[96] Privacy rights could undermine, rather than advance, data portability. The most comprehensive portability right could be achieved by data banks. Think about auto insurance. To make it easiest for policyholders to switch carriers, why not create a data intermediary for insurance, resembling credit bureaus that aggregate personal financial data and make them available to any authorized financial institution. A centralized repository would allow any auto insurer authorized by their customer to receive the customer's driving history at the level of granularity held by the insurer who collected these data. This would greatly shift power to individuals, especially those who drive safely, letting them capture an important benefit of their safety scores—low premiums. Of course, such intermediation bureaus do not officially exist and would likely require regulatory mandates. Now ask yourself: would advocates of data privacy support such a data bank, the ultimate concourse for data portability? Surely not. They would view such practice as the definition of Big-Brother-style institutionalized surveillance. Such solution—an industry-wide regulation, would naturally emerge from the data pollution framework.

The second problem of the domination-through-data thesis is the account of domination itself. Perhaps a narrative of empowerment, rather than subjugation, better suits the data story. It certainly captures the lived experience of people more accurately. Picture the driving data context, imagining that Consumer Reports offer a free drive-tracking app that provides daily safety scores. People who choose to enroll would be teaching themselves to drive better. They would cause less accidents and better to monitor their teen drivers. Would you say that such an app, like other "free" digital services, like most "free" apps, exerts dominion? Would this differ from insurance usage-based pricing? In both cases people will drive better, but since insurers also vary the premium based on the safety

[95] Oren Bar-Gill and Omri Ben-Shahar, "Exit from Contract," *Journal of Legal Analysis* 6 (2014): 151–18.

[96] Daniel J. Solove, "The Limitations of Privacy Rights," *Notre Dame Law Review* 98 (2023): 975, 1006–1007; Peter Swire & Yianni Lagos, "Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique," *Maryland Law Review* 72 (2013): 335, 373–74.

scores, people will enjoy their newfound ability to influence their premiums and benefit from discounts, as the vast majority of UBI participants do.[97]

Like many data technologies, tracking drivers has enormous social value accruing to all participants. It is introduced in a sector—*insurance*—where firms are already in the business of knowing people's ills and mishaps; where risk, loss, and misfortune are the "product;" and where traditional rating practices give firms information about people's income, family status, education, health, and much more. Insurers already know *who we are* in a meaningful, invasive way. Now, with tracking devices, they are gradually replacing all that with a different set of less sensitive data—*how we drive*. And, for good measure, this program demystifies the insurance transaction: policyholders can understand their risk ratings and scores, review the factors that explain why their premiums change, and, how should I say it, . . . die less often.

Privacy advocates choose to ignore the happy side of so many data technologies. Lost? GPS tracks *location*. Clueless? ChatGPT aids *education*. Weary? Let Gemini plan your *vacation*. Does any of this sound like *domination*? True, people may not see the profits generated from data, but they've somehow managed to place more faith in Google and Amazon than anyone else, except, maybe, their doctors.[98] By sharing their personal information, have they, perhaps, experienced *liberation*?

C. Discrimination

> "Algorithms can entrench troubling bias against women and minorities if the algorithms themselves encode such bias."
> — Danielle Citron (2022)[99]

The most defining feature of products driven by big data is *personalization*, which means that people are treated differently, based on what their information shows. There are good reasons to be alarmed about this. Disparate treatments could be unjust, especially where it has the effect of systematically disfavoring protected minority groups. Biased baked into the data or embedded by their collection processes could inadvertently infect algorithms. An irresistible response to this alarm is to advocate for greater data privacy: shut-off the information before it can be used to classify and discriminate.[100] This concern with discrimination, we saw, was central to the privacy resistance towards facial recognition technology. Performance was uneven across racial groups with identification errors concentrated among Black men.

Similar concerns underlie resistance to algorithmic criminal justice predictions. For instance, machine learning tools that predict crimes are deployed in sentencing and bail decisions to

[97] M. Soleymanian, C. B. Weinberg et al., "Sensor Data and Behavioral Tracking: Does Usage-Based Auto Insurance Benefit Drivers?" *Marketing Science* 38 (2019): 21–43, 22.
[98] Nicole Lyn Pesce, "Americans Trust Amazon and Google More Than the Police or the Government," *MarketWatch,* January 18, 2020.
[99] DANIELLE KEATS CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE 21 (2022).
[100] See, e.g., Jessica L. Roberts, "Protecting Privacy to Prevent Discrimination," *William & Mary Law Review* 56, no. 6 (2015): 2097–2174, 2112.

assess defendants' propensities to reoffend. In policing, predictive models help allocate enforcement resources by identifying crime hot spots. Despite the growing recognition that these predictive technologies are typically far more accurate than humans performing the same tasks, there is a nagging anxiety that the tools disfavor racial minorities.[101] There are both principled reasons (biases in the data) and practical ones (nontransparent algorithms) to worry about these tools, and (as we saw in chapter 3) they have catalyzed support for a central norm in data privacy law—data minimization—which restricts the collection and use of race and gender data.

