

E I G H T

t h e l i m i t s i n o p e n c o d e

I'VE TOLD A STORY ABOUT HOW REGULATION WORKS, AND ABOUT THE INCREASING regulability of the Internet that we should expect. These are, as I described, changes in the architecture of the Net that will better enable government's control by making behavior more easily monitored—or at least more traceable. These changes will emerge even if government does nothing. They are the by-product of changes made to enable e-commerce. But they will be cemented if (or when) the government recognizes just how it could make the network its tool.

That was Part I. In this part, I've focused upon a different regulability—the kind of regulation that is effected through the architectures of the space within which one lives. As I argued in Chapter 5, there's nothing new about this modality of regulation: Governments have used architecture to regulate behavior forever. But what is new is its significance. As life moves onto the Net, more of life will be regulated through the self-conscious design of the space within which life happens. That's not necessarily a bad thing. If there were a code-based way to stop drunk drivers, I'd be all for it. But neither is this pervasive code-based regulation benign. Due to the manner in which it functions, regulation by code can interfere with the ordinary democratic process by which we hold regulators accountable.

The key criticism that I've identified so far is transparency. Code-based regulation—especially of people who are not themselves technically expert—risks making regulation invisible. Controls are imposed for particular policy reasons, but people experience these controls as nature. And that experience, I suggested, could weaken democratic resolve.

Now that's not saying much, at least about us. We are already a pretty apathetic political culture. And there's nothing about cyberspace to suggest

things are going to be different. Indeed, as Castranova observes about virtual worlds: “How strange, then, that one does not find much democracy at all in synthetic worlds. Not a trace, in fact. Not a hint of a shadow of a trace. It’s not there. The typical governance model in synthetic worlds consists of isolated moments of oppressive tyranny embedded in widespread anarchy.”¹

But if we could put aside our own skepticism about our democracy for a moment, and focus at least upon aspects of the Internet and cyberspace that we all agree matter fundamentally, then I think we will all recognize a point that, once recognized, seems obvious: If code regulates, then in at least some critical contexts, the kind of code that regulates is critically important.

By “kind” I mean to distinguish between two types of code: open and closed. By “open code” I mean code (both software and hardware) whose functionality is transparent at least to one knowledgeable about the technology. By “closed code,” I mean code (both software and hardware) whose functionality is opaque. One can guess what closed code is doing; and with enough opportunity to test, one might well reverse engineer it. But from the technology itself, there is no reasonable way to discern what the functionality of the technology is.

The terms “open” and “closed” code will suggest to many a critically important debate about how software should be developed. What most call the “open source software movement,” but which I, following Richard Stallman, call the “free software movement,” argues (in my view at least) that there are fundamental values of freedom that demand that software be developed as free software. The opposite of free software, in this sense, is proprietary software, where the developer hides the functionality of the software by distributing digital objects that are opaque about the underlying design.

I will describe this debate more in the balance of this chapter. But importantly, the point I am making about “open” versus “closed” code is distinct from the point about how code gets created. I personally have very strong views about how code should be created. But whatever side you are on in the “free vs. proprietary software” debate in general, in at least the contexts I will identify here, you should be able to agree with me first, that open code is a constraint on state power, and second, that in at least some cases, code must, in the relevant sense, be “open.”

To set the stage for this argument, I want to describe two contexts in which I will argue that we all should agree that the kind of code deployed matters. The balance of the chapter then makes that argument.

BYTES THAT SNIFF

In Chapter 2, I described technology that at the time was a bit of science fiction. In the five years since, that fiction has become even less fictional. In 1997, the government announced a project called Carnivore. Carnivore was to be a technology that sifted through e-mail traffic and collected just those e-mails written by or to a particular and named individual. The FBI intended to use this technology, pursuant to court orders, to gather evidence while investigating crimes.

In principle, there's lots to praise in the ideals of the Carnivore design. The protocols required a judge to approve this surveillance. The technology was intended to collect data only about the target of the investigation. No one else was to be burdened by the tool. No one else was to have their privacy compromised.

But whether the technology did what it was said to do depends upon its code. And that code was closed.² The contract the government let with the vendor that developed the Carnivore software did not require that the source for the software be made public. It instead permitted the vendor to keep the code secret.

Now it's easy to understand why the vendor wanted its code kept secret. In general, inviting others to look at your code is much like inviting them to your house for dinner: There's lots you need to do to make the place presentable. In this case in particular, the DOJ may have been concerned about security.³ But substantively, however, the vendor might want to use components of the software in other software projects. If the code is public, the vendor might lose some advantage from that transparency. These advantages for the vendor mean that it would be more costly for the government to insist upon a technology that was delivered with its source code revealed. And so the question should be whether there's something the government gains from having the source code revealed.

And here's the obvious point: As the government quickly learned as it tried to sell the idea of Carnivore, the fact that its code was secret was costly. Much of the government's efforts were devoted to trying to build trust around its claim that Carnivore did just what it said it did. But the argument "I'm from the government, so trust me" doesn't have much weight. And thus, the efforts of the government to deploy this technology—again, a valuable technology if it did what it said it did—were hampered.

I don't know of any study that tries to evaluate the cost the government faced because of the skepticism about Carnivore versus the cost of developing Carnivore in an open way.⁴ I would be surprised if the government's strategy

made fiscal sense. But whether or not it was cheaper to develop closed rather than open code, it shouldn't be controversial that the government has an independent obligation to make its procedures—at least in the context of ordinary criminal prosecution—transparent. I don't mean that the investigator needs to reveal the things he thinks about when deciding which suspects to target. I mean instead the procedures for invading the privacy interests of ordinary citizens.

The only kind of code that can do that is “open code.” And the small point I want to insist upon just now is that where transparency of government action matters, so too should the kind of code it uses. This is not the claim that all government code should be public. I believe there are legitimate areas within which the government can act secretly. More particularly, where transparency would interfere with the function itself, then there's a good argument against transparency. But there were very limited ways in which a possible criminal suspect could more effectively evade the surveillance of Carnivore just because its code was open. And thus, again, open code should, in my view, have been the norm.

MACHINES THAT COUNT

Before November 7, 2000, there was very little discussion among national policy makers about the technology of voting machines. For most (and I was within this majority), the question of voting technology seemed trivial. Certainly, there could have been faster technologies for tallying a vote. And there could have been better technologies to check for errors. But the idea that anything important hung upon these details in technology was not an idea that made the cover of the front page of the *New York Times*.

The 2000 presidential election changed all that. More specifically, Florida in 2000 changed all that. Not only did the Florida experience demonstrate the imperfection in traditional mechanical devices for tabulating votes (exhibit 1, the hanging chad), it also demonstrated the extraordinary inequality that having different technologies in different parts of the state would produce. As Justice Stevens described in his dissent in *Bush v. Gore*, almost 4 percent of punch-card ballots were disqualified, while only 1.43 percent of optical scan ballots were disqualified.⁵ And as one study estimated, changing a single vote on each machine would have changed the outcome of the election.⁶

The 2004 election made things even worse. In the four years since the Florida debacle, a few companies had pushed to deploy new electronic voting machines. But these voting machines seemed to create more anxiety among voters than less. While most voters are not techies, everyone has a sense of the

obvious queasiness that a totally electronic voting machine produces. You stand before a terminal and press buttons to indicate your vote. The machine confirms your vote and then reports the vote has been recorded. But how do you know? How could anyone know? And even if you're not conspiracy-theory-oriented enough to believe that every voting machine is fixed, how can anyone know that when these voting machines check in with the central server, the server records their votes accurately? What's to guarantee that the numbers won't be fudged?

The most extreme example of this anxiety was produced by the leading electronic voting company, Diebold. In 2003, Diebold had been caught fudging the numbers associated with tests of its voting technology. Memos leaked to the public showed that Diebold's management knew the machines were flawed and intentionally chose to hide that fact. (The company then sued students who had published these memos—for copyright infringement. The students won a countersuit against Diebold.)

That incident seemed only to harden Diebold in its ways. The company continued to refuse to reveal anything about the code that its machines ran. It refused to bid in contexts in which such transparency was required. And when you tie that refusal to its chairman's promise to "deliver Ohio" for President Bush in 2004, you have all the makings of a perfect trust storm. You control the machines; you won't show us how they work; and you promise a particular result in the election. Is there any doubt people would be suspicious?⁷

Now it turns out that it is a very hard question to know how electronic voting machines should be designed. In one of my own dumbest moments since turning 21, I told a colleague that there was no reason to have a conference about electronic voting since all the issues were "perfectly obvious." They're not perfectly obvious. In fact, they're very difficult. It seems obvious to some that, like an ATM, there should at least be a printed receipt. But if there's a printed receipt, that would make it simple for voters to sell their votes. Moreover, there's no reason the receipt needs to reflect what was counted. Nor does the receipt necessarily reflect what was transmitted to any central tabulating authority. The question of how best to design these systems turns out not to be obvious. And having uttered absolute garbage about this point before, I won't enter here into any consideration of how best this might be architected.

But however a system is architected, there is an independent point about the openness of the code that comprises the system. Again, the procedures used to tabulate votes must be transparent. In the nondigital world, those procedures were obvious. In the digital world, however they're architected, we

need a way to ensure that the machine does what it is said it will do. One simple way to do that is either to open the code to those machines, or, at a minimum, require that that code be certified by independent inspectors. Many would prefer the latter to the former, just because transparency here might increase the chances of the code being hacked. My own intuition about that is different. But whether or not the code is completely open, requirements for certification are obvious. And for certification to function, the code for the technology must—in a limited sense at least—be open.

Both of these examples make a similar point. But that point, however, is not universal. There are times when code needs to be transparent, even if there are times when it does not. I'm not talking about all code for whatever purposes. I don't think Wal*Mart needs to reveal the code for calculating change at its check-out counters. I don't even think Yahoo! should have to reveal the code for its Instant Messaging service. But I do think we all should think that, in certain contexts at least, the transparency of open code should be a requirement.

This is a point that Phil Zimmermann taught by his practice more than 15 years ago. Zimmermann wrote and released to the Net a program called PGP (pretty good privacy). PGP provides cryptographic privacy and authentication. But Zimmermann recognized that it would not earn trust enough to provide these services well unless he made available the source code to the program. So from the beginning (except for a brief lapse when the program was owned by a company called NAI⁸) the source code has been available for anyone to review and verify. That publicity has built confidence in the code—a confidence that could never have been produced by mere command. In this case, open code served the purpose of the programmer, as his purpose was to build confidence and trust in a system that would support privacy and authentication. Open code worked.

The hard question is whether there's any claim to be made beyond this minimal one. That's the question for the balance of this chapter: How does open code affect regulability?

CODE ON THE NET

I've spent lots of time talking about "code." It's time to be a bit more specific about what "code" in the context of the Internet is, in what sense should we consider this code to be "open," and in what contexts its openness will matter.

As I've mentioned, the Internet is constructed by a set of protocols together referred to as TCP/IP. The TCP/IP suite includes a large number of

protocols that feed different “layers” of the network. The standard model for describing layers of a network is the open systems interconnect (OSI) reference model. It describes seven network layers, each representing a “function performed when data is transferred between cooperating applications across” the network. But the TCP/IP suite is not as well articulated in that model. According to Craig Hunt, “most descriptions of TCP/IP define three to five functional levels in the protocol architecture.” In my view, it is simplest to describe four functional layers in a TCP/IP architecture.⁹ From the bottom of the stack up, we can call these the data link, network, transport, and application layers.¹⁰

Three layers constitute the essential plumbing of the Internet, hidden in the Net’s walls. (The faucets work at the next layer; be patient.) At the very bottom, just above the physical layer of the Internet, in the data link layer, very few protocols operate, since that handles local network interactions exclusively. More protocols exist at the next layer up, the network layer, where the IP protocol is dominant. It routes data between hosts and across network links, determining which path the data should take. At the next layer up, the transport layer, two different protocols dominate—TCP and UDP. These negotiate the flow of data between two network hosts. (The difference between the two is reliability—UDP offers no reliability guarantee.)

The protocols together function as a kind of odd UPS. Data are passed from the application to the transport layer. There the data are placed in a (virtual) box and a (virtual) label is slapped on. That label ties the contents of the box to particular processes. (This is the work of the TCP or UDP protocols.) That box is then passed to the network layer, where the IP protocol puts the package into another package, with its own label. This label includes the origination and destination addresses. That box then can be further wrapped at the data link layer, depending on the specifics of the local network (whether, for example, it is an Ethernet network).

The whole process is thus a bizarre packaging game: A new box is added at each layer, and a new label on each box describes the process at that layer. At the other end, the packaging process is reversed: Like a Russian doll, each package is opened at the proper layer, until at the end the machine recovers the initial application data.

On top of these three layers is the application layer of the Internet. Here protocols “proliferate.”¹¹ These include the most familiar network application protocols, such as FTP (file transfer protocol, a protocol for transferring files), SMTP (simple mail transport protocol, a protocol for transferring mail), and HTTP (hyper text transfer protocol, a protocol to publish and read hypertext documents across the Web). These are rules for how a client (your computer)

will interact with a server (where the data are), or with another computer (in peer-to-peer services), and the other way around.¹²

These four layers of protocols are “the Internet.” Building on simple blocks, the system makes possible an extraordinary range of interaction. It is perhaps not quite as amazing as nature—think of DNA—but it is built on the same principle: keep the elements simple, and the compounds will astound.

When I speak about regulating the code, I’m not talking about changing these core TCP/IP protocols. (Though in principle, of course, they could be regulated, and others have suggested that they should be.)¹³ In my view these components of the network are fixed. If you required them to be different, you’d break the Internet. Thus rather than imagining the government changing the core, the question I want to consider is how the government might either (1) complement the core with technology that adds regulability, or (2) regulates applications that connect to the core. Both will be important, but my focus is on the code that plugs into the Internet. I will call that code the “application space” of the Internet. This includes all the code that implements TCP/IP protocols at the application layer—browsers, operating systems, encryption modules, Java, e-mail systems, P2P, whatever elements you want. The question for the balance of this chapter is: What is the character of that code that makes it susceptible to regulation?

A SHORT HISTORY OF CODE ON THE NET

In the beginning, of course, there were very few applications on the Net. The Net was no more than a protocol for exchanging data, and the original programs simply took advantage of this protocol. The file transfer protocol (FTP) was born early in the Net’s history;¹⁴ the electronic message protocol (SMTP) was born soon after. It was not long before a protocol to display directories in a graphical way (Gopher) was developed. And in 1991 the most famous of protocols—the hyper text transfer protocol (HTTP) and hyper text markup language (HTML)—gave birth to the World Wide Web.

Each protocol spawned many applications. Since no one had a monopoly on the protocol, no one had a monopoly on its implementation. There were many FTP applications and many e-mail servers. There were even a large number of browsers.¹⁵ The protocols were open standards, gaining their blessing from standards bodies such as the Internet Engineering Task Force (IETF) and, later, the W3C. Once a protocol was specified, programmers could build programs that utilized it.

Much of the software implementing these protocols was “open,” at least initially—that is, the source code for the software was available along with the

object code.¹⁶ This openness was responsible for much of the early Net's growth. Others could explore how a program was implemented and learn from that example how better to implement the protocol in the future.

The World Wide Web is the best example of this point. Again, the code that makes a web page appear as it does is called the hyper text markup language, or HTML.¹⁷ With HTML, you can specify how a web page will appear and to what it will be linked.

The original HTML was proposed in 1990 by the CERN researchers Tim Berners-Lee and Robert Cailliau.¹⁸ It was designed to make it easy to link documents at a research facility, but it quickly became obvious that documents on any machine on the Internet could be linked. Berners-Lee and Cailliau made both HTML and its companion HTTP freely available for anyone to take.

And take them people did, at first slowly, but then at an extraordinary rate. People started building web pages and linking them to others. HTML became one of the fastest-growing computer languages in the history of computing.

Why? One important reason was that HTML was always “open.” Even today, on most browsers in distribution, you can always reveal the “source” of a web page and see what makes it tick. The source remains open: You can download it, copy it, and improve it as you wish. Copyright law may protect the source code of a web page, but in reality it protects it very imperfectly. HTML became as popular as it did primarily because it was so easy to copy. Anyone, at any time, could look under the hood of an HTML document and learn how the author produced it.

Openness—not property or contract but free code and access—created the boom that gave birth to the Internet that we now know. And it was this boom that then attracted the attention of commerce. With all this activity, commerce rightly reasoned, surely there was money to be made.

Historically the commercial model for producing software has been different.¹⁹ Though the history began even as the open code movement continued, commercial software vendors were not about to produce “free” (what most call “open source”) software. Commercial vendors produced software that was closed—that traveled without its source and was protected against modification both by the law and by its own code.

By the second half of the 1990s—marked most famously by Microsoft's Windows 95, which came bundled Internet-savvy—commercial software vendors began producing “application space” code. This code was increasingly connected to the Net—it increasingly became code “on” the Internet. But for the most part, the code remained closed.

That began to change, however, around the turn of the century. Especially in the context of peer-to-peer services, technologies emerged that were dominant and “open.” More importantly, the protocols these technologies depended upon were unregulated. Thus, for example, the protocol that the peer-to-peer client Grokster used to share content on the Internet is itself an open standard that anyone can use. Many commercial entities tried to use that standard, at least until the Supreme Court’s decision in *Grokster*. But even if that decision inspires every commercial entity to abandon the StreamCast network, noncommercial implementations of the protocol will still exist.

The same mix between open and closed exists in both browsers and blogging software. Firefox is the more popular current implementation of the Mozilla technology—the technology that originally drove the Netscape browser. It competes with Microsoft’s Internet Explorer and a handful of other commercial browsers. Likewise, WordPress is an open-source blogging tool that competes with a handful of other proprietary blogging tools.

This recent growth in open code builds upon a long tradition. Part of the motivation for that tradition is ideological, or values based. Richard Stallman is the inspiration here. In 1984, Stallman began the Free Software Foundation with the aim of fueling the growth of free software. A MacArthur Fellow who gave up his career to commit himself to the cause, Stallman has devoted the last twenty years of his life to free software. That work began with the GNU project, which sought to develop a free operating system. By 1991, the GNU project had just about everything it needed, except a kernel. That final challenge was taken up by an undergraduate at the University of Helsinki. That year, Linus Torvalds posted on the Internet the kernel of an operating system. He invited the world to extend and experiment with it.

People took up the challenge, and slowly, through the early 1990s, marrying the GNU project with Torvald’s kernel, they built an operating system—GNU/Linux. By 1998, it had become apparent to all that GNU/Linux was going to be an important competitor to the Microsoft operating system. Microsoft may have imagined in 1995 that by 2000 there would be no other server operating system available except Windows NT, but when 2000 came around, there was GNU/Linux, presenting a serious threat to Microsoft in the server market. Now in 2007, Linux-based web servers continue to gain market share at the expense of Microsoft systems.

GNU/Linux is amazing in many ways. It is amazing first because it is theoretically imperfect but practically superior. Linus Torvalds rejected what computer science told him was the ideal operating system design,²⁰ and instead built an operating system that was designed for a single processor (an Intel 386) and not cross-platform-compatible. Its creative development, and

the energy it inspired, slowly turned GNU/Linux into an extraordinarily powerful system. As of this writing, GNU/Linux has been ported to at least eighteen different computer architecture platforms—from the original Intel processors, to Apple’s PowerPC chip, to Sun SPARC chips, and mobile devices using ARM processors.²¹ Creative hackers have even ported Linux to squeeze onto Apple’s iPod and old Atari systems. Although initially designed to speak only one language, GNU/Linux has become the lingua franca of free software operating systems.

What makes a system open is a commitment among its developers to keep its core code public—to keep the hood of the car unlocked. That commitment is not just a wish; Stallman encoded it in a license that sets the terms that control the future use of most free software. This is the Free Software Foundation’s General Public License (GPL), which requires that any code licensed with GPL (as GNU/Linux is) keep its source free. GNU/Linux was developed by an extraordinary collection of hackers worldwide only because its code was open for others to work on.

Its code, in other words, sits in the commons.²² Anyone can take it and use it as she wishes. Anyone can take it and come to understand how it works. The code of GNU/Linux is like a research program whose results are always published for others to see. Everything is public; anyone, without having to seek the permission of anyone else, may join the project.

This project has been wildly more successful than anyone ever imagined. In 1992, most would have said that it was impossible to build a free operating system from volunteers around the world. In 2002, no one could doubt it anymore. But if the impossible could become possible, then no doubt it could become impossible again. And certain trends in computing technology may create precisely this threat.

For example, consider the way Active Server Pages (ASP) code works on the network. When you go to an ASP page on the Internet, the server runs a program—a script to give you access to a database, for example, or a program to generate new data you need. ASPs are increasingly popular ways to provide program functionality. You use it all the time when you are on the Internet.

But the code that runs ASPs is not technically “distributed.” Thus, even if the code is produced using GPL’d code, there’s no GPL obligation to release it to anyone. Therefore, as more and more of the infrastructure of networked life becomes governed by ASP, less and less will be effectively set free by free license.

“Trusted Computing” creates another threat to the open code ecology. Launched as a response to virus and security threats within a networked environment, the key technical feature of “trusted computing” is that the platform

blocks programs that are not cryptographically signed or verified by the platform. For example, if you want to run a program on your computer, your computer would first verify that the program is certified by one of the authorities recognized by the computer operating system, and “incorporat[ing] hardware and software . . . security standards approved by the content providers themselves.”²³ If it isn’t, the program wouldn’t run.

In principle, of course, if the cost of certifying a program were tiny, this limitation might be unproblematic. But the fear is that this restriction will operate to effectively block open code projects. It is not easy for a certifying authority to actually know what a program does; that means certifying authorities won’t be keen to certify programs they can’t trust. And that in turn will effect a significant discrimination against open code.

REGULATING OPEN CODE

Open code projects—whether free software or open source software projects—share the feature that the knowledge necessary to replicate the project is intended always to be available to others. There is no effort, through law or technology, for the developer of an open code project to make that development exclusive. And, more importantly, the capacity to replicate and redirect the evolution of a project provided in its most efficient form is also always preserved.

How does this fact affect the regulability of code?

In Chapter 5, I sketched examples of government regulating code. But think again about those examples: How does such regulation work?

Consider two. The government tells the telephone company something about how its networks are to be designed, and the government tells television manufacturers what kinds of chips TVs are to have. Why do these regulations work?

The answer in each case is obvious. The code is regulable only because the code writers can be controlled. If the state tells the phone company to do something, the phone company is not likely to resist. Resistance would bring punishment; punishment is expensive; phone companies, like all other companies, want to reduce the cost of doing business. If the state’s regulation is rational (that is, effective), it will set the cost of disobeying the state above any possible benefit. If the target of regulation is a rational actor within the reach of the state, then the regulation is likely to have its intended effect. CALEA’s regulation of the network architecture for telephones is an obvious example of this (see Chapter 5).

An unmovable, and unmoving, target of regulation, then, is a good start toward regulability. And this statement has an interesting corollary: Regulable code is closed code. Think again about telephone networks. When the govern-

ment induces the telephone networks to modify their network software, users have no choice about whether to adopt this modification or not. You pick up the phone, you get the dial tone the phone company gives you. No one I know hacks the telephone company's code to build a different network design. The same with the V-chip—I doubt that many people would risk destroying their television by pulling out the chip, and I am certain that no one re-burns the chip to build in a different filtering technology.

In both cases the government's regulation works because when the target of the regulation complies, customers can do little but accept it.

Open code is different. We can see something of the difference in a story told by Netscape's former legal counsel, Peter Harter, about Netscape and the French.²⁴

In 1996, Netscape released a protocol (SSL v3.0) to facilitate secure electronic commerce on the Web. The essence of its function is to permit secure exchange between a browser and a server. The French were not happy with the security that SSL gave; they wanted to be able to crack SSL transactions. So they requested that Netscape modify SSL to enable their spying.

There are plenty of constraints on Netscape's ability to modify SSL—not the least of which being that Netscape has given SSL over to the public, in the form of a public standard. But assume for a second that it had not. Assume Netscape really did control the standards for SSL and in theory could modify the code to enable French spying. Would that mean that Netscape could comply with the French demand?

No. Technically, it could comply by modifying the code of Netscape Communicator and then posting a new module that enabled hacking by a government. But because Netscape (or more generally, the Mozilla project) is open source, anyone is free to build a competing module that would replace the Frenchified SSL module. That module would compete with other modules. The module that wins would be the one users wanted. Users don't typically want a module that enables spying by a government.

The point is simple, but its implication is profound. To the extent that code is open code, the power of government is constrained. Government can demand, government can threaten, but when the target of its regulation is plastic, it cannot rely on its target remaining as it wants.

Say you are a Soviet propagandist, and you want to get people to read lots of information about Papa Stalin. So you declare that every book published in the Soviet Union must have a chapter devoted to Stalin. How likely is it that such books will actually affect what people read?

Books are open code: They hide nothing; they reveal their source—they are their source! A user or adopter of a book always has the choice to read only the

chapters she wants. If it is a book on electronics, then the reader can certainly choose not to read the chapter on Stalin. There is very little the state can do to modify the reader's power in this respect.

The same idea liberates open code. The government's rules are rules only to the extent that they impose restrictions that adopters would want. The government may coordinate standards (like "drive on the right"), but it certainly cannot impose standards that constrain users in ways they do not want to be constrained. This architecture, then, is an important check on the government's regulatory power. Open code means open control—there is control, but the user is aware of it.²⁵

Closed code functions differently. With closed code, users cannot easily modify the control that the code comes packaged with. Hackers and very sophisticated programmers may be able to do so, but most users would not know which parts were required and which parts were not. Or more precisely, users would not be able to see the parts required and the parts not required because the source code does not come bundled with closed code. Closed code is the propagandist's best strategy—not a separate chapter that the user can ignore, but a persistent and unrecognized influence that tilts the story in the direction the propagandist wants.

So far I've played fast and loose with the idea of a "user." While some "users" of Firefox could change its code if they didn't like the way it functioned, the vast majority could not. For most of us, it is just as feasible to change the way Microsoft Word functions as it is to change the way GNU/Linux operates.

But the difference here is that there is—and legally can be—a community of developers who modify open code, but there is not—or legally cannot be—a community of developers who modify closed code, at least without the owner's permission. That culture of developers is the critical mechanism that creates the independence within open code. Without that culture, there'd be little real difference between the regulability of open and closed code.

This in turn implies a different sort of limit on this limit on the regulability of code. Communities of developers are likely to enable some types of deviations from rules imposed by governments. For example, they're quite likely to resist the kind of regulation by the French to enable the cracking of financial safety. They're less likely to disable virus protection or spam filters.

WHERE THIS LEADS

My argument so far has taken a simple path. In answer to those who say that the Net cannot be regulated, I've argued that whether it can be regulated depends on its architecture. Some architectures would be regulable, others

would not. I have then argued that government could take a role in deciding whether an architecture would be regulable or not. The government could take steps to transform an architecture from unregulable to regulable, both indirectly (by making behavior more traceable) and directly (by using code to directly effect the control the government wants).

The final step in this progression of regulability is a constraint that is only now becoming significant. Government's power to regulate code, to make behavior within the code regulable, depends in part on the character of the code. Open code is less regulable than closed code; to the extent that code becomes open, government's power is reduced.

Take for example the most prominent recent controversy in the area of copyright—peer-to-peer filesharing. As I've described, P2P filesharing is an application that runs on the network. Filesharing networks like StreamCast are simply protocols that P2P applications run. All these protocols are open; anyone can build to them. And because the technology for building to them is widely available, whether or not a particular company builds to them doesn't affect whether they will be built to—but demand does.

Thus, imagine for the moment that the recording industry is successful in driving out of business every business that supports P2P filesharing. The industry won't be successful in driving P2P out of existence. This is because open code has enabled noncommercial actors to sustain the infrastructure of P2P sharing, without the commercial infrastructure.

This is not, obviously, an absolute claim. I am discussing relative, not absolute, regulability. Even with open code, if the government threatens punishments that are severe enough, it will induce a certain compliance. And even with open code, the techniques of identity, tied to code that has been certified as compliant, will still give government plenty of power. Thus, much of the argument from Part I survives this point about open code—if the world becomes certificate-rich, regulability still increases. The same conclusion follows if more code were burned into hardware rather than left to exist as software. Then, even if the code were open, it would not be modifiable.²⁶

But when designing an architecture for cyberspace, the margins matter. The values of a given space are not only the values of speech, autonomy, access, or privacy. They may also be values of limited control. As John Perry Barlow puts it, they are the values of a certain bug being programmed into the architecture of the Net—a bug that inhibits the power of government to control the Net perfectly, even if it does not disable that power entirely.

For some, the objective is to build code that disables any possible governmental control. That is not my objective. I certainly believe that government must be constrained, and I endorse the constraints that open code imposes,

but it is not my objective to disable government generally. As I've argued already, and as the next part makes plain, some values can be achieved only if government intervenes. Government has a role, even if not as substantial a role as it would wish. We need to understand this role, as well as how our values might be advanced in the context of the Web.

One constraint seems clear in this account. As I argue more extensively later in the book, even if open code does not disable government's power to regulate completely, it certainly changes that power. On the margin, open code reduces the reward from burying regulation in the hidden spaces of code. It functions as a kind of Freedom of Information Act for network regulation. As with ordinary law, open code requires that lawmaking be public, and thus that lawmaking be transparent. In a sense that George Soros ought to understand, open code is a foundation to an open society.

Even this is an important—some might say an essential—check on the power of government. But whether or not one is for transparency generally, my aim so far is just to map out the links. Regulability is conditional on the character of the code, and open code changes that character. It is a limit on government's power to regulate—not necessarily by defeating the power to regulate, but by changing it.

T E N

i n t e l l e c t u a l p r o p e r t y

HAROLD REEVES IS AMONG THE BEST RESEARCH ASSISTANTS I HAVE HAD. (BUT ALAS, the law has now lost him—he’s become a priest!). Early into his second year at the University of Chicago Law School, he came to me with an idea he had for a student “comment”—an article that would be published in the law review.¹ The topic was trespass law in cyberspace—whether and how the law should protect owners of space in cyberspace from the kinds of intrusions that trespass law protects against in real space. His initial idea was simple: There should be no trespass law in cyberspace.² The law should grant “owners” of space in cyberspace no legal protection against invasion; they should be forced to fend for themselves.

Reeves’s idea was a bit nutty, and in the end, I think, wrong.³ But it contained an insight that was quite brilliant, and that should be central to thinking about law in cyberspace.

The idea—much more briefly and much less elegantly than Reeves has put it—is this: The question that law should ask is, What means would bring about the most efficient set of protections for property interests in cyberspace? Two sorts of protections are possible. One is the traditional protection of law—the law defines a space where others should not enter and punishes people who enter nonetheless. The other protection is a fence, a technological device (a bit of code) that (among other things) blocks the unwanted from entering. In real space, of course, we have both—law, in the form of trespass law, and fences that supplement that law. Both cost money, and the return from each is not necessarily the same. From a social perspective, we would want the mix that provides optimal protection at the lowest cost. (In economics-speak, we would want a mix such that the marginal cost of an additional unit of protection is equivalent to the marginal benefit.)

The implication of this idea in real space is that it sometimes makes sense to shift the burden of protection to citizens rather than to the state. If, for example, a farmer wants to store some valuable seed on a remote part of his farm, it is better for him to bear the cost of fencing in the seed than to require the police to patrol the area more consistently or to increase the punishment for those they catch. The question is always one of balance between the costs and benefits of private protection and state protection.

Reeves's insight about cyberspace follows the same line. The optimal protection for spaces in cyberspace is a mix between public law and private fences. The question to ask in determining the mix is which protection, on the margin, costs less. Reeves argues that the costs of law in this context are extremely high—in part because of the costs of enforcement, but also because it is hard for the law to distinguish between legitimate and illegitimate uses of cyberspaces. There are many “agents” that might “use” the space of cyberspace. Web spiders, which gather data for web search engines; browsers, who are searching across the Net for stuff to see; hackers (of the good sort) who are testing the locks of spaces to see that they are locked; and hackers (of the bad sort) who are breaking and entering to steal. It is hard, *ex ante*, for the law to know which agent is using the space legitimately and which is not. Legitimacy depends on the intention of the person granting access.

So that led Reeves to his idea: Since the intent of the “owner” is so crucial here, and since the fences of cyberspace can be made to reflect that intent cheaply, it is best to put all the incentive on the owner to define access as he wishes. The right to browse should be the norm, and the burden to lock doors should be placed on the owner.⁴

Now put Reeves's argument aside, and think for a second about something that will seem completely different but is very much the same idea. Think about “theft” and the protections that we have against it.

- I have a stack of firewood behind my house. No one steals it. If I left my bike out overnight, it would be gone.
- A friend told me that, in a favorite beach town, the city used to find it impossible to plant flowers—they would immediately be picked. But now, he proudly reports, after a long “community spirit” campaign, the flowers are no longer picked.
- There are special laws about the theft of automobiles, planes, and boats. There are no special laws about the theft of skyscrapers. Cars, planes, and boats need protection. Skyscrapers pretty much take care of themselves.

Many things protect property against theft—differently. The market protects my firewood (it is cheaper to buy your own than it is to haul mine away); the market is a special threat to my bike (which if taken is easily sold). Norms sometimes protect flowers in a park; sometimes they do not. Nature sometimes conspires with thieves (cars, planes, and boats) and sometimes against them (skyscrapers).

These protections are not fixed. I could lock my bike and thereby use real-space code to make it harder to steal. There could be a shortage of firewood; demand would increase, making it harder to protect. Public campaigns about civic beauty might stop flower theft; selecting a distinctive flower might do the same. Sophisticated locks might make stolen cars useless; sophisticated bank fraud might make skyscrapers vulnerable. The point is not that protections are given, or unchangeable, but that they are multiplied and their modalities different.

Property is protected by the sum of the different protections that law, norms, the market, and real-space code yield. This is the implication of the argument made in Chapter 7. From the point of view of the state, we need law only when the other three modalities leave property vulnerable. From the point of view of the citizen, real-space code (such as locks) is needed when laws and norms alone do not protect enough. Understanding how property is protected means understanding how these different protections work together.

Reeves's idea and these reflections on firewood and skyscrapers point to the different ways that law might protect "property" and suggest the range of kinds of property that law might try to protect. They also invite a question that has been asked by Justice Stephen Breyer and many others: Should law protect some kinds of property—in particular, intellectual property—at all?⁵

Among the kinds of property law might protect, my focus in this chapter will be on the property protected by copyright.⁶ Of all the different types of property, this type is said to be the most vulnerable to the changes that cyberspace will bring. Many believe that intellectual property cannot be protected in cyberspace. And in the terms that I've sketched, we can begin to see why one might think this, but we will soon see that this thought must be wrong.

ON THE REPORTS OF COPYRIGHT'S DEMISE

Roughly put, copyright gives a copyright holder certain exclusive rights over the work, including, most famously, the exclusive right to copy the work. I have a copyright in this book. That means, among other rights, and subject to some important exceptions, you cannot copy this book without my permission. The right is protected to the extent that laws (and norms) support it, and

it is threatened to the extent that technology makes it easy to copy. Strengthen the law while holding technology constant, and the right is stronger. Proliferate copying technology while holding the law constant, and the right is weaker.

In this sense, copyright has always been at war with technology. Before the printing press, there was not much need to protect an author's interest in his creative work. Copying was so expensive that nature itself protected that interest. But as the cost of copying decreased, and the spread of technologies for copying increased, the threat to the author's control increased. As each generation has delivered a technology better than the last, the ability of the copyright holder to protect her intellectual property has been weakened.

Until recently, the law's response to these changes has been measured and gradual. When technologies to record and reproduce sound emerged at the turn of the last century, composers were threatened by them. The law responded by giving composers a new, but limited, right to profit from recordings. When radio began broadcasting music, the composers were held to be entitled to compensation for the public performance of their work, but performers were not compensated for the "performance" of their recordings. Congress decided not to remedy that problem. When cable television started rebroadcasting television broadcasts, the copyright holders in the original broadcasts complained their work was being exploited without compensation. Congress responded by granting the copyright holders a new, but limited, right to profit from the rebroadcasts. When the VCR made it simple to record copyrighted content from off the air, copyright holders cried "piracy." Congress decided not to respond to that complaint. Sometimes the change in technology inspired Congress to create new rights, and sometimes not. But throughout this history, new technologies have been embraced as they have enabled the spread of culture.