The insurance area offers a surprisingly interesting environment to explore these intuitions. Insurance is where firms are in the business of classifying people by every measurable trait correlated with risk. If women live longer than men, then all else equal their life insurance is cheaper and their pension annuities more expensive. While Federal law prohibits such gender classification,[102] most insurance sectors are regulated by state laws that permit, and even encourage, actuarially relevant gender classification.[103] This is why, famously, young men pay more for auto insurance.

A big worry, however, is disparate impact along racial lines. Risk classification that relies on crude factors like credits scores or education, on which low-income people and members of racial minorities score less favorably, would yield racial disparities. With more personal data, wouldn't things get worse? Yes, critics say, "AI and big data are game changers when it comes to this risk of unintentional, but "rational," proxy discrimination."[104] And the Federal Insurance Office at the Department of Treasury concurs, warning that "big data methodologies may hide intentional or unintentional discrimination against protected classes" pointing specifically to usage of personal data in insurance.[105]

Let's pause for moment. While there is much evidence that demographic ratings in insurance hurt member of racial minorities, would the same effect occur when ratings are based on minute-by-minute tracking of drivers? Here, there is no supporting evidence, which invites speculation. One theory centers on heightened danger of night-time driving, more common among lower-income people who work night shifts. Sounds plausible. Another focuses on location-based ratings, which raise premiums for cars driven in accident prone neighborhoods, areas more likely to be poor. This, too, might be true.

[101] This view appears in a huge literature. See, e.g., Cathy O'Neil, *Weapons of Math Destruction* (New York: Crown Publishers, 2016), 84–90; Julia Angwin et al., "Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks," *ProPublica,* May 23, 2016.

[102] *City of Los Angeles, Department of Water and Power v. Manhart,* 435 U.S. 702, 716 (1978).

[103] *Telles v. Commissioner of Insurance*, 574 N.E.2d 358, 361 (Mass. 1991) (quoting *Life Insurance Association of Massachusetts v. Commissioner of Insurance*, 530 N.E.2d 168, 171 (Mass. 1988)); *Insurance Services Office v. Commissioner of Insurance*, 381 So. 2d 515, 517 (La. Ct. App. 1979).

[104] Anya Prince and Daniel Schwarcz, "Proxy Discrimination in the Age of Artificial Intelligence and Big Data," *Iowa Law Review* 105 (2020): 1257–1318; see also Ronen Avraham et al., "Understanding Insurance Anti-Discrimination Laws," *Southern California Law Review* 87, no. 2 (2014): 195–274.
Daniel Schwarcz 2020: 1257. See also Avraham, Logue, and Schwartz

[105] Douglas Heller et al., *Watch Where You're Going: What's Needed to Make Auto Insurance Telematics Work for Consumers*, May 26, 2021.

However, the speculation quickly overreaches, declaring from both theories that driver tracking is inherently discriminatory—that it's "similar to redlining practices."[106] It's "merely another data mining exercise" that "penalizes" people "because of where and when they drive as a function of work and housing segregation."[107] And suddenly, data privacy is invoked to prevent this discrimination.

There is every reason to think that the opposite is true—that tracking technology *reduces* racial disparities. We live in an era of insurance risk classification greatly and bluntly *disfavoring* low-income drivers, where discounts are dispensed to those with big homes, big incomes, and big resumés. But now we have a technology that refuses to give these factors preference. Being poor is no longer a proxy for risk. Insurers need no longer play census bureau—their algorithms assess behavior alone, blind to demographic markers. The moto is "price me by how I drive, not by who you think I am."[108] As a result, poorer people score comparatively well on many of the driving inputs that are tracked, including the weightiest ones (e.g., miles driven).

Also, factors on which poorer people score worse, like driving location (which prompts the "insurance redlining" conjecture), are given lesser weight. What matters now is not where one lives and parks their car overnight, but rather where the car travels throughout the day. So, while socioeconomics may split the rich and poor into their respective neighborhoods, school, and clubs, they mostly drive the same roads. Driver tracking evaluates individuals across these groups by the desegregated portions of their lives. Its rating features are "accessible" to all, affording policyholders equal opportunity to enjoy the additional benefit of learning to drive better and lower accident risk. When an algorithm reduces the safety score of a driver who accelerates over 85mph, there is no systemic error, prejudice, or historical bias.

In many areas, data privacy advocacy points to bias and discrimination as reasons to minimize and even prohibit data collection, but as in the insurance context it is often a straw man. To appreciate the mismatch between data privacy and bias reduction, consider facial recognition technology, whose discriminatory effect has been continually critiqued. Indeed, early version of the technology lead to cases of mistaken arrest, the victims of these mistakes where all black men, and cities, in response, imposed moratoria on its use. However, even back then, the incidence of disparate racial impact was not enough to conclude there being a deep-seated bias within the technology, especially when compared with the alternative—witness identification.

This aside, even diehard privacy advocates reluctantly acknowledge that, by now, the bias problem is solved. Kashmir Hill—reporting frequently in the New York Times on this "racist" technology (her words) and who wrote a book intensely devoted to its discriminatory effect—concedes that "the window of time for that criticism to be effective is closing as top developers have focused on addressing the problem of biased algorithms."[109] But rather

[106] Brandão, "Discrimination Issues in Usage-Based Insurance," 395.

[107] Karapiperis et al., "Usage-Based Insurance and Vehicle Telematics," 52.

[108] Consumer Reports, "Price Me by How I Drive, Not by Who You Think I Am!: Tell State Insurance Commissioners to Make Pricing Fair," accessed November 1, 2024.

[109] Hill, *Your Face Belongs to Us*, 239.

than celebrate progress, the author concedes that it was an advocacy mistake to focus "so much attention on a fixable problem" of discrimination, and quotes a privacy activist who laments the absence of the disparate effect, as this "provided facial recognition purveyors the opportunity to use greater accuracy across diverse groups as a justification to deploy the technology more widely."[110]

People, not algorithms, are biased. Yes, algorithms trained on data produced by people replicate decades of human bias, segregation, and discrimination. However, we can use data analysis to uncover this bias. And because algorithms are essentially complex equations, any bias uncovered can be removed once traced back to specific variables. For example, when an auto insurer realized that some of the driving factors it tracks reflect features that drivers cannot control, which are more likely to import racial or other biases into the safety scores, it designed the algorithm to shrink the weight of these uncontrollable factors.[111] Is there a comparable technique for removing bias from human minds?[112]

A striking case in point is Amazon's infamous AI recruiting algorithm from a decade ago, infamous for being biased against women. It remains cited as a canonical illustration of data technology's biases.[113] These references rarely note, though, that this algorithm was trained on past decisions of human hiring officers and instructed to mimic them. It was predicting what a human recruiter, looking at a resumé, would decide. A human bias that was invisible without AI had come to light—so clearly that Amazon quickly (and well before the media outcry) discarded this algorithm. Not much to celebrate in this Pyrrhic victory against AI algorithms; if the human discretion guiding the algorithm was biased, why revert to this discretion? We only make it harder on ourselves to eliminate bias, which is far more easily detected in software code than in human minds.[114]

Earlier in the book, I argued that problems of discrimination which emerge from data technologies are indeed central to the "why fear data" question. I claimed that discrimination is a very different problem than data privacy, as it afflicts society at large and requires different tools for intervention. My thesis was that regulation should approach issues like discrimination as pollutants. Now, however, I've demonstrated that anti-discrimination efforts and data privacy law aren't merely misaligned; they may be opposed. This means that data privacy laws could backfire as a means of bias mitigation. I am certainly not the first to recognize this conflict.[115] As a strategy for bias-mitigation, data privacy laws backfire. Data minimization complicates the anti-discrimination training of prediction

[110] Ibid.

[111] For example, the loss function can be estimated in two steps. First, only controllable variables are considered to establish their weight. Then, noncontrollable variables are added, resulting in artificially lower weight for the latter. Alternatively, insurers can use a model that minimizes not only the squared errors but also a loss function accounting for the size of the noncontrollables' coefficients. Both techniques effectively reduce the weight of noncontrollable demographic factors, such as nighttime driving, traffic density, and location.

[112] Joshua A. Kroll et al., "Accountable Algorithms," *University of Pennsylvania Law Review* 165 (2017): 633–705, 634; Kleinberg et al., "Discrimination in the Age of Algorithms," *Journal of Legal Analysis* 10 (2018): 113, 152–158.

[113] Jeffery Dastin, "Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," *Reuters*, October 10, 2018.

[114] See, e.g., Bennie Mols, "Keeping Algorithms Fair," *Communications of the ACM*, June 6, 2019.

[115] Xiang, "Being 'Seen' Versus 'Mis-Seen,'" 9.

algorithms at both the training and auditing stages. It actually poses no obstacle to bias—algorithms pick out proxy factors for race when these data are withheld. Instead, train algorithm with race data to neutralized race proxies, and later, at the screening phase, omit race entirely.[116] A complete removal of race labels would make it practically difficult to measure biases and to ensure algorithmic fairness.[117]

D. Error Reduction

Another argument that percolates beneath the data privacy philosophy focuses on systemic errors in the data and its analysis, and the resulting inaccuracy in how algorithms represent and treat people. The position goes something like this. The profiles that emerge from people's digital footprints represent mere slivers of their identities, constructing so-called "data doubles"—reduced forms which "flatten and distort" real humans. Through this, an individual becomes a vector of quantitative parameters, a "context-free numerical representation."[118] What is the harm? Sure, there's the nuisance of poorly targeted ads, irrelevant who-to-follow suggestions, and search results built on these low-resolution personae. Just press the "X" and move on, right? But the real harm isn't so easily dismissed. It may begin as a subtle insult to our full-fledged-ness but then continues to far more concrete consequences.