During the same period, norms about copyrighted content also evolved. But the single, defining feature of these norms can perhaps be summarized like this: that a consumer could do with the copyrighted content that he legally owned anything he wanted to do, without ever triggering the law of copyright. This norm was true almost by definition until 1909, since before then, the law didn't regulate "copies." Any use the consumer made of copyrighted content was therefore highly unlikely to trigger any of the exclusive rights of copyright. After 1909, though the law technically regulated "copies," the technologies to make copies were broadly available. There was a struggle about Xerox machines, which forced a bit of reform,⁷ but the first real conflict that copyright law had with consumers happened when cassette tapes made it easy to copy recorded music. Some of that copying was for the purpose of making a

“mixed tape,” and some was simply for the purpose of avoiding the need to buy the original recording. After many years of debate, Congress decided not to legislate a ban on home taping. Instead, in the Audio Home Recording Act, Congress signaled fairly clear exemptions from copyright for such consumer activity. These changes reinforced the norm among consumers that they were legally free to do whatever they wanted with copyrighted work. Given the technologies most consumers had access to, the stuff they wanted to do either did not trigger copyright (e.g., resell their books to a used bookstore), or if it did, the law was modified to protect it (e.g., cassette tapes).

Against the background of these gradual changes in the law, along with the practical norm that, in the main, the law didn't reach consumers, the changes of digital technology were a considerable shock. First, from the perspective of technology, digital technologies, unlike their analog sister, enabled perfect copies of an original work. The return from copying was therefore greater. Second, also from the perspective of technology, the digital technology of the Internet enabled content to be freely (and effectively anonymously) distributed across the Internet. The availability of copies was therefore greater. Third, from the perspective of norms, consumers who had internalized the norm that they could do with “their content” whatever they wanted used these new digital tools to make “their content” available widely on the Internet. Companies such as Napster helped fuel this behavior, but the practice existed both before and after Napster. And fourth, from the perspective of law, because the base technology of the Internet didn't reveal anything about the nature of the content being shared on the Internet, or about who was doing the sharing, there was little the law could do to stop this massive “sharing” of content. Thus fifth, and from the perspective of copyright holders, digital technologies and the Internet were the perfect storm for their business model: If they made money by controlling the distribution of “copies” of copyrighted content, you could well understand why they viewed the Internet as a grave threat.

Very quickly, and quite early on, the content industry responded to this threat. Their first line of defense was a more aggressive regime of regulation. Because, the predictions of cyberspace mavens notwithstanding, not everyone was willing to concede that copyright law was dead. Intellectual property lawyers and interest groups pushed early on to have law shore up the protections of intellectual property that cyberspace seemed certain to erase.

LAW TO THE RESCUE

The initial response to this push was a White Paper produced by the Commerce Department in 1995. The paper outlined a series of modifications

aimed, it said, at restoring “balance” in intellectual property law. Entitled “Intellectual Property and the National Information Infrastructure,” the report sought to restate existing intellectual property law in terms that anyone could understand, as well as to recommend changes in the law in response to the changes the Net would bring. But as scholars quickly pointed out, the first part was a bust.⁸ The report no more “restated” existing law than Soviet historians “retold” stories of Stalin’s administration. The restatement had a tilt, very definitely in the direction of increased intellectual property protection, but it pretended that its tilt was the natural lay of the land.

For our purposes, however, it is the recommendations that were most significant. The government proposed four responses to the threat presented by cyberspace. In the terms of Chapter 7, these responses should be familiar.

The first response was traditional. The government proposed changes in the law of copyright to “clarify” the rights that it was to protect.⁹ These changes were intended to better define the rights granted under intellectual property law and to further support these rights with clarified (and possibly greater) legal penalties for their violation.

The second response addressed norms, specifically copying norms. The report recommended increased educational efforts, both in schools and among the general public, about the nature of intellectual property and the importance of protecting it. In the terms of Chapter 7, this is the use of law to change norms so that norms will better support the protection of intellectual property. It is an indirect regulation of behavior by direct regulation of norms.

The third and fourth responses mixed technology and the market. The report called for legal support—through financial subsidies and special legal protection—of “copyright management schemes.” These “schemes” were simply technologies that would make it easier to control access to and use of copyrighted material. We will explore these “schemes” at some length later in this chapter, but I mention them now as another example of indirect regulation—using the market to subsidize the development of a certain software tool, and using law to regulate the properties of other software tools. Copyright management systems would be supported by government funding and by the threat of criminal sanctions for anyone deploying software to crack them.¹⁰

Congress followed the recommendations of the 1995 White Paper in some respects. The most important was the enactment of the Digital Millennium Copyright Act in 1998. That statute implemented directly the recommendation that “technological protection measures” be protected by law. Code that someone implements to control either access to or use of a copyrighted work got

special legal protection under the DMCA: Circumvention of that code, subject to a few important exceptions, constituted a violation of the law.

We will return to the DMCA later. The point just now, however, is to recognize something important about the presumption underlying the White Paper. The 1995 package of proposals was a scattershot of techniques—some changes in law, some support for changing norms, and lots of support for changing the code of cyberspace to make it better able to protect intellectual property. Perhaps nothing better than this could have been expected in 1995—the law promised a balance of responses to deal with the shifting balance brought on by cyberspace.

Balance is attractive, and moderation seems right. But something is missing from this approach. The White Paper proceeds as if the problem of protecting intellectual property in cyberspace was just like the problem of protecting intellectual property in real space. It proceeds as if the four constraints would operate in the same proportions as in real space, as if nothing fundamental had changed.

But something fundamental has changed: the role that code plays in the protection of intellectual property. Code can, and increasingly will, displace law as the primary defense of intellectual property in cyberspace. Private fences, not public law.

The White Paper did not see this. Built into its scattershot of ideas is one that is crucial to its approach but fundamentally incorrect—the idea that the nature of cyberspace is anarchy. The White Paper promises to strengthen law in every area it can. But it approaches the question like a ship battening down for a storm: Whatever happens, the threat to copyright is real, damage will be done, and the best we can do is ride it out.

This is fundamentally wrong. We are not entering a time when copyright is more threatened than it is in real space. We are instead entering a time when copyright is more effectively protected than at any time since Gutenberg. The power to regulate access to and use of copyrighted material is about to be perfected. Whatever the mavens of the mid-1990s may have thought, cyberspace is about to give holders of copyrighted property the biggest gift of protection they have ever known.

In such an age, the real question for law is not, how can law aid in that protection? but rather, is the protection too great? The mavens were right when they predicted that cyberspace will teach us that everything we thought about copyright was wrong.¹¹ But the lesson in the future will be that copyright is protected far too well. The problem will center not on copy-right but on copy-duty—the duty of owners of protected property to make that property accessible.

That's a big claim. To see it, however, and to see the consequences it entails, we need consider three examples. The first is a vision of a researcher from Xerox PARC (appropriately enough), Mark Stefik, and his idea of "trusted systems."¹² The second is an implication of a world dominated by trusted systems. The third is an unreckoned cost to the path we are now on to "protect intellectual property." The examples will throw into relief the threat that these changes present for values that our tradition considers fundamental. They should force us to make a choice about those values, and about their place in our future.

THE PROMISE FOR INTELLECTUAL PROPERTY IN CYBERSPACE

It all depends on whether you really understand the idea of trusted systems. If you don't understand them, then this whole approach to commerce and digital publishing is utterly unthinkable. If you do understand them, then it all follows easily.

Ralph Merkle, quoted in Stefik, "Letting Loose the Light" (1996)

In what we can call the first generation of digital technologies, content owners were unable to control who copied what. If you have a copy of a copyrighted photo rendered in a graphics file, you could make unlimited copies of that file with no effect on the original. When you make the one-hundredth copy, nothing would indicate that it was the one-hundredth copy rather than the first. And as we've described again and again, in the original code of the Internet, there was nothing to regulate how or to whom copyrighted content was distributed. The function of "copying" as it was developed by the coders who built it, either in computers or networks, aimed at "copying"—not at "copying" with specified permissions.

This character to the function "copy" was not unique to cyberspace. We have seen a technology that presented the same problem, and I've already described how a solution was subsequently built into the technology.¹³ Digital Audio Tape (DAT) technology was thought to be a threat to copyright owners. A number of solutions to this threat were proposed. Some people argued for higher penalties for illegal copying of tapes (direct regulation by law). Some, such as Richard Stallman, argued for a tax on blank tapes, with the proceeds compensating copyright holders (indirect regulation of the market by law). Some argued for better education to stop illegal copies of tapes (indirect regulation of norms by law). But some argued for a change in the code of DAT machines that would block unlimited perfect copying.

The tax and code regulators won. In late 1992, as a compromise between the technology and content industries, Congress passed the Audio Home Recording Act. The act first imposed a tax on both recorders and blank DAT media, with the revenues to be used to compensate copyright holders for the expected copyright infringement enabled by the technology. But more interestingly, the Act required manufacturers of DAT technology to include a Serial Copy Management System, which would limit the ability of DAT technology to copy. That limit was effected through a code inserted in copies made using DAT technology. From an original, the technology would always permit a copy. But from a copy made on a DAT recorder, no further digital copy could be made. (An analog copy could be made, thus degrading the quality of the copy, but not a perfect digital copy.) The technology was thus designed to break the “copy” function under certain conditions, so as to indirectly protect copyright owners. The net effect of these two changes was to minimize any harm from the technology, as well as to limit the functionality of the technology where it would be expected that functionality would encourage the violation of copyright. (Many think the net effect of this regulation also killed DAT technology.)

Something like the same idea animated Stefik’s vision.¹⁴ He was not keen to make the quality of copies decrease. Rather, his objective was to make it possible to track and control the copies of digital content that are made.¹⁵

Think of the proposal like this. Today, when you buy a book, you may do any number of things with it. You can read it once or one hundred times. You can lend it to a friend. You can photocopy pages in it or scan it into your computer. You can burn it, use it as a paperweight, or sell it. You can store it on your shelf and never once open it.

Some of these things you can do because the law gives you the right to do them—you can sell the book, for example, because the copyright law explicitly limits the copyright owner’s right to control your use of the physical book after the “first sale.” Other things you can do because there is no effective way to stop you. A book seller might sell you the book at one price if you promise to read it once, and at a different price if you want to read it one hundred times, but there is no way for the seller to know whether you have obeyed the contract. In principle, the seller could sell a police officer with each book to follow you around and make sure you use the book as you promised, but the costs of this control would plainly exceed any benefit.

But what if each of these rights could be controlled, and each unbundled and sold separately? What if, that is, the software itself could regulate whether you read the book once or one hundred times; whether you could cut and paste from it or simply read it without copying; whether you could send it as

an attached document to a friend or simply keep it on your machine; whether you could delete it or not; whether you could use it in another work, for another purpose, or not; or whether you could simply have it on your shelf or have it and use it as well?

Stefik describes a network that makes such unbundling of rights possible. He describes an architecture that would allow owners of copyrighted materials to sell access to those materials on the terms they want and would enforce those contracts.

The details of the system are not important here (it builds on the encryption architecture I described in Chapter 4),¹⁶ but its general idea is easy enough to describe. As the Net is now, basic functions like copying and access are crudely regulated in an all-or-nothing fashion. You generally have the right to copy or not, to gain access or not.

But a more sophisticated system of rights could be built into the Net—not into a different Net, but on top of the existing Net. This system would function by discriminating in the intercourse it has with other systems. A system that controlled access in this more fine-grained way would grant access to its resources only to another system that controlled access in the same way. A hierarchy of systems would develop, and copyrighted material would be traded only among systems that properly controlled access.

In such a world, then, you could get access, say, to the *New York Times* and pay a different price depending on how much of it you read. The *Times* could determine how much you read, whether you could copy portions of the newspaper, whether you could save it on your hard disk, and so on. But if the code you used to access the *Times* site did not enable the control the *Times* demanded, then the *Times* would not let you onto its site at all. In short, systems would exchange information only with others that could be trusted, and the protocols of trust would be built into the architectures of the systems.

Stefik calls this “trusted systems,” and the name evokes a helpful analog. Think of bonded couriers. Sometimes you want to mail a letter with something particularly valuable in it. You could simply give it to the post office, but the post office is not a terribly reliable system; it has relatively little control over its employees, and theft and loss are not uncommon. So instead of going to the post office, you could give your letter to a bonded courier. Bonded couriers are insured, and the insurance is a cost that constrains them to be reliable. This reputation then makes it possible for senders of valuable material to be assured about using their services. As Stefik writes:

with trusted systems, a substantial part of the enforcement of a digital contract is carried out by the trusted system. [T]he consumer does not have the option of

disregarding a digital contract by, for example, making unauthorized copies of a work. A trusted system refuses to exercise a right that is not sanctioned by the digital contract.¹⁷

This is what a structure of trusted systems does for owners of intellectual property. It is a bonded courier that takes the thing of value and controls access to and use of it according to the orders given by the principal.

Imagine for a moment that such a structure emerged generally in cyberspace. How would we then think about copyright law?

An important point about copyright law is that, though designed in part to protect authors, the control it was designed to create was never to be perfect. As the Supreme Court noted, copyright “protection has never accorded the copyright owner complete control over all possible uses of his work.”¹⁸ Thus, the law grants only particular exclusive rights, and those rights are subject to important limitations, such as “fair use,” limited terms, and the first sale doctrine. The law threatened to punish violators of copyright laws—and it was this threat that induced a fairly high proportion of people to comply—but the law was never designed to simply do the author’s bidding. It had public purposes as well as the author’s interest in mind.

Trusted systems provide authors with the same sort of protection. Because authors can restrict unauthorized use of their material, they can extract money in exchange for access. Trusted systems thus achieve what copyright law aims to, but they can achieve this protection without the law doing the restricting. It permits a much more fine-grained control over access to and use of protected material than the law permits, and it can do so without the aid of the law.

What copyright seeks to do using the threat of law and the push of norms, trusted systems do through the code. Copyright orders others to respect the rights of the copyright holder before using his property; trusted systems give access only if rights are respected in the first place. The controls needed to regulate this access are built into the systems, and no users (except hackers) have a choice about whether to obey them. The code complements the law by codifying the rules, making them more efficient.

Trusted systems in this sense are a privatized alternative to copyright law. They need not be exclusive; there is no reason not to use both law and trusted systems. Nevertheless, the code is effectively doing the work that the law was designed to do. It implements the law’s protection, through code, far more effectively than the law did.

What could be wrong with this? We do not worry when people put double bolts on their doors to supplement the work of the neighborhood cop. We

do not worry when they lock their cars and take their keys. It is not an offense to protect yourself rather than rely on the state. Indeed, in some contexts it is a virtue. Andrew Jackson's mother, for example, told him, "Never tell a lie, nor take what is not your own, nor sue anybody for slander, assault and battery. Always settle them cases yourself."¹⁹ Self-sufficiency is strength and going to the law a sign of weakness.

There are two steps to answering this question. The first rehearses a familiar but forgotten point about the nature of "property"; the second makes a less familiar, but central, point about the nature of intellectual property. Together they suggest why perfect control is not the control that law has given owners of intellectual property. And together they suggest the potential problem that copyright law in cyberspace will create.

THE LIMITS ON THE PROTECTION OF PROPERTY

The realists in American legal history (circa 1890–1930) were scholars who (in part) emphasized the role of the state in what was called "private law."²⁰ At the time they wrote, it was the "private" in private law that got all the emphasis. Forgotten was the "law," as if "property" and "contract" existed independent of the state.

The realists' aim was to undermine this view. Contract and property law, they argued, gave private parties power.²¹ If you breach a contract with me, I can have the court order the sheriff to force you to pay; the contract gives me access to the state power of the sheriff. If your contract with your employer says that it may dismiss you for being late, then the police can be called in to eject you if you refuse to leave. If your lease forbids you to have cats, then the landlord can use the power of the courts to evict you if you do not get rid of the cats. These are all instances where contract and property, however grounded in private action, give a private person an entitlement to the state.

No doubt this power is justified in many cases; to call it "law" is not to call it unjust. The greatest prosperity in history has been created by a system in which private parties can order their lives freely through contract and property. But whether justified in the main or not, the realists argued that the contours of this "law" should be architected to benefit society.²²

This is not communism. It is not an attack on private property, and it is not to say that the state creates wealth (put your Ayn Rand away). These are claims about the relationship between private law and public law, and they should be uncontroversial.

Private law creates private rights to the extent that these private rights serve some collective good. If a private right is harmful to a collective good,

then the state has no reason to create it. The state's interests are general, not particular. It has a reason to create rights when those rights serve a common, rather than particular, end.

The institution of private property is an application of this point. The state has an interest in defining rights to private property because private property helps produce a general, and powerful, prosperity. It is a system for ordering economic relations that greatly benefits all members of society. No other system that we have yet devised better orders economic relations. No other system, some believe, could.²³

But even with ordinary property—your car, or your house—property rights are never absolute. There is no property that does not have to yield at some point to the interests of the state. Your land may be taken to build a highway, your car seized to carry an accident victim to the hospital, your driveway crossed by the postman, your house inspected by health inspectors. In countless ways, the system of property we call “private property” is a system that balances exclusive control by the individual against certain common state ends. When the latter conflict with the former, it is the former that yields.

This balance, the realists argued, is a feature of all property. But it is an especially important feature of intellectual property. The balance of rights with intellectual property differs from the balance with ordinary real or personal property. “Information,” as Boyle puts it, “is different.”²⁴ And a very obvious feature of intellectual property shows why.

When property law gives me the exclusive right to use my house, there's a very good reason for it. If you used my house while I did, I would have less to use. When the law gives me an exclusive right to my apple, that too makes sense. If you eat my apple, then I cannot. Your use of my property ordinarily interferes with my use of my property. Your consumption reduces mine.

The law has a good reason, then, to give me an exclusive right over my personal and real property. If it did not, I would have little reason to work to produce it. Or if I did work to produce it, I would then spend a great deal of my time trying to keep you away. It is better for everyone, the argument goes, if I have an exclusive right to my (rightly acquired) property, because then I have an incentive to produce it and not waste all my time trying to defend it.²⁵

Things are different with intellectual property. If you “take” my idea, I still have it. If I tell you an idea, you have not deprived me of it.²⁶ An unavoidable feature of intellectual property is that its consumption, as the economists like to put it, is “nonrivalrous.” Your consumption does not lessen mine. If I write a song, you can sing it without making it impossible for me to sing it. If I write a book, you can read a copy of it (please do) without disabling me from reading another copy of it. Ideas, at their core, can be shared with no reduction in

the amount the “owner” can consume. This difference is fundamental, and it has been understood since the founding.

Jefferson put it better than I:

If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possess the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lites his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density at any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation. Inventions then cannot, in nature, be a subject of property.²⁷

Technically, Jefferson presents two concepts: One is the possibility of excluding others from using or getting access to an idea, which he defines as “action of the thinking power . . . which an individual may exclusively possess as long as he keeps it to himself.” This is the question whether ideas are “excludable”; Jefferson affirms that an idea is “excludable” until “the moment it is divulged.”

The other concept is whether my use of a divulged idea lessens your use of the same idea. This is the question of whether divulged ideas are “rivalrous.”²⁸ Again, Jefferson suggests that, once they are divulged, ideas are not “rivalrous.” Jefferson believes that the act of divulging/sharing has made ideas both nonexcludable and nonrivalrous, and that there is little that man can do to change this fact.²⁹

In fact, shared ideas are both nonexcludable and nonrivalrous. I can exclude people from my secret ideas or writings—I can keep them secret, or build fences to keep people out. How easily, or how effectively, I can do so is a technical question. It depends on the architecture of protection that a given context provides. But given the proper technology, there is no doubt that I can keep people out. What I cannot do is to exclude people from my shared ideas or writings simply because they are not my secrets anymore.

My shared ideas are “nonrivalrous” goods, too. No technology (that we know of) will erase an idea from your head as it passes into my head. My

knowing what you know does not lessen your knowing the same thing. That fact is a given in the world, and it makes intellectual property different. Unlike apples, and unlike houses, once shared, ideas are something I can take from you without diminishing what you have.

It does not follow, however, that there is no need for property rights over expressions or inventions.³⁰ Just because you can have what I have without lessening what I have does not mean that the state has no reason to create rights over ideas, or over the expression of ideas.

If a novelist cannot stop you from copying (rather than buying) her book, then she may have very little incentive to produce more books. She may have as much as she had before you took the work she produced, but if you take it without paying, she has no monetary incentive to produce more.

Now, of course, the incentives an author faces are quite complex, and it is not possible to make simple generalizations.³¹ But generalizations do not have to be perfect to make a point: Even if some authors write for free, it is still the case that the law needs some intellectual property rights. If the law did not protect authorship at all, there would be fewer authors. The law has a reason to protect the rights of authors, at least insofar as doing so gives them an incentive to produce. With ordinary property, the law must both create an incentive to produce and protect the right of possession; with intellectual property, the law need only create the incentive to produce.

This is the difference between these two very different kinds of property, and this difference fundamentally affects the nature of intellectual property law. While we protect real and personal property to protect the owner from harm and give the owner an incentive, we protect intellectual property to ensure that we create a sufficient incentive to produce it. “Sufficient incentive,” however, is something less than “perfect control.” And in turn we can say that the ideal protections of intellectual property law are something less than the ideal protections for ordinary or real property.

This difference between the nature of intellectual property and ordinary property was recognized by our Constitution, which in article I, section 8, clause 8, gives Congress the power “to promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”

Note the special structure of this clause. First, it sets forth the precise reason for the power—to promote the progress of science and useful arts. It is for those reasons, and those reasons only, that Congress may grant an exclusive right. And second, note the special temporality of this right: “for limited Times.” The Constitution does not allow Congress to grant authors and inventors permanent exclusive rights to their writings and discoveries, only

limited rights. (Though apparently those limited times can be extended.³²) It does not give Congress the power to give them a perpetual “property” in their writings and discoveries, only an exclusive right over them for a limited time.

The Constitution’s protection for intellectual property then is fundamentally different from its protection of ordinary property. I’ve said that all property is granted subject to the limit of the public good. But even so, if the government decided to nationalize all property after a fifteen-year term of ownership, the Constitution would require it to compensate the owners. By contrast, if Congress set the copyright term at fifteen years, there would be no claim that the government pay compensation after the fifteen years were up. Intellectual property rights are a monopoly that the state gives to producers of intellectual property in exchange for their production of it. After a limited time, the product of their work becomes the public’s to use as it wants. This is Communism at the core of our Constitution’s protection of intellectual property. This “property” is not property in the ordinary sense of that term.

And this is true for reasons better than tradition as well. Economists have long understood that granting property rights over information is dangerous (to say the least).³³ This is not because of leftist leanings among economists; it is because economists are consequentialists, and their objective in granting any property right is simply to facilitate production. But there is no way to know, in principle, whether increasing or decreasing the rights granted under intellectual property law will lead to an increase in the production of intellectual property. The reasons are complex, but the point is not: Increasing intellectual property’s protection is not guaranteed to “promote the progress of science and useful arts”—indeed, often doing so will stifle it.

The balance that intellectual property law traditionally strikes is between the protections granted the author and the public use or access granted everyone else. The aim is to give the author sufficient incentive to produce. Built into the law of intellectual property are limits on the power of the author to control use of the ideas she has created.³⁴

A classic example of these limits and of this public use dimension is the right of “fair use.” Fair use is the right to use copyrighted material, regardless of the wishes of the owner of that material. A copyright gives the owner certain rights; fair use is a limitation on those rights. It gives you the right to criticize this book, cut sections from it, and reproduce them in an article attacking me. In these ways and in others, you have the right to use this book independent of how I say it should be used.

Fair use does not necessarily work against the author’s interest—or more accurately, fair use does not necessarily work against the interests of authors as a class. When fair use protects the right of reviewers to criticize books without

the permission of authors, then more critics criticize. And the more criticism there is, the better the information is about what books people should buy. The better the information is about what to buy, the more people will buy it. Authors as a whole benefit from the system of fair use, even if particular authors do not.

The law of copyright is filled with such rules. Another is the “first sale” doctrine. If you buy this book, you can sell it to someone else free of any constraint I might impose on you.³⁵ This doctrine differs from the tradition in, for example, Europe, where there are “moral rights” that give the creator power over subsequent use.³⁶ I’ve already mentioned another example—limited term. The creator cannot extend the term for which the law will provide protection (even if Congress can); that is fixed by the statute and runs out when the statute runs out.

Taken together, these rules give the creator significant—but not perfect—control over the use of what he produces. They give the public some access, but not complete access. They are balanced differently from the balance the law strikes for ordinary property—by design. They are constitutionally structured to help build an intellectual and cultural commons.

The law strikes this balance. It is not a balance that would exist in nature. Without the law, and before cyberspace, authors would have very little protection; with the law, they have significant, but not perfect, protection. The law gives authors something they otherwise would not have in exchange for limits on their rights, secured to benefit the intellectual commons as a whole.

PRIVATE SUBSTITUTES FOR PUBLIC LAW

So copyright law strikes a balance between control and access. What about that balance when code is the law? Should we expect that any of the limits will remain? Should we expect code to mirror the limits that the law imposes? Fair use? Limited term? Would private code build these “bugs” into its protections?

The point should be obvious: When intellectual property is protected by code, nothing requires that the same balance be struck. Nothing requires the owner to grant the right of fair use. She might allow individuals to browse for free, as a bookstore does, but she might not. Whether she grants this right depends on whether it profits her. Fair use becomes contingent upon private gain. More importantly, it becomes contingent upon the private gain of authors individually rather than authors as a class.

Thus, as privatized law, trusted systems regulate in the same domain that copyright law regulates. But unlike copyright law, they do not guarantee the

same limits on copyright's protection. Trusted systems give the producer maximum control over the uses of copyrighted work—admittedly at a cheaper cost, thus perhaps permitting many more authors to publish. But they give authors almost perfect control in an area in which the law did not. Code thus displaces the balance that copyright law strikes by displacing the limits the law imposes. As Daniel Benloliel puts it,

[D]ecentralized content providers are . . . privatizing the enforcement authority with strict technological standards, under which individuals would be banned from access and use of particular digital content in a way that might override legitimate fair use.³⁷

So far my description simply sets law against code: the law of copyright either complemented by, or in conflict with, private code. You may not yet be convinced that we should consider this a conflict, because it has always been the case that one can exercise more control over a copyrighted work than the law gives you the right to exercise over the copyright. For example, if you own a painting that is in the public domain, there's no requirement for you to let anyone see it. You could lock it in your bedroom and never let anyone see it ever. In a sense, you've thus deprived the world of the value of this painting being in the "public domain." But no one has ever thought that this interaction between the law of trespass and copyright has created any important conflict. So why should anyone be troubled if copyright owners use code to lock up their content beyond the balance the law of copyright strikes?

If this is where you're stuck, then let me add one more part to the story. As I mentioned above, the DMCA contains an anti-circumvention provision. That part of the law forbids the circumvention of some technical protection measures; it forbids the development of tools to circumvent technical protection as well. Most important, it forbids these circumventions regardless of the purpose of the circumvention. Thus, if the underlying use you would make of a copyrighted work—if you could get access to it—is a "fair use," the DMCA still makes it an offense to circumvent technical protections to get access to it. Thus one part of the law of copyright grants "fair use," while another part of copyright removes at least some fair use liberty where the fair use has been removed by technical means.³⁸

But so what, the skeptic will ask. What the law gives, the law can take away, can't it?

No it can't, and that's the point. As the Supreme Court has indicated, copyright law is consistent with the First Amendment only because of certain

important limitations built into the law. Removing those limitations would then raise important First Amendment questions. Thus, when the law acts with code to remove the law's protection for fair use, this should raise an important question—at least for those concerned about maintaining the balance that copyright law strikes.

But maybe this conflict is just temporary. Couldn't the code be changed to protect fair use?

The answer to that hopeful (and again, hopeful because my main point is about whether incentives to protect fair use exist) question is no, not directly. Fair use inherently requires a judgment about purpose, or intent. That judgment is beyond the ken of even the best computers. Indirectly, however, fair use could be protected. A system that allowed an individual to unlock the trusted system if he claimed the use was fair (perhaps marking the used work with a tag to make it possible to trace the use back to the user) could protect fair use. Or as Stefik describes, a system that granted users a "fair use license," allowing them to unlock the content and use insurance backing the license to pay for any misuse, might also protect fair use.³⁹ But these alternatives again rely on structures beyond code. With the code itself, there is no way adequately to police fair use.

Some will respond that I am late to the party: Copyright law is already being displaced, if not by code then by the private law of contract. Through the use of click-wrap, or shrink-wrap, licenses, authors are increasingly demanding that purchasers, or licensees, waive rights that copyright law gave them. If copyright law gives the right to reverse-engineer, then these contracts might extract a promise not to reverse-engineer. If copyright law gives the right to dispose of the book however the purchaser wants after the first sale, then a contract might require that the user waive that right. And if these terms in the contract attached to every copyright work are enforceable merely by being "attached" and "knowable," then already we have the ability through contract law to rewrite the balance that copyright law creates.

I agree that this race to privatize copyright law through contract is already far along, fueled in particular by decisions such as Judge Frank Easterbrook's in *ProCD v. Zeidenberg*. But contracts are not as bad as code. Contracts are a form of law. If a term of a contract is inconsistent with a value of copyright law, you can refuse to obey it and let the other side get a court to enforce it. In some cases, courts have expressly refused to follow a contract term precisely because it is inconsistent with a copyright law value.⁴⁰ The ultimate power of a contract depends upon the decision by a court to enforce the contract or not. Although courts today are relatively eager to find ways to enforce these contracts, there is at least hope that if the other side makes its case very clear,

courts could shift direction again.⁴¹ As Stefik writes, trusted systems “differ from an ordinary contract in critical ways.”

[I]n an ordinary contract, compliance is not automatic; it is the responsibility of the agreeing parties. There may be provisions for monitoring and checking on compliance, but the actual responsibility for acting in accordance with the terms falls on the parties. In addition, enforcement of the contract is ultimately the province of the courts.⁴²

The same is not true of code. Whatever problems there are when contracts replace copyright law, the problems are worse when code displaces copyright law. Again—where do we challenge the code? When the software protects without relying in the end on the state, where can we challenge the nature of the protection? Where can we demand balance when the code takes it away?

I don’t mean to enter the extremely contentious debate about whether this change in control is good or appropriate. I’ve said too much about that elsewhere.⁴³ For our purposes here, the point is simply to recognize a significant change. Code now makes possible increasingly perfect control over how culture is spread. Regulations have “been fairly consistent . . . on the side of expanding the power of the owners to control the use of their products.”⁴⁴ And these regulations invite a demand for perfect control over how culture is spread.

The rise of contracts qualifying copyright law and the rise of code qualifying copyright law raise a question that the law of copyright has not had to answer before. We have never had to choose whether authors should be permitted perfectly to control the use of their intellectual property independent of the law, for such control was not possible. The balance struck by the law was the best that authors could get. But now, code gives authors a better deal. The question for legal policy is whether this better deal makes public sense.

Here we confront the first latent ambiguity within the law of copyright. There are those who would say that copyright law already decides this question—whether against code-based control, or for it. But in my view, this is a choice the law has yet to make. I have my own views about how the law should decide the question. But what technology has done is force us to see a choice that was not made before. See the choice, and then make it.

Put most directly: There has always been a set of uses of copyrighted work that was unregulated by the law of copyright. Even within the boundary of uses that were regulated by the law of copyright, “fair use” kept some uses free. The core question is why? Were these transactions left free because it

was too costly to meter them? Or were these transactions left free because keeping them free was an important public value tied to copyright?

This is a question the law never had to resolve, though there is support for both views.⁴⁵ Now the technology forces us to resolve it. The question, then, is how.

A nice parallel to this problem exists in one part of constitutional law. The framers gave Congress the power to regulate interstate commerce and commerce that affects interstate commerce.⁴⁶ At the founding, that was a lot of commerce, but because of the inefficiencies of the market, not all of it. Thus, the states had a domain of commerce that they alone could regulate.⁴⁷

Over time, however, the scope of interstate commerce has changed so that much less commerce is now within the exclusive domain of the states. This change has produced two sorts of responses. One is to find other ways to give states domains of exclusive regulatory authority. The justification for this response is the claim that these changes in interstate commerce are destroying the framers' vision about state power.

The other response is to concede the increasing scope of federal authority, but to deny that it is inconsistent with the framing balance.⁴⁸ Certainly, at the founding, some commerce was not interstate and did not affect interstate commerce. But that does not mean that the framers intended that there must always be such a space. They tied the scope of federal power to a moving target; if the target moves completely to the side of federal power, then that is what we should embrace.⁴⁹

In both contexts, the change is the same. We start in a place where balance is given to us by the mix of frictions within a particular regulatory domain: Fair use is a balance given to us because it is too expensive to meter all use; state power over commerce is given to us because not all commerce affects interstate commerce. When new technology disturbs the balance, we must decide whether the original intent was that there be a balance, or that the scope of one side of each balance should faithfully track the index to which it was originally tied. Both contexts, in short, present ambiguity.

Many observers (myself included) have strong feelings one way or the other. We believe this latent ambiguity is not an ambiguity at all. In the context of federal power, we believe either that the states were meant to keep a domain of exclusive authority⁵⁰ or that the federal government was to have whatever power affected interstate commerce.⁵¹ In the context of fair use, we believe that either fair use is to be a minimum of public use, guaranteed regardless of the technology,⁵² or that it is just an efficient compromise in response to an inefficient technology, to be removed as soon as efficiency can be achieved.

But in both cases, this may make the problem too easy. The best answer in both contexts may be that the question was unresolved at the framing: Perhaps no one thought of the matter, and hence there is no answer to the question of what they would have intended if some central presupposition had changed. And if there was no original answer, we must decide the question by our own lights. As Stefik says of trusted systems—and, we might expect, of the implications of trusted systems—“It is a tool never imagined by the creators of copyright law, or by those who believe laws governing intellectual property cannot be enforced.”⁵³

The loss of fair use is a consequence of the perfection of trusted systems. Whether you consider it a problem or not depends on your view of the value of fair use. If you consider it a public value that should exist regardless of the technological regime, then the emergence of this perfection should trouble you. From your perspective, there was a value latent in the imperfection of the old system that has now been erased.

But even if you do not think that the loss of fair use is a problem, trusted systems threaten other values latent in the imperfection of the real world. Consider a second.

THE ANONYMITY THAT IMPERFECTION ALLOWS

I was a student at an English university for a number of years. In the college I attended, there was a “buttery”—a shop inside the college that basically sold alcohol. During the first week I was there I had to buy a large amount of Scotch (a series of unimaginative gifts, as I remember). About a week after I made these purchases, I received a summons from my tutor to come talk with him in his office. When I arrived, the tutor asked me about my purchases. This was, to his mind, an excessive amount of alcohol, and he wanted to know whether I had a good reason for buying it.

Needless to say, I was shocked at the question. Of course, technically, I had made a purchase at the college, and I had not hidden my name when I did so (indeed, I had charged it on my college account), so, formally, I had revealed my alcohol purchases to the college and its agents. Still, it shocked me that this information would be monitored by college authorities and then checked up on. I could see why they did it, and I could see the good that might come from it. It just never would have occurred to me that these data would be used in this way.

If this was an invasion, of course, it was a small one. Later it was easy for me to hide my binges simply by buying from a local store rather than the college buttery. (Though I later learned that the local store rented its space from the college, so who knows what deal they had struck?) And in any case,

I was not being punished. The college was just concerned. But the example suggests a more general point: We reveal to the world a certain class of data about ourselves that we ordinarily expect the world not to use. What happens when they use it?

Trusted systems depend on such data—they depend on the ability to know how people use the property that is being protected. To set prices most efficiently, the system ideally should know as much about individuals and their reading habits as possible. It needs to know this data because it needs an efficient way to track use and so to charge for it.⁵⁴

But this tracking involves a certain invasion. We live now in a world where we think about what we read in just the way that I thought about what I bought as a student in England—we do not expect that anyone is keeping track. We would be shocked if we learned that the library was keeping tabs on the books that people checked out and then using this data in some monitoring way.

Such tracking, however, is just what trusted systems require. And so the question becomes: Should there be a right against this kind of monitoring? The question is parallel to the question of fair use. In a world where this monitoring could not effectively occur, there was, of course, no such right against it. But now that monitoring can occur, we must ask whether the latent right to read anonymously, given to us before by imperfections in technologies, should be a legally protected right.

Julie Cohen argues that it should, and we can see quite directly how her argument proceeds.⁵⁵ Whatever its source, it is a value in this world that we can explore intellectually on our own. It is a value that we can read anonymously, without fear that others will know or watch or change their behavior based on what we read. This is an element of intellectual freedom; it is a part of what makes us as we are.⁵⁶

But this element is potentially erased by trusted systems. These systems need to monitor, and this monitoring destroys anonymity. We need to decide whether, and how, to preserve values from today in a context of trusted systems.

This could first be a question of translation: namely, how should changes in technology be accommodated to preserve values from an earlier context in a new context? It is the same question that Brandeis asked about wiretapping.⁵⁷ It is the question the Court answers in scores of contexts all the time. It is fundamentally a question about preserving values when contexts change.