Characterizing people based on observed sets of quantifiable parameters dehumanizes by overlooking their complex essence. Critics aptly observe that more refined data does not alleviate the problem but often deepens it, entrenching people's experience of technocratic dominance. A sense of violation results from "the *pretense* of the system to represent who we really are, leaving us with a lesser sense of our individuality."[119]

There is some obvious truth to this critique. No matter how robust the data scraping and collection campaigns, the resulting profiles are inexact along important dimensions. They are certainly incomplete, and potentially distorted. The primary and most common context for this critique concerns information harvested from social media. Personal data captured in these environs provide snapshots or people's statements and preferences, expressions that are often thoughtless, mischievous impulses, reflecting the cognitive patience and deliberation of a short elevator ride: "like" a post; view a "feed"; blurb a comment; snap a "pic." Even prudent people, who pursue other activities with deliberation and care, have social media silhouettes overflowing with infinitesimal manifestations of mindless spontaneity. Thoughtful actions are slow and few, whereas thoughtlessness is quick and common; it is no surprise which of the two populates the databases.

In this way, social media data captures mostly the sorrier halves of the personalities. Consequently, the algorithmic outputs—ads, recommendations, personalized news—are

[116] Crystal S. Yang et al., "Equal Protection Under Algorithms: A New Statistical and Legal Framework," *Michigan Law Review* 119 (2020): 291, 346–48.

[117] McKane Andrus et al., "What We Can't Measure, We Can't Understand": Challenges to Demographic Data Procurement in the Pursuit of Fairness," FAccT '21, March 3–10, 2021.

[118] Dan L. Burk, "Algorithmic Legal Metrics," *Notre Dame Law Review* 96, no. 3 (2021): 1147–1204.

[119] David Heyd, "Personalized Law and the Depersonalization of the Person," *Jerusalem Review of Legal Studies* 29, no. 1 (June 2024): 22–31.

tailored to attract and cultivate these sides. Imagine using social media data to predict how much insurance or retirement savings a person wants or needs. Insurance and retirement decisions are the stuff of System 2, our faculty for careful deliberation. The inputs collected from social media data belong to the less analytical System 1.[120] Surely, recommendations generated by data reflecting System 1 preferences should not be used to inform decisions that warrant System 2 rigor. This merging of systems fuses thought and action, collapsing the cognitive space meant for deliberation. It widens the gap between preferences and choice. And it could be weaponized to exploit cognitive lapses, deepen polarization and out-group bias, and erode self-control.[121] These are bad for the individuals and bad for society.

This critique is important, but does it justify data privacy law? Is the solution to simply switch off the data spigot? It's easy to see why concerned observers think so. But while data removal eliminates harmful personalized treatments, it also disrupts the beneficial ones. Socially valuable data technologies like facial recognition, usage-based insurance, or electronic health records bring enormous benefits while occasionally misfiring.

For instance, the problems of inaccuracy could be addressed at the point of use, not at the point of data collection. If police are too quick to rely on a tentative facial match in making an arrest—a problem that is real but not common—plenty of targeted safeguards are available to reduce such false positives without squandering the benefits of the technology. Or, if platforms design data-based applications that injure users, tort liability could be imposed without squandering the benefits of the platform immunity norm.[122]

Or, when a trained automaton rather than a human is tasked with identifying distinct features of a person, some error and dehumanization is present—by definition. But is this a price worth paying? The answer should obviously consider also the upside. If personalized medicine improves the efficacy of medical diagnosis or treatment and can save numerous lives, wouldn't some dehumanization in the delivery of care be justified? Also, let's not forget what would come in place of the data technology—human discretion, with its multitude of biases, far more stubborn.

There is a popular longing for the human touch. "Man alone can do the impossible: he can distinguish, he chooses and judges; ... he alone may reward the good and punish the wicked," wrote Goethe in "The Godlike".[123] Look no further than the American criminal justice system to be reminded how *actual* human judges constantly fall short of Goethe's self-congratulatory ideal. And yet that longing—a hubris that yields a sense of human exceptionalism—is unshakeable. Confronted with AI algorithms that recommend smaller criminal imprisonment to fewer people and achieve less racial bias, critics unfairly concentrate on their errors. Sure, there are intangibles that AI misses, for now—subtleties

[120] Amanda Agan et al., "Automating Automaticity: How the Context of Human Choice Affects the Extent of Algorithmic Bias," Becker Friedman Institute Working Paper No. 2023-19, February 2023.

[121] Jon Kleinberg et al., "The Challenge of Understanding What Users Want: Inconsistent Preferences and Engagement Optimization," *EC '22: Proceedings of the 23rd ACM Conference on Economics and Computation* (2022).