In the context of both fair use and reading, Cohen has a consistent answer to this question of translation. She argues that there is a right to resist, or “hack,” trusted systems to the extent that they infringe on traditional fair use. (Others have called this the “Cohen Theorem.”) As for reading, she argues that

copyright management schemes must protect a right to read anonymously—that if they monitor, they must be constructed so that they preserve anonymity. The strategy is the same: Cohen identifies a value yielded by an old architecture but now threatened by a new architecture, and then argues in favor of an affirmative right to protect the original value.

But here again we might view the question more ambiguously. I share Cohen’s view, but the argument on the other side is not silly. If it’s permissible to use technology to make copyrighted works available, why isn’t it permissible to gather data about who uses what works? That data gathering is not part of the copyright itself; it is a byproduct of the technology. And as our tradition has never had this technical capacity before, it is hard to say a choice was made about it in the past.

PERMISSION CULTURE VS. FREE

I’ve already described the limits copyright law places on itself. These limits, as I argued, reflect important values. They express the balance that copyright law aims to be.

But what is too often missed in this discussion of balance is any sense of perspective. We focus on the gradual shifts in the law but miss the profound sense in which the significance of the law has changed.

This change is produced by the unintended interaction between the architecture of digital technologies and the architecture of the law.

Copyright law at its core regulates “copies.” In the analog world, there were very few contexts in which one produced “copies.” As Jessica Litman described more than a decade ago,

At the turn of the century, U.S. copyright law was technical, inconsistent, and difficult to understand, but it didn’t apply to very many people or very many things. If one were an author or publisher of books, maps, charts, paintings, sculpture, photographs or sheet music, a playwright or producer of plays, or a printer, the copyright law bore on one’s business. Booksellers, piano-roll and phonograph record publishers, motion picture producers, musicians, scholars, members of Congress, and ordinary consumers could go about their business without ever encountering a copyright problem.⁵⁸

Thus there were many ways in which you could use creative work in the analog world without producing a copy.

Digital technology, at its core, makes copies. Copies are to digital life as breathing is to our physical life. There is no way to use any content in a digital

context without that use producing a copy. When you read a book stored on your computer, you make a copy (at least in the RAM memory to page through the book). When you do anything with digital content, you technically produce a copy.

This technical fact about digital technologies, tied to the technical architecture of the law, produces a profound shift in the scope or reach of the law of copyright that too many simply miss: While in the analog world, life was sans copyright law; in the digital world, life is subject to copyright law. Every single act triggers the law of copyright. Every single use is either subject to a license or illegal, unless deemed to be “fair use.” The emergence of digital technologies has thus radically increased the domain of copyright law—from regulating a tiny portion of human life, to regulating absolutely every bit of life on a computer.

Now if all you think about is protecting the distribution of professionally created culture, this might not concern you much. If you’re trying to stop “piracy,” then a regime that says every use requires permission is a regime that gives you a fairly broad range of tools for stamping out piracy.

But though you wouldn’t notice this listening to the debates surrounding copyright law just now, in fact, protecting the distribution of professionally created culture is not the only, or even, I suggest, the most important part of culture. And indeed, from a historical perspective, top-down, professionally produced culture is but a tiny part of what makes any culture sing. The 20th century may have been an exception to this rule, but no Congress voted to make professional culture the only legal culture within our society.

Standing alongside professional culture is amateur culture—where amateur doesn’t mean inferior or without talent, but instead culture created by people who produce not for the money, but for the love of what they do. From this perspective, there is amateur culture everywhere—from your dinner table, where your father or sister tell jokes that take off from the latest political scandal or the latest *Daily Show*; from your basement, where your brother and his three best friends are causing permanent damage to their eardrums as they try to become the next Rolling Stones; from your neighbors who gather each Thursday and Sunday to sing in a church choir; from your neighborhood schools, where kids and teachers create art or music in the course of learning about our culture; from the kids at your neighborhood school, who tear their pants or wear their shirts in some odd way, all as a way to express and make culture.

This amateur culture has always been with us, even if it is to us today, as Dan Hunter and Greg Lastowska put it, “hidden.”⁵⁹ It is precisely how the

imagination of kids develops;⁶⁰ it is how culture has always developed. As Siva Vaidhyanathan writes,

widespread democratic cultural production (peer-to-peer production, one might say) . . . merely echoes how cultural texts have flowed through and been revised by discursive communities everywhere for centuries. Texts often undergo a process similar to a game of “telephone,” through which a text is substantially—sometimes almost unintentionally—distorted through many small revisions. . . . Such radical textual revisions have occurred in other contexts and have helped build political critiques, if not movements. For instance, historian Lawrence Levine (1988) has documented how working-class players and audiences in nineteenth-century America adapted and revised the works of William Shakespeare to their local contexts, concerns and ideologies. And historian Eric Lott (1993) has shown how *Uncle Tom’s Cabin* was reworked by working-class white communities to aid the cause of racial dominance instead of the Christian liberationist message the book was intended to serve.⁶¹

Importantly, too, this kind of cultural remix has historically been free of regulation. No one would think that as you tell a joke around your dinner table, or sing songs with your friends, or practice to become the next Rolling Stones, you need a lawyer standing next to you, clearing the rights to “use” the culture as you make your creative remix. The law of copyright, historically, has been focused on commercial life. It has left the noncommercial, or beyond commercial, creativity free of legal regulation.

All this has now changed, and digital technologies are responsible. First, and most important, digital technologies have radically expanded the scope of this amateur culture. Now the clever remix of some political event or the latest song by your favorite band are not just something you can share with your friends. Digital technologies have made it simple to capture and share this creativity with the world. The single most important difference between the Internet circa 1999 and the Internet circa today is the explosion of user-generated creativity—from blogs, to podcasts, to videocasts, to mashups, the Internet today is a space of extraordinary creativity.

Second, digital technologies have democratized creativity. Technology has given a wide range of potential creators the capacity to become real. “People are waking from their consumerist coma,” one commentator describes.⁶² As DJ Danger Mouse put it at the Web 2.0 conference in 2004,

Mashing is so easy. It takes years to learn how to play the guitar and write your own songs. It takes a few weeks of practice with a turntable to make people

dance and smile. It takes a few hours to crank out something good with some software. So with such a low barrier to entry, everyone jumps in and starts immediately being creative.⁶³

But third, and directly relevant to the story of this chapter, to the extent this creativity finds its expression on the Net, it is now subject to the regulation of copyright law. To the extent it uses others' creativity, it needs the permission of others. To the extent it builds upon the creativity of others, it needs to be sure that that creativity can be built upon legally. A whole system of regulation has now been grafted upon an economy of creativity that until now has never known regulation. Amateur culture, or bottom up culture, or the culture that lives outside of commercial transactions—all of this is subject to regulation in a way that 30 years ago it was not.

A recent example of this conflict makes the point very concisely. There's a genre of digital creativity called Anime Music Videos (AMVs). AMVs are remixes of anime cartoons and music. Kids spend hundreds, sometimes thousands of hours reediting the anime cartoons to match them perfectly to music. The result is, in a word, extraordinary. It is among the most creative uses of digital technology that I have seen.

While this genre of creativity is not small, it's also not huge. Basically one site dominates activity around AMVs. That site has more than 500,000 members, and some 30,000 creators upload AMV content to the site.

In November 2005, one prominent record label, Wind-Up Records, informed this website that it wanted all Wind-Up Records artists removed from the site. That was some 3,000 videos, representing at least 250,000 hours of volunteer work by creators across the world—work that would have just one real effect: to promote the underlying artists' work.

From the perspective of the law as it is, this is an easy case. What the kids are doing is making a derivative work of the anime; they are distributing full copies of the underlying music; and they are synchronizing the music to video—all without the permission of the copyright owners.

But from the perspective of culture, this should be a very hard case. The creativity demonstrated by this work is extraordinary. I can't show you that creativity in a book, but the notes point you to an example that you can see.⁶⁴ It is noncommercial, amateur creative work—precisely the sort that has never been subject to the regulation of the law, but which now, because it is living in digital context, is monitored, and regulated, by the law.

Here again, I have strong feelings about what the right answer should be. But we should recognize the latent ambiguity this conflict presents:

Because of the changes in digital technology, it is now possible for the law to regulate every single use of creative work in a digital environment. As life increasingly moves into a digital environment, this means that the law will regulate more and more of the use of culture.

Is this consistent with our values?

The answer again could be found first by trying to translate framing values into the current context. From that perspective, it would be extraordinarily difficult to imagine that the framing vision would have included the level of legal regulation that the current regime entails.

Again, that conclusion could be questioned by recognizing that the possibility of such extensive regulation didn't exist, and so the choice about whether such extensive regulation should be allowed wasn't made. That choice, when made, should recognize that while there is extensive and new regulation of amateur culture, that regulation creates new wealth for professional culture. There's a choice to be made about which form of culture we should protect. That choice has not yet been made directly. It is one more choice we have yet to make.

THE PROBLEMS THAT PERFECTION MAKES

These three examples reveal a common pattern—one that will reach far beyond copyright. At one time we enjoyed a certain kind of liberty. But that liberty was not directly chosen; it was a liberty resulting from the high costs of control.⁶⁵ That was the conclusion we drew about fair use—that when the cost of control was high, the space for fair use was great. So too with anonymous reading: We read anonymously in real space not so much because laws protect that right as because the cost of tracking what we read is so great. And it was the same with amateur culture: That flourished free of regulation because regulation could not easily reach it.

When costs of control fall, however, liberty is threatened. That threat requires a choice—do we allow the erosion of an earlier liberty, or do we erect other limits to re-create that original liberty?

The law of intellectual property is the first example of this general point. As the architecture of the Internet changes, it will allow for a greater protection of intellectual property than real-space architectures allowed; this greater protection will force a choice on us that we do not need to make in real space. Should the architecture allow perfect control over intellectual property, or should we build into the architecture an incompleteness that guarantees a certain aspect of public use or a certain space for individual freedom?

Ignoring these questions will not make them go away. Pretending that the framers answered them is no solution either. In this context (and this is just the first) we will need to make a judgment about which values the architecture will protect.

CHOICES

I've argued that cyberspace will open up three important choices in the context of intellectual property: whether to allow intellectual property in effect to become completely propertized (for that is what a perfect code regime for protecting intellectual property would do); and whether to allow this regime to erase the anonymity latent in less efficient architectures of control; and whether to allow the expansion of intellectual property to drive out amateur culture. These choices were not made by our framers. They are for us to make now.

I have a view, in this context as in the following three, about how we should exercise that choice. But I am a lawyer. Lawyers are taught to point elsewhere—to the framers, to the United Nations charter, to an act of Congress—when arguing about how things ought to be. Having said that there is no such authority here, I feel as if I ought to be silent.

Cowardly, not silent, however, is how others might see it. They say that I should say what I think. So in each of these three applications (intellectual property, privacy, and free speech), I will offer my view about how these choices should be made. But I do this under some duress and encourage you to simply ignore what I believe. It will be short, and summary, and easy to discard. It is the balance of the book—and, most importantly, the claim that we have a choice to make—that I really want to stick.

Anonymity

Cohen, it seems to me, is plainly right about anonymity, and the Cohen Theorem is inspirational. However efficient the alternative may be, we should certainly architect cyberspaces to ensure anonymity—or more precisely, pseudonymity—first. If the code is going to monitor what I do, then at least it should not know that it is “I” that it is monitoring. I am less troubled if it knows that “14AH342BD7” read such and such; I am deeply troubled if that number is tied back to my name.

Cohen is right for a second reason as well: All of the good that comes from monitoring could be achieved while protecting privacy. It may take a bit more coding to build in routines for breaking traceability; it may take more planning to ensure that privacy is protected. But if those rules are embedded

up front, the cost would not be terribly high. It is far cheaper to architect privacy protections now rather than retrofit for them later.

The Commons

By “the Commons” I mean a resource that anyone within a relevant community can use without seeking the permission of anyone else. Such permission may not be required because the resource is not subject to any legal control (it is, in other words, in the public domain). Or it may not be required because permission to use the resource has already been granted. In either case, to use or to build upon this resource requires nothing more than access to the resource itself.⁶⁶

In this sense, the questions about the scope and reach of copyright law ask whether our future will protect the intellectual commons that it did in the past. Again, it did so in the past because the friction of control was too great. But now that that friction is gone, will we preserve or destroy the commons that used to exist?

My view is that it ought to be preserved.

We can architect cyberspace to preserve a commons or not. (Jefferson thought that nature had already done the architecting, but Jefferson wrote before there was code.) We should choose to architect it with a commons. Our past had a commons that could not be designed away; that commons gave our culture great value. What value the commons of the future could bring us is something we are just beginning to see. Intellectual property scholars saw it—long before cyberspace came along—and laid the groundwork for much of the argument we need to have now.⁶⁷ The greatest work in the law of cyberspace has been written in the field of intellectual property. In a wide range of contexts, these scholars have made a powerful case for the substantive value of an intellectual commons.⁶⁸

James Boyle puts the case most dramatically in his extraordinary book *Shamans, Software, and Spleens*.⁶⁹ Drawing together both cyberspace and non-cyberspace questions, he spells out the challenge we face in an information society—particularly the political challenge.⁷⁰ Elsewhere he identifies our need for an “environmental movement” in information policy—a rhetoric that gets people to see the broad range of values put at risk by this movement to propertize all information. Boyle’s work has inspired many others to push a similar agenda of freedom.⁷¹

That freedom would limit the law’s regulation over the use and reuse of culture. It would resist perfect control over use; it would free a wide range of reuse. It would build through affirmative protections for freedom the liberty

that friction gave us before. It would do so because it believes in the values this freedom stands for, and it would demonstrate the value in that freedom by enabling the communities that freedom would itself enable.

But this freedom could be constructed either through changes in the law or voluntarily. That is, the law could be rebalanced to encourage the freedom thought important, or this property could be redeployed to effect the freedom thought important.

The second strategy was the technique of the Free Software Movement, described in Chapter 8. Using copyright law, Stallman deployed a software license that both preserved the four freedoms of free software, and also required that those modifying and distributing free software distribute the modifications freely. This license thus effects a software commons, since the software is available to all to use, and this software commons has become a critical raw material fueling the digital age.

More recently, Stallman's idea has been copied by others seeking to rebuild a commons in cyberspace. The Wikipedia project, for example, has built—to the astonishment of most—an extraordinary online encyclopedia solely through the volunteer efforts of thousands, contributing essays and edits in a public wiki. The product of that work is now protected perpetually (yes, I know, only for a “limited time,” but don't correct *me* about that little detail) through a copyright license that, like the GPL, requires any modification to be distributed freely as well. (More on Wikipedia in Chapter 12.)

And so too has Creative Commons used private law to build an effective public commons. Again, following Stallman, Creative Commons offers copyright holders a simple way to mark their creative work with the freedoms they intend it to carry. That mark is a license which reserves to the author some rights, while dedicating to the public rights that otherwise would have been held privately. As these licenses are nonexclusive and public, they too effectively build a commons of creative resources that anyone can build upon.

Though I have spent a great deal of my time helping to build the Creative Commons, I still believe private action alone is not enough. Yet there is value in learning something from what this private action produces, as its lesson may help policy makers recraft copyright law in the future.

E L E V E N

p r i v a c y

THE CONCLUSION OF PART 1 WAS THAT CODE COULD ENABLE A MORE REGULABLE cyberspace; the conclusion of Part 2 was that code would become an increasingly important regulator in that more regulable space. Both conclusions were central to the story of the previous chapter. Contrary to the early panic by copyright holders, the Internet will become a space where intellectual property can be more easily protected. As I've described, that protection will be effected through code.

Privacy is a surprisingly similar story. Indeed, as Jonathan Zittrain argued in an essay published in the *Stanford Law Review*,¹ the problems of privacy and copyright are exactly the same. With both, there's a bit of "our" data that "we've" lost control over. In the case of copyright, it is the data constituting a copy of our copyrighted work; in the case of privacy, it is the data representing some fact about us. In both cases, the Internet has produced this loss of control: with copyright, because the technology enables perfect and free copies of content; with privacy, as we'll see in this chapter, because the technology enables perpetual and cheap monitoring of behavior. In both cases, the question policy makers should ask is what mix of law and technology might restore the proper level of control. That level must balance private and public interests: With copyright, the balance is as I described in the last chapter; with privacy, it is as we'll explore in this chapter.

The big difference between copyright and privacy, however, is the political economy that seeks a solution to each problem. With copyright, the interests threatened are powerful and well organized; with privacy, the interests threatened are diffuse and disorganized. With copyright, the values on the other side of protection (the commons, or the public domain) are neither compelling nor well understood. With privacy, the values on the other side of protection

(security, the war against terrorism) *are* compelling and well understood. The result of these differences, as any political theorist would then predict, is that over the past ten years, while we've seen a lot of legislative and technical changes to solve the problems facing copyright, we've seen very few that would solve the problems of privacy.

Yet as with copyright, we could restrike the balance protecting privacy. There are both changes in law and changes in technology that could produce a much more private (and secure) digital environment. Whether we will realize these changes depends upon recognizing both the dynamics to regulation in cyberspace and the importance of the value that privacy is.

We will think about three aspects of privacy, and how cyberspace has changed each of them. Two of these three will be the focus of this chapter, but I begin with the third to help orient the balance.

PRIVACY IN PRIVATE

The traditional question of “privacy” was the limit the law placed upon the ability of others to penetrate your private space. What right does the government have to enter your home, or search your papers? What protection does the law of trespass provide against others beyond the government snooping into your private stuff? This is one meaning of Brandeis's slogan, “the right to be left alone.”² From the perspective of the law, it is the set of legal restrictions on the power of others to invade a protected space.

Those legal restrictions were complemented by physical barriers. The law of trespass may well say it's illegal to enter my house at night, but that doesn't mean I won't lock my doors or bolt my windows. Here again, the protection one enjoys is the sum of the protections provided by the four modalities of regulation. Law supplements the protections of technology, the protections built into norms, and the protections from the costliness of illegal penetration.

Digital technologies have changed these protections. The cost of parabolic microphone technology has dropped dramatically; that means it's easier for me to listen to your conversation through your window. On the other hand, the cost of security technologies to monitor intrusion has also fallen dramatically. The net of these changes is difficult to reckon, but the core value is not rendered ambiguous by this difficulty. The expectation of privacy in what is reasonably understood to be “private” spaces remains unchallenged by new technologies. This sort of privacy doesn't present a “latent ambiguity.”

PRIVACY IN PUBLIC: SURVEILLANCE

A second kind of privacy will seem at first oxymoronic—privacy in public. What kind of protection is there against gathering data about me while I'm on a public street, or boarding an airplane?

The traditional answer was simple: None. By stepping into the public, you relinquished any rights to hide or control what others came to know about you. The facts that you transmitted about yourself were as “free as the air to common use.”³ The law provided no legal protection against the use of data gathered in public contexts.

But as we've seen again and again, just because the law of privacy didn't protect you it doesn't follow that you weren't protected. Facts about you while you are in public, even if not legally protected, are effectively protected by the high cost of gathering or using those facts. Friction is thus privacy's best friend.

To see the protection that this friction creates, however, we must distinguish between two dimensions along which privacy might be compromised.

There is a part of anyone's life that is *monitored*, and there is a part that can be *searched*. The monitored is that part of one's daily existence that others see or notice and can respond to, if response is appropriate. As I walk down the street, my behavior is monitored. If I walked down the street in a small village in western China, my behavior would be monitored quite extensively. This monitoring in both cases would be transitory. People would notice, for example, if I were walking with an elephant or in a dress, but if there were nothing special about my walk, if I simply blended into the crowd, then I might be noticed for the moment but forgotten soon after—more quickly in San Francisco, perhaps, than in China.

The *searchable* is the part of your life that leaves, or is, a record. Scribbles in your diary are a record of your thoughts. Stuff in your house is a record of what you possess. The recordings on your telephone answering machine are a record of who called and what they said. Your hard drive is you. These parts of your life are not ephemeral. They instead remain to be reviewed—at least if technology and the law permit.

These two dimensions can interact, depending upon the technology in each. My every action in a small village may be monitored by my neighbors. That monitoring produces a record—in their memories. But given the nature of the recording technology, it is fairly costly for the government to search that record. Police officers need to poll the neighbors; they need to triangulate on the inevitably incomplete accounts to figure out what parts are true, and what parts are not. That's a familiar process, but it has its limits. It might be easy to

poll the neighbors to learn information to help locate a lost person, but if the government asked questions about the political views of a neighbor, we might expect (hope?) there would be resistance to that. Thus, in principle, the data are there. In practice, they are costly to extract.

Digital technologies change this balance—radically. They not only make more behavior monitorable; they also make more behavior searchable. The same technologies that gather data now gather it in a way that makes it searchable. Thus, increasingly life becomes a village composed of parallel processors, accessible at any time to reconstruct events or track behavior.

Consider some familiar examples:

The Internet

In Part I, I described the anonymity the Internet originally provided. But let's be clear about something important: That relative anonymity of the "old days" is now effectively gone. Everywhere you go on the Internet, the fact that IP address xxx.xxx.xxx.xxx went there is recorded. Everywhere you go where you've allowed a cookie to be deposited, the fact that the machine carrying that cookie went there is recorded—as well as all the data associated with that cookie. They know you from your mouse droppings. And as businesses and advertisers work more closely together, the span of data that can be aggregated about you becomes endless.

Consider a hypothetical that is completely technically possible under the existing architectures of the Net. You go to a web page of a company you trust, and you give that company every bit of your private data—your name, address, social security number, favorite magazines and TV shows, etc. That company gives you a cookie. You then go to another site, one you don't trust. You decide not to give that site any personal data. But there's no way for you to know whether these companies are cooperating about the data they collect. It's perfectly possible they synchronize the cookies data they create. And thus, there's no technical reason why once you've given your data once, it isn't known by a wide range of sites that you visit.

In the section that follows, we'll consider more extensively how we should think about privacy in any data I've affirmatively provided others, such as my name, address, or social security number. But for the moment, just focus upon the identity data they've collected as I move around in "public." Unless you've taken extraordinary steps—installing privacy software on your computer, or disabling cookies, etc.—there's no reason you should expect that the fact that you visited certain sites, or ran certain searches, isn't knowable by someone. It is. The layers of technology designed

to identify “the customer” have produced endless layers of data that can be traced back to you.

Searches

In January 2006, Google surprised the government by doing what no other search company had done: It told the government “no.” The Justice Department had launched a study of pornography on the Net as a way to defend Congress’s latest regulation of pornography. It thus wanted data about how often, and in what form, people search for porn on the Internet. It asked Google to provide 1,000,000 random searches from its database over a specified period. Google—unlike Yahoo! and MSN—refused.

I suspect that when most first heard about this, they asked themselves an obvious question—Google keeps search requests? It does. Curiosity is monitored, producing a searchable database of the curious. As a way to figure out better how to do its job, Google—and every other search engine⁴—keeps a copy of every search it’s asked to make. More disturbingly, Google links that search to a specific IP address, and, if possible, to a Google users’ account. Thus, in the bowels of Google’s database, there is a list of all searches made by you when you were logged into your gmail account, sitting, waiting, for someone to ask to see it.

The government did ask. And in the normal course of things, the government’s request would be totally ordinary. It is unquestioned that the government gets to ask those with relevant evidence to provide it for an ongoing civil or criminal investigation (there are limits, but none really significant). Google has evidence; the government would ordinarily have the right to get it.

Moreover, the government in this case explicitly promised it would not use this evidence for anything more than evaluating patterns of consumption around porn. In particular, it promised it wouldn’t trace any particularly suspicious searches. It would ignore that evidence—which ordinarily it would be free to use for whatever purpose it chose—just so it could get access to aggregate data about searches for porn.

So what’s the problem this example illustrates?

Before search engines, no one had any records of curiosity; there was no list of questions asked. Now there is. People obsessively pepper search engines with questions about everything. The vast majority of these are totally benign (“mushrooms AND ragout”). Some of them show something less benign about the searcher (“erotic pictures AND children”). Now there’s a list of all these questions, with some providing evidence of at least criminal intent.

The government's interest in that list will increase. At first, its demands will seem quite harmless—so what if it counts the number of times people ask Google to point them to erotic pictures? Then, when not so harmless, the demands will link to very harmful behavior—searches that suggest terrorism, or abuse. Who could argue against revealing that? Finally, when not so harmless, and when the crime is not so harmful, the demands will simply insist this is an efficient way to enforce the law. “If you don't like the law, change it. But until you do, let us enforce it.” The progression is obvious, inevitable, and irresistible.

E-mail

Electronic mail is a text-based message stored in digital form. It is like a transcribed telephone call. When sent from one person to another, e-mail is copied and transmitted from machine to machine; it sits on these different machines until removed either by routines—decisions by machines—or by people.

The content of many e-mail messages is like the content of an ordinary telephone call—unplanned, unthinking, the ordinary chatter of friends. But unlike a telephone call, this content is saved in a searchable form. Companies now invest millions in technologies that scan the conversations of employees that before were effectively private. Both in real time and in retrospect, the content of conversations can become known. On the theory that they “own the computer,”⁵ employers increasingly snoop in the e-mail of employees, looking for stuff they deem improper.⁶

In principle, such monitoring and searching are possible with telephone calls or letters. In practice, these communications are not monitored. To monitor telephones or regular mail requires time and money—that is, human intervention. And this cost means that most won't do it. Here again, the costs of control yield a certain kind of freedom.

Controlling employees (or spouses) is one important new use of e-mail technologies. Another is the better delivery of advertisement. Google is again the leader here with its new Gmail service. Gmail can advertise to you as you read your e-mail. But the advance is that the advertisement is triggered by the content of the e-mail. Imagine a television that shifted its advertisement as it heard what you were talking about on the phone. The content of the e-mail—and perhaps the content of your inbox generally—helps determine what is shown to you.

To make this system work well, Google needs you to keep lots of data on its servers. Thus the only thing within Gmail that is difficult to do—and it is really really difficult—is to delete content from a Google Gmail account.

Gmail lets you delete one screen at a time. But when you have 20,000 e-mails in your inbox, who has time? Would it be difficult for Gmail to enable a “delete all” function? Of course not. This is Google! Thus, through the clever use of architecture, Google assures more data is kept, and that data then becomes a resource for other purposes. If you ever get involved in a lawsuit, the first question of the lawyer from the other side should be—do you have a Gmail account? Because, if you do, your life sits open for review.

V-mail

If e-mail becomes a permanent record, why not v-mail? Voice mail systems archive messages and record the communication attributes of the conversations. As technologies for voice recognition improve, so does the ability to search voice records. As voice mail systems shift to digital systems, archiving content on central servers rather than \$50 devices connected to the phone at home, they become practical search resources. In principle, every night the government could scan all the stored voice recordings at every telephone company in the nation. This search would impose no burden on the user; it could be targeted on and limited to specific topics, and it could operate in the background without anyone ever knowing.

Voice

And why stop with recordings? According to one report, the NSA monitors over 650 million telephone conversations *a day*.⁷ That monitoring is automatic. It used to be of foreigners only, but now apparently the system monitors an extraordinary range of communication, searching for that bit or clue that triggers investigative concern. The system produces something akin to a weather report as well as particularized indicators. There are, for example, measures of “chatter” that may signal a storm.

This monitoring, like each of the examples before, creates no burden for those using a telephone. Those using the phone don’t know something is listening on the other end. Instead, the system works quietly in the background, searching this monitored communication in real time.

Video

In each of the examples so far, someone has chosen to use a technology, and that technology has made their privacy vulnerable. The change is produced as that technology evolves to make it simpler to monitor and search behavior.

But the same evolution is happening outside networks as well. Indeed, it is happening in the quintessentially public place—the streets, or in public venues. This monitoring is the production of the current version of video technology. Originally, video cameras were a relatively benign form of monitoring. Because the product of their monitoring relied solely upon human interpretation, there were relatively few contexts in which it paid to have someone watch. And where someone wasn't watching in real time, then the use of these technologies is to trace bad behavior after it happens. Few seem upset when a convenience store video camera makes it possible to identify the criminal who has murdered the attendant.

Digital technology has changed the video, however. It is now a tool of intelligence, not just a tool to record. In London, as I've described, cameras are spread through the city to monitor which cars drive in the city. This is because nonresidents must pay a special tax to drive in "congestion zones." The cameras record and interpret license plates, and then determine whether the right tax was paid for that car. The objective of the system was to minimize congestion in London. Its consequence is a database of every car that enters London, tied to a particular time and location.

But the more ambitious use of video surveillance is human face recognition. While the technology received some very bad press when first introduced in Tampa,⁸ the government continues to encourage companies to develop the capacity to identify who someone is while that someone is in a traditionally anonymous place. As one vendor advertises, "[f]ace recognition technology is the least intrusive and fastest biometric technology. . . . There is no intrusion or delay, and in most cases the subjects are entirely unaware of the process. They do not feel 'under surveillance' or that their privacy has been invaded."⁹

These technologies aren't yet reliable. But they continue to be funded by both private investors and the government. Indeed, the government runs evaluation tests bi-annually to rate the reliability of the technologies.¹⁰ There must at least be someone who expects that someday it will possible to use a camera to identify who is in a crowd, or who boarded a train.

Body Parts

Criminals leave evidence behind, both because they're usually not terribly rational and because it's extremely hard not to. And technology is only making it harder not to. With DNA technology, it becomes increasingly difficult for a criminal to avoid leaving his mark, and increasingly easy for law enforcement to identify with extremely high confidence whether X did Y.

Some nations have begun to capitalize on this new advantage. And again, Britain is in the lead.¹¹ Beginning in 1995, the British government started collecting DNA samples to include in a national registry. The program was initially promoted as a way to fight terrorism. But in a decade, its use has become much less discriminating.

In December 2005, while riding public transportation in London, I read the following on a public announcement poster:

Abuse, Assault, Arrest: Our staff are here to help you. Spitting on DLR staff is classified as an assault and is a criminal offence. Saliva Recovery Kits are now held on every train and will be used to identify offenders against the national DNA database.

And why not? Spitting may be harmless. But it is insulting. And if the tools exist to identify the perpetrator of the insult, why not use them?

In all these cases, technologies designed either without monitoring as their aim or with just limited monitoring as their capacity have now become expert technologies for monitoring. The aggregate of these technologies produces an extraordinary range of searchable data. And, more importantly, as these technologies mature, there will be essentially no way for anyone living within ordinary society to escape this monitoring. Monitoring to produce searchable data will become the default architecture for public space, as standard as street lights. From the simple ability to trace back to an individual, to the more troubling ability to know what that individual is doing or likes at any particular moment, the maturing data infrastructure produces a panopticon beyond anything Bentham ever imagined.

“Orwell” is the word you’re looking for. And while I believe that analogies to Orwell are just about always useless, let’s make one comparison here nonetheless. While the ends of the government in 1984 were certainly vastly more evil than anything our government would ever pursue, it is interesting to note just how inefficient, relative to the current range of technologies, Orwell’s technologies were. The central device was a “telescreen” that both broadcasted content and monitored behavior on the other side. But the great virtue of the telescreen was that you knew what it, in principle, could see. Winston knew where to hide, because the perspective of the telescreen was transparent.¹² It was easy to know what it couldn’t see, and hence easy to know where to do the stuff you didn’t want it to see.

That’s not the world we live in today. You can’t know whether your search on the Internet is being monitored. You don’t know whether a camera is trying to identify who you are. Your telephone doesn’t make funny clicks as the

NSA listens in. Your e-mail doesn't report when some bot has searched it. The technologies of today have none of the integrity of the technologies of 1984. None are decent enough to let you know when your life is being recorded.

There's a second difference as well. The great flaw to the design of 1984 was in imagining just how it was that behavior was being monitored. There were no computers in the story. The monitoring was done by gaggles of guards watching banks of televisions. But that monitoring produced no simple way for the guards to connect their intelligence. There was no search across the brains of the guards. Sure, a guard might notice that you're talking to someone you shouldn't be talking to or that you've entered a part of a city you shouldn't be in. But there was no single guard who had a complete picture of the life of Winston.

Again, that "imperfection" can now be eliminated. We can monitor everything and search the product of that monitoring. Even Orwell couldn't imagine that.

I've surveyed a range of technologies to identify a common form. In each, the individual acts in a context that is technically public. I don't mean it should be treated by the law as "public" in the sense that privacy should not be protected there. I'm not addressing that question yet. I mean only that the individual is putting his words or image in a context that he doesn't control. Walking down 5th Avenue is the clearest example. Sending a letter is another. In both cases, the individual has put himself in a stream of activity that he doesn't control.

The question for us, then, is what limits there should be—in the name of "privacy"—on the ability to surveil these activities. But even that question puts the matter too broadly. By "surveil," I don't mean surveillance generally. I mean the very specific kind of surveillance the examples above evince. I mean what we could call "digital surveillance."

"Digital surveillance" is the process by which some form of human activity is analyzed by a computer according to some specified rule. The rule might say "flag all e-mail talking about Al Qaeda." Or it might say "flag all e-mail praising Governor Dean." Again, at this point I'm not focused upon the normative or legal question of whether such surveillance should be allowed. At this point, we're just working through definitions. In each of the cases above, the critical feature in each is that a computer is sorting data for some follow-up review by some human. The sophistication of the search is a technical question, but there's no doubt that its accuracy is improving substantially.

So should this form of monitoring be allowed?

I find when I ask this question framed precisely like this that there are two polar opposite reactions. On the one hand, friends of privacy say that there's nothing new here. There's no difference between the police reading your mail, and the police's computer reading your e-mail. In both cases, a legitimate and reasonable expectation of privacy has been breached. In both cases, the law should protect against that breach.

On the other hand, friends of security insist there is a fundamental difference. As Judge Richard Posner wrote in the *Washington Post*, in an article defending the Bush Administration's (extensive¹³) surveillance of domestic communications, "[m]achine collection and processing of data cannot, as such, invade privacy." Why? Because it is a machine that is processing the data. Machines don't gossip. They don't care about your affair with your co-worker. They don't punish you for your political opinions. They're just logic machines that act based upon conditions. Indeed, as Judge Posner argues, "[t]his initial sifting, far from invading privacy (a computer is not a sentient being), keeps most private data from being read by any intelligence officer." We're better off having machines read our e-mail, Posner suggests, both because of the security gain, and because the alternative snoop—an intelligence officer—would be much more nosey.

But it would go too far to suggest there isn't some cost to this system. If we lived in a world where our every communication was monitored (if?), that would certainly challenge the sense that we were "left alone." We would be left alone in the sense a toddler is left in a playroom—with parents listening carefully from the next room. There would certainly be something distinctively different about the world of perpetual monitoring, and that difference must be reckoned in any account of whether this sort of surveillance should be allowed.

We should also account for the "best intentions" phenomenon. Systems of surveillance are instituted for one reason; they get used for another. Jeff Rosen has cataloged the abuses of the surveillance culture that Britain has become:¹⁴ Video cameras used to leer at women or for sensational news stories. Or in the United States, the massive surveillance for the purpose of tracking "terrorists" was also used to track domestic environmental and antiwar groups.¹⁵

But let's frame the question in its most compelling form. Imagine a system of digital surveillance in which the algorithm was known and verifiable: We knew, that is, exactly what was being searched for; we trusted that's all that was being searched for. That surveillance was broad and indiscriminate. But before anything could be done on the basis of the results from that surveillance, a court would have to act. So the machine would spit out bits of data implicating X in some targeted crime, and a court would decide

whether that data sufficed either to justify an arrest or a more traditional search. And finally, to make the system as protective as we can, the only evidence that could be used from this surveillance would be evidence directed against the crimes being surveilled for. So for example, if you're looking for terrorists, you don't use the evidence to prosecute for tax evasion. I'm not saying what the targeted crimes are; all I'm saying is that we don't use the traditional rule that allows all evidence gathered legally to be usable for any legal end.

Would such a system violate the protections of the Fourth Amendment? Should it?

The answer to this question depends upon your conception of the value protected by the Fourth Amendment. As I described in Chapter 6, that amendment was targeted against indiscriminate searches and "general warrants"—that is, searches that were not particularized to any particular individual and the immunity that was granted to those engaging in that search. But those searches, like any search at that time, imposed burdens on the person being searched. If you viewed the value the Fourth Amendment protected as the protection from the unjustified burden of this indiscriminate search, then this digital surveillance would seem to raise no significant problems. As framed above, they produce no burden at all unless sufficient evidence is discovered to induce a court to authorize a search.

But it may be that we understand the Fourth Amendment to protect a kind of dignity. Even if a search does not burden anyone, or even if one doesn't notice the search at all, this conception of privacy holds that the very idea of a search is an offense to dignity. That dignity interest is only matched if the state has a good reason to search *before* it searches. From this perspective, a search without justification harms your dignity whether it interferes with your life or not.

I saw these two conceptions of privacy play out against each other in a tragically common encounter in Washington, D.C. A friend and I had arranged a "police ride-along"—riding with District police during their ordinary patrol. The neighborhood we patrolled was among the poorest in the city, and around 11:00 P.M. a report came in that a car alarm had been tripped in a location close to ours. When we arrived near the scene, at least five police officers were attempting to hold three youths; three of the officers were holding the suspects flat against the wall, with their legs spread and their faces pressed against the brick.