[122] See, e.g., Karan Lala, "TikTok the Tortfeasor: A Framework to Discuss Social-Platform Externalities and Arguments Favoring Ex Ante Mitigations," *University of Chicago Law Review* 91 (2024): 1765-1808.

[123] Goethe, "The Godlike," trans. Vernon Watkins, in *Selected Poems*, *Goethe: The Collected Works*, Vol. 1 (Princeton, NJ: Princeton University Press, 1994).

like eye contact and body language and inputs derived from intuition or empathy. I worry that the yearning for small-data human decisions resembles Dr. Johnson's description of second marriages: "the triumph of hope over experience."

In the end, my hunch is that fears around big data, especially coming from data privacy warriors, do not actually concern *inaccuracy* but exactly the opposite. The problem they see, what creeps many of us out, is uncanny *accuracy* of these digital predictive tools. Facial recognition may have had some false positives in the past, but the deeper worry is how good it has become. It is when the technology reaches such heights of accuracy, when it is terrifyingly close to recording people as they really are, anticipating what they'll do, and inferring who they like, that the privacy alarms toll. Data technologies are resented for their "intimate invasion," for planting "deep body periscopes" in private spaces to observe personal lives.[124] As accuracy grows exponentially, so does the perceived infringement of autonomy—to which I now turn.

E. Autonomy

> "Imposing privacy is disallowing people to demean their self-worth by yielding appropriate concern for the formation of reputation and self-concept."
>
> — Anita Allen (2011)

From their inception, privacy rights were a critical means to advance autonomy and dignity.[125] They stand for non-interruption of bodies and personal space, enabling people to flourish privately, develop intimate relationships, avoid embarrassment, explore more forms of personal expression, experience safety, and be themselves in manners separate from their public personae. Without privacy right, these good things might be chilled or sabotaged. Privacy is therefore valuable—not as an end, but to the extent that it facilitates the activities and capabilities that form the basis of an autonomous life.

To this privacy-as-means account, data privacy rights seem like a natural extension. So many of people's nonpublic explorations leave digital footprints which, if made public, might violate core autonomy values, or in turn lead to self-censorship and inhibition. Take, for example, health privacy. Bodies, genes, and illnesses are at the center of people's privacy interests, and publicizing them would breach core senses of dignity and safety. Information about personal health thus demands protection, and medical privacy laws like HIPAA intend to do exactly that, to afford people safe environments to pursue their health care.

But what happens when personal information, including personal health records, become "data"? Does the account of privacy-as-means-to-advance-autonomy change when digital records of many individuals are compiled into a database for the purpose of observing *aggregate* attributes? Are the core interests—to be left alone and flourish privately—equally threatened? And should the intensity of other social and private interests, which benefit from access to these datasets, shift the scope of data privacy rights?

[124] Hugo Jeanningros and Liz McFall, "The Value of Sharing: Branding and Behaviour in a Life and Health Insurance Company," *Big Data & Society* 7 (2020): 1–15.
[125] Samuel D. Warren and Louis Brandeis., "The Right to Privacy," *Harvard Law Review* 4 (1890): 193–220.

Before we discuss this in the context of health data, let's consider driving data—an area in which the loss of data privacy is said to undermine individuals' autonomy. Earlier in the chapter I mentioned Karen Levy's bleak account of the domination experienced by fleet drivers when their every move is tracked. Levy sees a clear derogation of drivers' autonomy. They told her: "a computer does not know when we are tired, fatigued, or anything else. . . . I am a grown man and have been on my own for many many years making responsible decisions." They also say that when their safety scores are made publicly visible, they feel shamed or embarrassed in front of co-workers—"the butt of next week's jokes."[126] These long-haul truckers embody an ethos of individualism and freedom, and their grievances regarding the impact of digital tracking on their sense of autonomy has plenty of credibility. The tradeoff is real and difficult—accident reduction versus a sovereign work environment.

But the tradeoff is not nearly as hard when people enroll in usage-based auto insurance, not least because they *choose* to opt in (more on consent in a minute). For car owners, driving is not a job, it is not a way of life, and it is certainly *not* an intimate act. Their driving is performed in public roads. It is guided by very useful navigation apps that already store location data. It is constrained by traffic laws, highway patrol, and unpleasant interaction with other drivers. It directly and dangerously impacts others in a very non-private way. It is not a leisure pursuit but a means to reach places. The digital monitoring of this activity is not akin to public exposure of the private and intimate aspects of one's life nor a violation of workplace autonomy. As succinctly observed by my wise colleague, privacy law expert Lior Strahilevitz, "There are plenty of privacy causes worth defending in contemporary society. [...] motorist obscurity is simply not one of them."[127]

When other colleagues at the University of Chicago heard my skepticism regarding drivers' data privacy, they urged me to recognize something that my foreign upbringing may have blurred—America's nearly-spiritual obsession with cars. Cars, they indicated, are sacred ground. Drivers would experience personal violation if tracking devices infiltrated their space.