These three were "suspects"—they were near a car alarm when it went off—and yet, from the looks of things, you would have thought they had been caught holding the Hope diamond.

And then an extraordinary disruption broke out. To the surprise of everyone, and to my terror (for this seemed a tinder box, and what I am about to describe seemed the match), one of the three youths, no older than seventeen, turned around in a fit of anger and started screaming at the cops. “Every time anything happens in this neighborhood, I get thrown against the wall, and a gun pushed against my head. I’ve never done anything illegal, but I’m constantly being pushed around by cops with guns.”

His friend then turned around and tried to calm him down. “Cool it, man, they’re just trying to do their job. It’ll be over in a minute, and everything will be cool.”

“I’m not going to cool it. Why the fuck do I have to live this way? I am not a criminal. I don’t deserve to be treated like this. Someday one of these guns is going to go off by accident—and then I’ll be a fucking statistic. What then?”

At this point the cops intervened, three of them flipping the indignant youth around against the wall, his face again flat against the brick. “This will be over in a minute. If you check out, you’ll be free to go. Just relax.”

In the voice of rage of the first youth was the outrage of dignity denied. Whether reasonable or not, whether minimally intrusive or not, there was something insulting about this experience—all the more insulting when repeated, one imagines, over and over again. As Justice Scalia has written, wondering whether the framers of the Constitution would have considered constitutional the police practice known as a “Terry stop”—stopping and frisking any individual whenever the police have a reasonable suspicion—“I frankly doubt . . . whether the fiercely proud men who adopted our Fourth Amendment would have allowed themselves to be subjected, on mere suspicion of being armed and dangerous, to such indignity.”¹⁶

And yet again, there is the argument of minimal intrusion. If privacy is a protection against unjustified and excessive disruption, then this was no invasion of privacy. As the second youth argued, the intrusion was minimal; it would pass quickly (as it did—five minutes later, after their identification checked out, we had left); and it was reasonably related to some legitimate end. Privacy here is simply the protection against unreasonable and burdensome intrusions, and this search, the second youth argued, was not so unreasonable and burdensome as to justify the fit of anger (which also risked a much greater danger).

From this perspective, the harm in digital surveillance is even harder to reckon. I’m certain there are those who feel an indignity at the very idea that records about them are being reviewed by computers. But most would recognize a very different dignity at stake here. Unlike those unfortunate kids against the wall, there is no real interference here at all. Very much as with

those kids, if nothing is found, nothing will happen. So what is the indignity? How is it expressed?

A third conception of privacy is about neither preserving dignity nor minimizing intrusion. It is instead substantive—privacy as a way to constrain the power of the state to regulate. Here the work of William Stuntz is a guide.¹⁷ Stuntz argues that the real purpose of the Fourth and Fifth Amendments is to make some types of regulation too difficult by making the evidence needed to prosecute such violations effectively impossible to gather.

This is a hard idea for us to imagine. In our world, the sources of evidence are many—credit card records, telephone records, video cameras at 7-Elevens—so it's hard for us to imagine any crime that there wouldn't be some evidence to prosecute. But put yourself back two hundred years when the only real evidence was testimony and things, and the rules of evidence forbade the defendant from testifying at all. Imagine in that context the state wanted to punish you for "sedition." The only good evidence of sedition would be your writings or your own testimony about your thoughts. If those two sources were eliminated, then it would be practically impossible to prosecute sedition successfully.

As Stuntz argues, this is just what the Fourth and Fifth Amendments do. Combined, they make collecting the evidence for a crime like sedition impossible, thereby making it useless for the state to try to prosecute it. And not just sedition—as Stuntz argues, the effect of the Fourth, Fifth, and Sixth Amendments was to restrict the scope of regulation that was practically possible. As he writes: "Just as a law banning the use of contraceptives would tend to encourage bedroom searches, so also would a ban on bedroom searches tend to discourage laws prohibiting contraceptives."¹⁸

But were not such searches already restricted by, for example, the First Amendment? Would not a law punishing seditious libel have been unconstitutional in any case? In fact, that was not at all clear at the founding; indeed, it was so unclear that in 1798 Congress passed the Alien and Sedition Acts, which in effect punished sedition quite directly.¹⁹ Many thought these laws unconstitutional, but the Fourth and Fifth Amendments would have been effective limits on their enforcement, whether the substantive laws were constitutional or not.

In this conception, privacy is meant as a substantive limit on government's power.²⁰ Understood this way, privacy does more than protect dignity or limit intrusion; privacy limits what government can do.

If this were the conception of privacy, then digital surveillance could well accommodate it. If there were certain crimes that it was inappropriate to prosecute, we could remove them from the search algorithm. It would be

hard to identify what crimes constitutionally must be removed from the algorithm—the First Amendment clearly banishes sedition from the list already. Maybe the rule simply tracks constitutional limitation.

Now the key is to recognize that, in principle, these three distinct conceptions of privacy could yield different results depending on the case. A search, for example, might not be intrusive but might offend dignity. In that case, we would have to choose a conception of privacy that we believed best captured the Constitution's protection.

At the time of the founding, however, these different conceptions of privacy would not, for the most part, have yielded different conclusions. Any search that reached beyond the substantive limits of the amendment, or beyond the limits of dignity, would also have been a disturbance. Half of the framers could have held the dignity conception and half the utility conception, but because every search would have involved a violation of both, all the framers could have endorsed the protections of the Fourth Amendment.

Today, however, that's not true. Today these three conceptions could yield very different results. The utility conception could permit efficient searches that are forbidden by the dignity and substantive conceptions. The correct translation (as Brandeis employed the term in the *Olmstead* wiretapping case) depends on selecting the proper conception to translate.

In this sense, our original protections were the product of what Cass Sunstein calls an "incompletely theorized agreement."²¹ Given the technology of the time, there was no reason to work out which theory underlay the constitutional text; all three were consistent with existing technology. But as the technology has changed, the original context has been challenged. Now that technologies such as the worm can search without disturbing, there is a conflict about what the Fourth Amendment protects.

This conflict is the other side of Sunstein's incompletely theorized agreement. We might say that in any incompletely theorized agreement ambiguities will be latent, and we can describe contexts where these latencies emerge. The latent ambiguities about the protection of privacy, for example, are being rendered patent by the evolution of technology. And this in turn forces us to choose.

Some will once again try to suggest that the choice has been made—by our Constitution, in our past. This is the rhetoric of much of our constitutional jurisprudence, but it is not very helpful here. I do not think the framers worked out what the amendment would protect in a world where perfectly noninvasive searches could be conducted. They did not establish a constitution to apply in all possible worlds; they established a constitution for their world. When their world differs from ours in a way that reveals a choice they did not have to make, then we need to make that choice.

PRIVACY IN PUBLIC: DATA

The story I've told so far is about limits on government: What power should the government have to surveil our activities, at least when those activities are in public? That's the special question raised by cyberspace: What limits on "digital surveillance" should there be? There are, of course, many other more traditional questions that are also important. But my focus was "digital surveillance."

In this part, I consider a third privacy question that is closely related, but very distinct. This is the question of what presumptive controls we should have over the data that we reveal to others. The issue here is not primarily the control of the government. The question is thus beyond the ordinary reach of the Fourth Amendment. Instead, the target of this control is private actors who have either gathered data about me as they've observed me, or collected data from me.

Again, let's take this from the perspective of real space first. If I hire a private detective to follow you around, I've not violated anyone's rights. If I compile a list of places you've been, there's nothing to stop me from selling that list. You might think this intrusive. You might think it outrageous that the law would allow this to happen. But again, the law traditionally didn't worry much about this kind of invasion because the costs of such surveillance were so high. Celebrities and the famous may wish the rules were different, but for most of us, for most of our history, there was no need for the law to intervene.

The same point could be made about the data I turned over to businesses or others in the days before the Internet. There was nothing in the law to limit what these entities did with that data. They could sell it to mailing list companies or brokers; they could use it however they wanted. Again, the practical cost of doing things with such data was high, so there wasn't that much done with this data. And, more importantly, the invasiveness of any such use of data was relatively low. Junk mail was the main product, and junk mail in physical space is not a significant burden.

But here, as with "digital surveillance," things have changed dramatically. Just a couple stories will give us a taste of the change:

- In the beginning of 2006, the *Chicago Sun-Times* reported²² that there were websites selling the records of telephone calls made from cell phones. A blog, AmericaBlog, demonstrated the fact by purchasing the cell phone records of General Wesley Clark. For around \$120, the blog was able to prove what most would have thought impossible: that anyone with a credit card could find

something so personal as the list (and frequency and duration) of people someone calls on a cell phone.

This conduct was so outrageous that no one really stood up to defend it. But the defense isn't hard to construct. Wesley Clark "voluntarily" dialed the numbers on his cell phone. He thus voluntarily turned that data over to the cell phone company. Because the cell phone company could sell data, it made it easier for the company to keep prices low(er). Clark benefited from those lower prices. So what's his complaint?

- A number of years ago I received a letter from AT&T. It was addressed to an old girlfriend, but the letter had not been forwarded. The address was my then-current apartment. AT&T wanted to offer her a new credit card. They were a bit late: She and I had broken up eight years before. Since then, she had moved to Texas, and I had moved to Chicago, to Washington, back to Chicago, on to New Haven, back to Chicago, and finally to Boston, where I had moved twice. My peripateticism, however, did not deter AT&T. With great faith in my constancy, it believed that a woman I had not even seen in many years was living with me in this apartment.

How did AT&T maintain such a belief? Well, floating about in cyberspace is lots of data about me. It has been collected from me ever since I began using credit cards, telephones, and who knows what else. The system continuously tries to update and refine this extraordinary data set—that is, it profiles who I am and, using that profile, determines how it will interact with me.

These are just the tip of the iceberg. Everything you do on the Net produces data. That data is, in aggregate, extremely valuable, more valuable to commerce than it is to the government. The government (in normal times) really cares only that you obey some select set of laws. But commerce is keen to figure out how you want to spend your money, and data does that. With massive amounts of data about what you do and what you say, it becomes increasingly possible to market to you in a direct and effective way. Google Gmail processes the data in your e-mail to see what it should try to sell. Amazon watches what you browse to see what special "Gold Box" offers it can make. There's an endless list of entities that want to know more about you to better serve (at least) their interests. What limits, or restrictions, ought there to be on them?

We should begin with an obvious point that might help direct an answer. There's a big difference between (1) collecting data about X to suss out a crime or a criminal, (2) collecting data about X that will be sold to Y simply to reveal facts about X (such as his cell phone calls), and (3) collecting data about X to better market to X. (1) and (2) make X worse off, though if we

believe the crime is properly a crime, then with (1), X is not worse off relative to where he should be. (3) in principle could make you better off—it facilitates advertising that is better targeted and better designed to encourage voluntary transactions. I say “in principle” because even though it’s possible that the ads are better targeted, there are also more of them. On balance, X might be worse off with the flood of well-targeted offers than with a few less well-targeted offers. But despite that possibility, the motive of (3) is different from (1) and (2), and that might well affect how we should respond.

So let’s begin with the focus on (3): What is the harm from this sort of “invasion”? Arguments rage on both sides of this question.

The “no harm” side assumes that the balance of privacy is struck at the line where you reveal information about yourself to the public. Sure, information kept behind closed doors or written in a private diary should be protected by the law. But when you go out in public, when you make transactions there or send material there, you give up any right to privacy. Others now have the right to collect data about your public behavior and do with it what suits them.

Why is that idea not troubling to these theorists? The reasons are many:

- First, the harm is actually not very great. You get a discount card at your local grocery store; the store then collects data about what you buy. With that data, the store may market different goods to you or figure out how better to price its products; it may even decide that it should offer different mixes of discounts to better serve customers. These responses, the argument goes, are the likely ones, because the store’s business is only to sell groceries more efficiently.
- Second, it is an unfair burden to force others to ignore what you show them. If data about you are not usable by others, then it is as if you were requiring others to discard what you have deposited on their land. If you do not like others using information about you, do not put it in their hands.
- Third, these data actually do some good. I do not know why Nike thinks I am a good person to tell about their latest sneakers, and I do not know why Keds does not know to call. In both cases, I suspect the reason is bad data about me. I would love it if Nike knew enough to leave me alone. And if these data were better collected and sorted, it would.
- Finally, in general, companies don’t spend money collecting these data to actually learn anything about you. They want to learn about people *like* you. They want to know your type. In principle, they would be happy to know your type even if they could not then learn who you are. What the merchants want is a way to discriminate—only in the sense of being able to tell the difference between sorts of people.

The other side of this argument, however, also has a point. It begins, again, by noticing the values that were originally protected by the imperfection of monitoring technology. This imperfection helped preserve important substantive values; one such value is the benefit of innocence. At any given time, there are innocent facts about you that may appear, in a particular context or to a particular set, guilty. Peter Lewis, in a *New York Times* article called “Forget Big Brother,” puts the point well:

Surveillance cameras followed the attractive young blond woman through the lobby of the midtown Manhattan hotel, kept a glassy eye on her as she rode the elevator up to the 23rd floor and peered discreetly down the hall as she knocked at the door to my room. I have not seen the videotapes, but I can imagine the digital readout superimposed on the scenes, noting the exact time of the encounter. That would come in handy if someone were to question later why this woman, who is not my wife, was visiting my hotel room during a recent business trip. The cameras later saw us heading off to dinner and to the theater—a middle aged, married man from Texas with his arm around a pretty East Village woman young enough to be his daughter.

“As a matter of fact,” Lewis writes, “she is my daughter.”²³

One lesson of the story is the burden of these monitored facts. The burden is on you, the monitored, first to establish your innocence, and second to assure all who might see these ambiguous facts that you are innocent. Both processes, however, are imperfect; say what you want, doubts will remain. There are always some who will not believe your plea of innocence.

Modern monitoring only exacerbates this problem. Your life becomes an ever-increasing record; your actions are forever held in storage, open to being revealed at any time, and therefore at any time demanding a justification.

A second value follows directly from this modern capacity for archiving data. We all desire to live in separate communities, or among or within separate normative spaces. Privacy, or the ability to control data about yourself, supports this desire. It enables these multiple communities and disables the power of one dominant community to norm others into oblivion. Think, for example, about a gay man in an intolerant small town.

The point comes through most clearly when contrasted with an argument advanced by David Brin.²⁴ Brin argues against this concern with privacy—at least if privacy is defined as the need to block the production and distribution of data about others. He argues against it because he believes that such an end is impossible; the genie is out of the bottle. Better, he suggests, to find ways to ensure that this data-gathering ability is generally available. The solution to

your spying on me is not to block your spying, but to let me spy on you—to hold you accountable, perhaps for spying, perhaps for whatever else you might be doing.

There are two replies to this argument. One asks: Why do we have to choose? Why can't we both control spying and build in checks on the distribution of spying techniques?

The other reply is more fundamental. Brin assumes that this counter spying would be useful to hold others "accountable." But according to whose norms? "Accountable" is a benign term only so long as we have confidence in the community doing the accounting. When we live in multiple communities, accountability becomes a way for one community to impose its view of propriety on another. Because we do not live in a single community, we do not live by a single set of values. And perfect accountability can only undermine this mix of values.

The imperfection in present monitoring enables this multiplication of normative communities. The ability to get along without perfect recording enables a diversity that perfect knowledge would erase.

A third value arises from a concern about profiling. If you search within Google for "mortgage" in a web search engine, advertising for mortgages appears on your computer screen. The same for sex and for cars. Advertising is linked to the search you submit. Data is collected, but not just about the search. Different sites collect just about every bit of personal information about you that they can.²⁵ And when you link from the Google search to a web page, the search you just performed is passed along to the next site.

Data collection is the dominant activity of commercial websites. Some 92 percent of them collect personal data from web users, which they then aggregate, sort, and use.²⁶ Oscar Gandy calls this the "panoptic sort"—a vast structure for collecting data and discriminating on the basis of that data—and it is this discrimination, he says, that ought to concern us.²⁷

But why should it concern us? Put aside an important class of problems—the misuse of the data—and focus instead on its ordinary use. As I said earlier, the main effect is simply to make the market work more smoothly: Interests and products are matched to people in a way that is better targeted and less intrusive than what we have today. Imagine a world where advertisers could tell which venues paid and which did not; where it was inefficient to advertise with billboards and on broadcasts; where most advertising was targeted and specific. Advertising would be more likely to go to those people for whom it would be useful information. Or so the argument goes. This is discrimination, no doubt, but not the discrimination of Jim Crow. It is the wonderful sort of discrimination that spares me Nike ads.

But beyond a perhaps fleeting concern about how such data affect the individual, profiling raises a more sustained collective concern about how it might affect a community.

That concern is manipulation. You might be skeptical about the power of television advertising to control people's desires: Television is so obvious, the motives so clear. But what happens when the motive is not so obvious? When options just seem to appear right when you happen to want them? When the system seems to know what you want better and earlier than you do, how can you know where these desires really come from?

Whether this possibility is a realistic one, or whether it should be a concern, are hard and open questions. Steven Johnson argues quite effectively that in fact these agents of choice will facilitate a much greater range and diversity—even, in part, chaos—of choice.²⁸ But there's another possibility as well—profiles will begin to normalize the population from which the norm is drawn. The observing will affect the observed. The system watches what you do; it fits you into a pattern; the pattern is then fed back to you in the form of options set by the pattern; the options reinforce the pattern; the cycle begins again.

A second concern is about equality. Profiling raises a question that was latent in the market until quite recently. For much of the nineteenth century in the United States economic thought was animated by an ideal of equality. In the civil space individuals were held to be equal. They could purchase and sell equally; they could approach others on equal terms. Facts about individuals might be known, and some of these facts might disqualify them from some economic transactions—your prior bankruptcy, for example, might inhibit your ability to make transactions in the future. But in the main, there were spaces of relative anonymity, and economic transactions could occur within them.²⁹

Over time this space of equality has been displaced by economic zonings that aim at segregation.³⁰ They are laws, that is, that promote distinctions based on social or economic criteria.³¹ The most telling example is zoning itself. It was not until this century that local law was used to put people into segregated spaces.³² At first, this law was racially based, but when racially based zoning was struck down, the techniques of zoning shifted.³³

It is interesting to recall just how contentious this use of law was.³⁴ To many, rich and poor alike, it was an affront to the American ideal of equality to make where you live depend on how much money you had. It always does, of course, when property is something you must buy. But zoning laws add the support of law to the segregation imposed by the market. The effect is to recreate in law, and therefore in society, distinctions among people.