They were not making this up—it is, I learned, a popular literary theme, and a cherished one. "Americans," wrote Saroyan, "have found the healing of God in a variety of things, the most pleasant of which is probably automobile drives."[128] Law professor Sarah Seo added a normative veneer to this theme, hailing driving as "a manifestation of freedom", a pleasurable activity where Americans can be "spontaneous and independent" unchained by "the dictates of social convention", where drivers experience "satisfaction of a deep desire that [is] vital to human flourishing."[129] Am I so sociologically clueless, so incognizant of the rituals of my adoptive land, that I have failed to recognize the misalignment between technology and folklore?

[126] Levy, *Data Driven*, 46–47, 70

[127] Lior J. Strahilevitz, "How's My Driving? For Everyone (and Everything)" *New York University Law Review* 81 81 (2006): 1699, 1744.

[128] William Saroyan, *Short Drive, Sweet Chariot* (New York, NY: Pocket Books, 1967).

[129] Sarah Seo, *Policing the Open Road: How Cars Transformed American Freedom* (Cambridge, MA: Harvard University Press, 2021).

Or maybe the poetic infatuation with driving the open road is a castle in the air? The bumper-to-bumper drivers *I* exchange honks with as we crawl along from O'Hare surely are not at the height of "human flourishing" and the only "healing of God" that comes to their mind is, perhaps, the stiff drink they will guzzle when the agony ends. In our day-to-day lives, we don't ride our convertibles along the Côte d'Azur to "satisfy our deep desires." We *commute*. In rush hour. Already running late. And it is a scientifically confirmed misery. Daniel Kahneman and co-authors surveyed and ranked people's satisfaction with daily activities. Guess which came last. Commuting, respondents say, is the "least enjoyable" activity, worse than housework. When asked why, respondents explain that what they find most agonizing on the road is the discourteous and dangerous actions of others.[130]

Ok, maybe driving is a slam dunk case. The safety benefits of tracking are massive, and the loss of autonomy when drivers give up some privacy rights and agree to be tracked is questionable. But what about medical privacy, an area where, supposedly, autonomy would be at risk were it not for the protections in current data privacy regime? The previous chapter discussed the privacy restrictions on use of electronic health records (EHRs) in treatment and in research. Protection of autonomy is, of course, the heart and soul of these privacy rules, and particularly the right of people to make informed decisions about their medical treatment and to control how their medical information is shared. But how does it extend to the use of databases compiled by EHRs?

The widely shared view among privacy advocates and bioethicists, which has become a cornerstone of medical privacy law, is that without informed consent by each patient whose data is contained in the electronic records, the use these records would violate the patient's autonomy. While most jurisdictions now allow providers to *store* health records electronically without patient consent, any sharing of these records— even for direct patient care and, less surprisingly, for any secondary uses—must still be approved by the patient. Societal interests, such as the improvement of care to others, the advancement of medical knowledge, or internal review of hospital performance, cannot be pursued absent specific consent.

In the abstract, this might seem like a proper scheme, especially under the view that a person owns their health information. Also supportive are references to privacy violations that would profoundly affect a person's autonomy, like the disclosure of an abortion procedure to a red-state government, or of an illness that carries social stigma. But this occasional heightened privacy need should not obscure the privacy/health tradeoff. Unfortunately, the tradeoff is often overlooked, with the autonomy basis for medical privacy creating a blanket regulatory priority for privacy over public health.

Let's carefully think what is at stake. In the previous chapter I mentioned the Malawi study that examined AIDS patients' right to opt-out of the use of EHRs, and the estimate that over 5000 lives were saved by the adoption of electronic records that allowed caregivers to better monitor the treatments.[131] Given the heavy social stigma associated with AIDS

[130] Daniel Kahneman et al., "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method," *Science* 306 (2004): 1776–1780.

[131] Laura Derksen, Anita McGahan, and Leandro Pongeluppe, "Privacy at What Cost? Using Electronic Medical Records to Recover Lapsed Patients Into HIV Care," 4 (mimeo., 2022).

diagnosis, this is a scenario of the utmost privacy interest; it is hard to think of a greater exercise of personal autonomy than by the choice to keep the diagnosis confidential, undocumented electronically. But there is a problem. In fact, two big problems. First, the decision of a patient to opt out of electronic records risks community spread of the disease and puts others—especially children—at the risk of deadly infection. In the name of data self-determination, the consent rule allows parents to deny their young AIDS-stricken children what has been shown to be a significant path to survival.[132] Second, the decision puts patients themselves at a heightened risk of death which they do not properly consider. In the Malawian AIDS context, the power to make privacy versus own-health tradeoff was bestowed upon individuals whose discretion was being shaped by the shock and trauma of a severe health diagnosis, by irrationally exaggerated fears of information leaks, and perhaps by instinctive distrust in modern healthcare and data systems. Patients' choices reflect their own perspectives, but also misinformation and misjudgments about how a data technology might impact their privacy.

This is an extreme example, but it serves to remind that the autonomy principle underlying medical privacy rights is not a trump card. There are contexts in which medical autonomy is profound, others, where it is superfluous, and still others where the sacrifices it requires are intolerable. One's health is very private, but through contagion becomes a public good.[133] When a data practice has vital effectiveness, when it poses only trivial threat to privacy, and when the consent—when given—is hardly ever informed, a cost-benefit framework is an ethically-compelling avenue for the calibration of privacy rights.

Finally, it might seem ironic that autonomy arguments are being used to shut down data collection, wholesale, to deny people the right to barter away the protection. This is the view of California with respect to the *voluntary* enrollment in usage-based auto insurance. Policyholders are aware of the bargain, including the ways that insurers' tracking might make them feel uncomfortable. While they might not be aware of what exactly is done with the tracking data, they are choosing to enroll despite their imperfect information. At any time, they can painlessly reverse their choice. The general autonomy-based argument against surveillance technologies seems misapplied in this context. Still, I have great sympathy for the view that consent does not purify every data transaction . . .

Non-Consent

. . . which brings us, finally, to a the most common and least effective method of privacy protection—the informed consent doctrine.

Our daily lives are full of transactions between parties unequally sophisticated. Consumers, patients, users, and employees interact with sellers, clinics, platforms, and employers, in manners varied in all aspects besides their *asymmetry*. Individuals have neither the information, understanding, or the stakes possessed by their experienced and well-

[132] Ibid, 4, stating "honoring patient requests for privacy can significantly hamper the effectiveness of an EMR system, and comes at the cost of disruptions in care and even deaths."
[133] Joseph R. Buckman et al., "Privacy Regulation and Barriers to Public Health," *Management Science* 69, no. 1 (2023): 342–350.

counseled counterparties. This imbalance causes much trouble to which we've dedicated and entire areas of transactional law—including all of data privacy law.

Asymmetry's most common regulatory solution, cutting across all areas of substantive laws, is informed consent. Behind this solution is an irresistible, uncontroversial premise: if people err in the face unfamiliar and complex decisions, supply them information until the task is familiar and comprehensible. Then, with the terms laid out before them, just ask if they agree to the relationship. Thus informed, their consent becomes a precaution, ensuring that people don't cede valuables like financial stability, physical health, and, yes, personal information, without thinking things through.

Much of the law created to protect health data privacy is that of informed consent. Electronic health records may be shared only after patients give permission. Human subjects research is approved, and it yields may be studied, only after the subjects receive full disclosure of the risks and express specific consent. Certainly, if people are considered to own their health information, informed consent to sharing is the number one rule to protect their autonomy.

But while informed consent had remained the darling of consumer protection regulation (as in Louis Brandeis' "sunlight is the best of disinfectants"), its critical limits are increasingly understood. People's affirmations of consent, even after full disclosure of the terms of engagement, are hardly ever informed. Why? Because the mastery required to choose wisely is just too tall an order. Things are too dense and complex. Every app, website, and device come with a data policy that could be thousands of words long. The content is unfamiliar and nuanced, leaving novices without the practical experience or intuition to navigate it competently. Informed consent has thus become a kabuki dance, a classic ritual of privacy theater, and everyone now cynically recognizes this.

Well, not *everyone*. In consumer protection law, most advocates and lawmakers believe that this paperwork problem is solvable. Its past failures are always attributed to an unfortunate format: the disclosures were too narrow, or too long, or too technical, or too obscure, or too early, or too late, or in the wrong place, or not conspicuous. The law thus tasks itself with requiring better drafted forms—expanded, simplified, tightened, emphasized, repeated, summarized, segregated, nudged, bursting with color—whatever it takes to make people digest the information. Sunstein is a big believer: "properly designed disclosure requirements can significantly improve the operation of markets, leading consumers to make more informed decisions."[134]

What "properly designed" means bounces between two polar views—the full-disclosure versus the pared-down template. In the area of medical privacy, the full-disclosure approach has its way (despite mountains of empirical evidence that it has never worked). Disclosure forms required by HIPAA are carefully drafted to provide comprehensive information. IRBs that oversee biomedical research tinker with every sentence of the multi-page consent forms with the stated goal of making them "clear, simple, unclouded,

[134] Cass R. Sunstein, "Empirically Informed Regulation," *University of Chicago Law Review* 78 (2011): 1349, 1366.

unhurried, and sensitive disclosure[s] that [give] the potential [research] participant all the information a reasonable person would need to make a well-informed decision."[135] By contrast, online data privacy—we saw in chapter 3—now wants to adopt the pared-down route. Inspired by the (perceived) success of the Nutrition Facts labels on processed food products, which greatly simplify the dietary information consumers seek (but, alas, have not improve Americans' eating habits), scholars advocate for simplified, "targeted," and "succinct" disclosures.[136]

You can read between the lines what I think about these strategies. I have never hidden my strong views here. In previous co-authored work, I explained why these re-engineering techniques have not worked in the past and will continue to flop: the quixotic search for simplification fails because the complex just isn't simple and can't be made so.[137] In the end, the evidence is clear cut: no matter how the data privacy notices are designed, people do not read them. Their "I Agree" clicks are the apex of uninformed contracting. At present, data privacy's notice-and-consent strategy is an exercise in the use of brute force—bombard people with repeated reminders, endless pop-up boxes, over-and-over-and-over, each with its consent button. Unfortunately, no matter how many times you multiply by zero, you still get zero.

Lately, the data privacy commentary has begun to recognize the unconquerable challenge of informed consent.[138] If consent may never be informed, lawmakers are now urged to deploy the more ambitious model of paternalistic regulation, which outright prohibits the collection of some types of data. This strategy, imposing mandatory limits on contracting, is not new. It is in the DNA of progressive legislation in areas like consumer credit and employment laws, prohibiting certain one-sided terms from loan or work contracts. In the same spirit, some data should not be collected even if people were to agree. This is a regulatory avenue increasingly favored by the EU and by some in the data privacy professoriate.[139]

To some in this camp, this view is based on distributive justice. Those with lower income will be more likely to bargain away their personal data, whereas the affluent will afford to pay in cash. Privacy should not be the currency of poverty, where people must trade their personal

[135] Institute of Medicine, *Responsible Research: A Systems Approach to Protecting Research Participants* (National Academies Press, 2003), viii.

[136] See, e.g., Ian Ayres and Alan Schwartz, "The No Reading Problem in Consumer Contract Law," *Stanford Law Review* 66 (2014): 545–610; Archon Fung et al., *Full Disclosure: The Perils and Promise of Transparency* (Cambridge: Cambridge University Press, 2007).

[137] Omri Ben-Shahar and Carl E. Schneider, *More Than You Wanted to Know: The Failure of Mandated Disclosure* (Princeton: Princeton University Press, 2014); Omri Ben-Shahar and Carl E. Schneider, "The Futility of Cost-Benefit Analysis in Financial Disclosure Regulation," *Journal of Legal Studies* 43 (2014): S253-S271; Omri Ben-Shahar and Adam Chilton, "Simplification of Privacy Disclosures: An Experimental Test," *Journal Legal Studies* 45 (2016): 541.

[138] Ignacio Cofone, *The Privacy Fallacy: Harm and Power in the Information Economy* (Cambridge University Press, 2023) 8, 12.

[139] See, e.g.,, Daniel J. Solove, "Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data," *Northwestern University Law Review* 118 (2024): 1081–1138; Bart Custers and Gianclaudio Malgiery, "Priceless data: why the EU fundamental right to data protection is at odds with trade in personal data," Computer Law & Security Review 45 (2022): 105683; Margaret Radin, *Boilerplate: The Fine Print, Vanishing Rights, and the Rule of Law* (Princeton University Press, 2012); Cofone, *The Privacy Fallacy*.

information for services they cannot otherwise afford. Fair, but I have not seen evidence of a disparate effect on the poor. Privacy is the new money: everyone seems to pay with data. Besides, businesses are more interested in the personal data of the affluent, so that they can send personalized ads direct to those with greater purchasing power.

A more interesting basis for mandatory privacy is courageously offered by Anita Allen in her excellent and prophetic book *Unpopular Privacy*. Privacy, Allen says, is not only a right, but also a duty, it is "ethically mandatory because respecting privacy is respecting civility norms of deference and demeanor." Contracting one's privacy away deprives them "of highly valued states that promote their vital interests, and those of fellow human beings with whom they associate."[140] Privacy is a "foundational good" for society at large because it supports trust and interpersonal relations. Like other fundamental human rights—the bodily autonomy, basic freedoms, and due process—it must remain inalienable.

Promotion of "vital interests" is, I recognize, a normative, not descriptive, concept. People may feel one way about what is vital for them, while moral theorists may think otherwise. There is surely a realm of inalienable rights, but is data privacy part of it? If so, when? The data privacy imperative must have some limiting criteria that tell us when other interests should supersede. When such trade-offs are present—when having more data privacy violates these criteria—mandatory privacy becomes a dogma, an orthodoxy, rather than wisdom.

Throughout the past two chapters, I worked to show that mandatory protection of data privacy does sacrifice other foundational goods, that it leads to more traffic casualties, that it weakens our fight against human trafficking, and that it creates impediments to critical medical progress. My disenchantment with the account of mandatory privacy that Allen offers is not in her description of data privacy's value. It is in the lack of pragmatism—the puzzling indifference to the good things that would have to be sacrificed.

[140] Anita L. Allen, *Unpopular Privacy: What Must We Hide* (Oxford: Oxford University Press, 2011), 172.