There was a time when we would have defined our country as a place that aimed to erase these distinctions. The historian Gordon Wood describes this goal as an important element of the revolution that gave birth to the United States.³⁵ The enemy was social and legal hierarchy; the aim was a society of equality. The revolution was an attack on hierarchies of social rank and the special privileges they might obtain.

All social hierarchies require information before they can make discriminations of rank. Having enough information about people required, historically, fairly stable social orders. Making fine class distinctions—knowing, for instance, whether a well-dressed young man was the gentleman he claimed to be or only a dressed-up tradesman—required knowledge of local fashions, accents, customs, and manners. Only where there was relatively little mobility could these systems of hierarchy be imposed.

As mobility increased, then, these hierarchical systems were challenged. Beyond the extremes of the very rich and very poor, the ability to make subtle distinctions of rank disappeared as the mobility and fluidity of society made them too difficult to track.

Profiling changes all this. An efficient and effective system for monitoring makes it possible once again to make these subtle distinctions of rank. Collecting data cheaply and efficiently will take us back to the past. Think about frequent flyer miles. Everyone sees the obvious feature of frequent flyer miles—the free trips for people who fly frequently. This rebate program is quite harmless on its own. The more interesting part is the power it gives to airlines to discriminate in their services.

When a frequent flyer makes a reservation, the reservation carries with it a customer profile. This profile might include information about which seat she prefers or whether she likes vegetarian food. It also tells the reservation clerk how often this person flies. Some airlines would then discriminate on the basis of this information. The most obvious way is through seat location—frequent flyers get better seats. But such information might also affect how food is allocated on the flight—the frequent flyers with the most miles get first choice; those with the fewest may get no choice.

In the scheme of social justice, of course, this is small potatoes. But my point is more general. Frequent flyer systems permit the re-creation of systems of status. They supply information about individuals that organizations might value, and use, in dispensing services.³⁶ They make discrimination possible because they restore information that mobility destroyed. They are ways of defeating one benefit of anonymity—the benefit of equality.

Economists will argue that in many contexts this ability to discriminate—in effect, to offer goods at different prices to different people—is overall a

benefit.³⁷ On average, people are better off if price discrimination occurs than if it does not. So we are better off, these economists might say, if we facilitate such discrimination when we can.

But these values are just one side of the equation. Weighed against them are the values of equality. For us they may seem remote, but we should not assume that because they are remote now they were always remote.

Take tipping: As benign (if annoying) as you might consider the practice of tipping, there was a time at the turn of the century when the very idea was an insult. It offended a free citizen's dignity. As Viviana Zelizer describes it:

In the early 1900s, as tipping became increasingly popular, it provoked great moral and social controversy. In fact, there were nationwide efforts, some successful, by state legislatures to abolish tipping by turning it into a punishable misdemeanor. In countless newspaper editorials and magazine articles, in etiquette books, and even in court, tips were closely scrutinized with a mix of curiosity, amusement, and ambivalence—and often open hostility. When in 1907, the government officially sanctioned tipping by allowing commissioned officers and enlisted men of the United States Navy to include tips as an item in their travel expense vouchers, the decision was denounced as an illegitimate endorsement of graft. Periodically, there were calls to organize anti-tipping leagues.³⁸

There is a conception of equality that would be corrupted by the efficiency that profiling embraces. That conception is a value to be weighed against efficiency. Although I believe this value is relatively weak in American life, who am I to say? The important point is not about what is strong or weak, but about the tension or conflict that lay dormant until revealed by the emerging technology of profiling.

The pattern should be familiar by now, because we have seen the change elsewhere. Once again, the code changes, throwing into relief a conflict of values. Whereas before there was relative equality because the information that enabled discrimination was too costly to acquire, now it pays to discriminate. The difference—what makes it pay—is the emergence of a code. The code changes, the behavior changes, and a value latent in the prior regime is displaced.

We could react by hobbling the code, thus preserving this world. We could create constitutional or statutory restrictions that prevent a move to the new world. Or we could find ways to reconcile this emerging world with the values we think are fundamental.

SOLUTIONS

I've identified two distinct threats to the values of privacy that the Internet will create. The first is the threat from "digital surveillance"—the growing capacity of the government (among others) to "spy" on your activities "in public." From Internet access, to e-mail, to telephone calls, to walking on the street, digital technology is opening up the opportunity for increasingly perfect burdenless searches.

The second threat comes from the increasing aggregation of data by private (among other) entities. These data are gathered not so much to "spy" as to facilitate commerce. Some of that commerce exploits the source of the data (Wesley Clark's cell phone numbers). Some of that commerce tries to facilitate commerce with the source of that data (targeted ads).

Against these two different risks, we can imagine four types of responses, each mapping one of the modalities that I described in Chapter 7:

- **Law:** Legal regulation could be crafted to respond to these threats. We'll consider some of these later, but the general form should be clear enough. The law could direct the President not to surveil American citizens without reasonable suspicion, for example. (Whether the President follows the law is a separate question.) Or the law could ban the sale of data gathered from customers without express permission of the customers. In either case, the law threatens sanctions to change behavior directly. The aim of the law could either be to enhance the power of individuals to control data about them, or to disable such power (for example, by making certain privacy-related transactions illegal).
- **Norms:** Norms could be used to respond to these threats. Norms among commercial entities, for example, could help build trust around certain privacy protective practices.
- **Markets:** In ways that will become clearer below, the market could be used to protect the privacy of individuals.
- **Architecture/Code:** Technology could be used to protect privacy. Such technologies are often referred to as "Privacy Enhancing Technologies." These are technologies designed to give the user more technical control over data associated with him or her.

As I've argued again and again, there is no single solution to policy problems on the Internet. Every solution requires a mix of at least two modalities. And in the balance of this chapter, my aim is to describe a mix for each of these two threats to privacy.

No doubt this mix will be controversial to some. But my aim is not so much to push any particular mix of settings on these modality dials, as it is to demonstrate a certain approach. I don't insist on the particular solutions I propose, but I do insist that solutions in the context of cyberspace are the product of such a mix.

Surveillance

The government surveils as much as it can in its fight against whatever its current fight is about. When that surveillance is human—wiretapping, or the like—then traditional legal limits ought to apply. Those limits impose costs (and thus, using the market, reduce the incidence to those most significant); they assure at least some review. And, perhaps most importantly, they build within law enforcement a norm respecting procedure.

When that surveillance is digital, however, then it is my view that a different set of restrictions should apply. The law should sanction “digital surveillance” if, *but only if*, a number of conditions apply:

1. The purpose of the search enabled in the algorithm is described.
2. The function of the algorithm is reviewed.
3. The purpose and the function match is certified.
4. No action—including a subsequent search—can be taken against any individual on the basis of the algorithm without judicial review.
5. With very limited exceptions, no action against any individual can be pursued for matters outside the purpose described. Thus, if you're looking for evidence of drug dealing, you can't use any evidence discovered for prosecuting credit card fraud.

That describes the legal restrictions applied against the government in order to enhance privacy. If these are satisfied, then in my view such digital surveillance should not conflict with the Fourth Amendment. In addition to these, there are privacy enhancing technologies (PETs) that should be broadly available to individuals as well. These technologies enable individuals to achieve anonymity in their transactions online. Many companies and activist groups help spread these technologies across the network.

Anonymity in this sense simply means non-traceability. Tools that enable this sort of non-traceability make it possible for an individual to send a message without the content of that message being traced to the sender. If implemented properly, there is absolutely no technical way to trace that message. That kind of anonymity is essential to certain kinds of communication.

It is my view that, at least so long as political repression remains a central feature of too many world governments, free governments should recognize a protected legal right to these technologies. I acknowledge that view is controversial. A less extreme view would acknowledge the differences between the digital world and real world,³⁹ and guarantee a right to pseudonymous communication but not anonymous communication. In this sense, a pseudonymous transaction doesn't obviously or directly link to an individual without court intervention. But it contains an effective fingerprint that would allow the proper authority, under the proper circumstances, to trace the communication back to its originator.

In this regime, the important question is who is the authority, and what process is required to get access to the identification. In my view, the authority must be the government. The government must subject its demand for revealing the identity of an individual to judicial process. And the executive should never hold the technical capacity to make that link on its own.

Again, no one will like this balance. Friends of privacy will be furious with any endorsement of surveillance. But I share Judge Posner's view that a sophisticated surveillance technology might actually increase effective privacy, if it decreases the instances in which humans intrude on other humans. Likewise, friends of security will be appalled at the idea that anyone would endorse technologies of anonymity. "Do you know how hard it is to crack a drug lord's encrypted e-mail communication?" one asked me.

The answer is no, I don't have a real sense. But I care less about enabling the war on drugs than I do about enabling democracies to flourish. Technologies that enable the latter will enable the former. Or to be less cowardly, technologies that enable Aung San Suu Kyi to continue to push for democracy in Burma will enable Al Qaeda to continue to wage its terrorist war against the United States. I acknowledge that. I accept that might lead others to a less extreme position. But I would urge the compromise in favor of surveillance to go no further than protected pseudonymity.

Control of Data

The problem of controlling the spread or misuse of data is more complex and ambiguous. There are uses of personal data that many would object to. But many is not all. There are some who are perfectly happy to reveal certain data to certain entities, and there are many more who would become happy if they could trust that their data was properly used.

Here again, the solution mixes modalities. But this time, we begin with the technology.⁴⁰

As I described extensively in Chapter 4, there is an emerging push to build an Identity Layer onto the Internet. In my view, we should view this Identity Layer as a PET (private enhancing technology): It would enable individuals to more effectively control the data about them that they reveal. It would also enable individuals to have a trustable pseudonymous identity that websites and others should be happy to accept. Thus, with this technology, if a site needs to know I am over 18, or an American citizen, or authorized to access a university library, the technology can certify this data without revealing anything else. Of all the changes to information practices that we could imagine, this would be the most significant in reducing the extent of redundant or unnecessary data flowing in the ether of the network.

A second PET to enable greater control over the use of data would be a protocol called the Platform for Privacy Preferences (or P3P for short).⁴¹ P3P would enable a *machine-readable* expression of the privacy preferences of an individual. It would enable an automatic way for an individual to recognize when a site does not comply with his privacy preferences. If you surf to a site that expresses its privacy policy using P3P, and its policy is inconsistent with your preferences, then depending upon the implementation, either the site or you are made aware of the problem created by this conflict. The technology thus could make clear a conflict in preferences. And recognizing that conflict is the first step to protecting preferences.

The critical part of this strategy is to make these choices machine-readable. If you Google “privacy policy,” you’ll get close to 2.5 *billion* hits on the Web. And if you click through to the vast majority of them (not that you could do that in this lifetime), you will find that they are among the most incomprehensible legal texts around (and that’s saying a lot). These policies are the product of pre-Internet thinking about how to deal with a policy problem. The government was pushed to “solve” the problem of Internet privacy. Its solution was to require “privacy policies” be posted everywhere. But does anybody read these policies? And if they do, do they remember them from one site to another? Do you know the difference between Amazon’s policies and Google’s?

The mistake of the government was in not requiring that those policies also be understandable by a computer. Because if we had 2.5 billion sites with both a human readable and machine readable statement of privacy policies, then we would have the infrastructure necessary to encourage the development of this PET, P3P. But because the government could not think beyond its traditional manner of legislating—because it didn’t think to require changes in code as well as legal texts—we don’t have that infrastructure now. But, in my view, it is critical.

These technologies standing alone, however, do nothing to solve the problem of privacy on the Net. It is absolutely clear that to complement these technologies, we need legal regulation. But this regulation is of three very different sorts. The first kind is substantive—laws that set the boundaries of privacy protection. The second kind is procedural—laws that mandate fair procedures for dealing with privacy practices. And the third is enabling—laws that make enforceable agreements between individuals and corporations about how privacy is to be respected.

(1) Limits on Choice

One kind of legislation is designed to limit individual freedom. Just as labor law bans certain labor contracts, or consumer law forbids certain credit arrangements, this kind of privacy law would restrict the freedom of individuals to give up certain aspects of their privacy. The motivation for this limitation could either be substantive or procedural—substantive in that it reflects a substantive judgment about choices individuals should not make, or procedural in that it reflects the view that systematically, when faced with this choice, individuals will choose in ways that they regret. In either case, the role of this type of privacy regulation is to block transactions deemed to weaken privacy within a community.

(2) The Process to Protect Privacy

The most significant normative structure around privacy practices was framed more than thirty years ago by the HEW (Health, Education, Welfare) Advisory Committee on Automated Data Systems. This report set out five principles that were to define the “Code of Fair Information Practices.”⁴² These principles require:

1. There must be no personal data record-keeping systems whose very existence is secret.
2. There must be a way for a person to find out what information about the person is in a record and how it is used.
3. There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person’s consent.
4. There must be a way for a person to correct or amend a record of identifiable information about the person.

5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.

These principles express important substantive values—for example, that data not be reused beyond an original consent, or that systems for gathering data be reliable—but they don't interfere with an individual's choice to release his or her own data for specified purposes. They are in this sense individual autonomy enhancing, and their spirit has guided the relatively thin and ad hoc range of privacy legislation that has been enacted both nationally and at the state level.⁴³

(3) Rules to Enable Choice About Privacy

The real challenge for privacy, however, is how to enable a meaningful choice in the digital age. And in this respect, the technique of the American government so far—namely, to require text-based privacy policy statements—is a perfect example of how not to act. Cluttering the web with incomprehensible words will not empower consumers to make useful choices as they surf the Web. If anything, it drives consumers away from even attempting to understand what rights they give away as they move from site to site.

P3P would help in this respect, but only if (1) there were a strong push to spread the technology across all areas of the web and (2) the representations made within the P3P infrastructure were enforceable. Both elements require legal action to be effected.

In the first edition of this book, I offered a strategy that would, in my view, achieve both (1) and (2): namely, by protecting personal data through a property right. As with copyright, a privacy property right would create strong incentives in those who want to use that property to secure the appropriate consent. That content could then be channeled (through legislation) through appropriate technologies. But without that consent, the user of the privacy property would be a privacy pirate. Indeed, many of the same tools that could protect copyright in this sense could also be used to protect privacy.

This solution also recognizes what I believe is an important feature of privacy—that people value privacy differently.⁴⁴ It also respects those different values. It may be extremely important to me not to have my telephone number easily available; you might not care at all. And as the law's presumptive preference is to use a legal device that gives individuals the freedom to be different—meaning the freedom to have and have respected wildly different subjective values—that suggests the device we use here is property. A property

system is designed precisely to permit differences in value to be respected by the law. If you won't sell your Chevy Nova for anything less than \$10,000, then the law will support you.

The opposite legal entitlement in the American legal tradition is called a "liability rule."⁴⁵ A liability rule also protects an entitlement, but its protection is less individual. If you have a resource protected by a liability rule, then I can take that resource so long as I pay a state-determined price. That price may be more or less than you value it at. But the point is, I have the right to take that resource, regardless.

An example from copyright law might make the point more clearly. A derivative right is the right to build upon a copyrighted work. A traditional example is a translation, or a movie based on a book. The law of copyright gives the copyright owner a property right over that derivative right. Thus, if you want to make a movie out of John Grisham's latest novel, you have to pay whatever Grisham says. If you don't, and you make the movie, you've violated Grisham's rights.

The same is not true with the derivative rights that composers have. If a songwriter authorizes someone to record his song, then anyone else has a right to record that song, so long as they follow certain procedures and pay a specified rate. Thus, while Grisham can choose to give only one filmmaker the right to make a film based on his novel, the Beatles must allow anyone to record a song a member of the Beatles composed, so long as that person pays. The derivative right for novels is thus protected by a property rule; the derivative right for recordings by a liability rule.

The law has all sorts of reasons for imposing a liability rule rather than a property rule. But the general principle is that we should use a property rule, at least where the "transaction costs" of negotiating are low, and where there is no contradicting public value.⁴⁶ And it is my view that, with a technology like P3P, we could lower transaction costs enough to make a property rule work. That property rule in turn would reinforce whatever diversity people had about views about their privacy—permitting some to choose to waive their rights and others to hold firm.

There was one more reason I pushed for a property right. In my view, the protection of privacy would be stronger if people conceived of the right as a property right. People need to take ownership of this right, and protect it, and propertizing is the traditional tool we use to identify and enable protection. If we could see one fraction of the passion defending privacy that we see defending copyright, we might make progress in protecting privacy.

But my proposal for a property right was resoundingly rejected by critics whose views I respect.⁴⁷ I don't agree with the core of these criticisms. For the

reasons powerfully marshaled by Neil Richards, I especially don't agree with the claim that there would be a First Amendment problem with propertizing privacy.⁴⁸ In any case, William McGeeveran suggested an alternative that reached essentially the same end that I sought, without raising any of the concerns that most animated the critics.⁴⁹

The alternative simply specifies that a representation made by a website through the P3P protocol be considered a binding offer, which, if accepted by someone using the website, becomes an enforceable contract.⁵⁰ That rule, tied to a requirement that privacy policies be expressed in a machine-readable form such as P3P, would both (1) spread P3P and (2) make P3P assertions effectively law. This would still be weaker than a property rule, for reasons I will leave to the notes.⁵¹ And it may well encourage the shrink-wrap culture, which raises its own problems. But for my purposes here, this solution is a useful compromise.

To illustrate again the dynamic of cyberlaw: We use law (a requirement of policies expressed in a certain way, and a contract presumption about those expressions) to encourage a certain kind of technology (P3P), so that that technology enables individuals to better achieve in cyberspace what they want. It is LAW helping CODE to perfect privacy POLICY.

This is not to say, of course, that we have no protections for privacy. As we have seen throughout, there are other laws besides federal, and other regulators besides the law. At times these other regulators may protect privacy better than law does, but where they don't, then in my view law is needed.

PRIVACY COMPARED

The reader who was dissatisfied with my argument in the last chapter is likely to begin asking pointed questions. "Didn't you reject in the last chapter the very regime you are endorsing here? Didn't you reject an architecture that would facilitate perfect sale of intellectual property? Isn't that what you've created here?"

The charge is accurate enough. I have endorsed an architecture here that is essentially the same architecture I questioned for intellectual property. Both are regimes for trading information; both make information "like" "real" property. But with copyright, I argued against a fully privatized property regime; with privacy, I am arguing in favor of it. What gives?

The difference is in the underlying values that inform, or that should inform, information in each context. In the context of intellectual property, our bias should be for freedom. Who knows what "information wants";⁵² whatever it wants, we should read the bargain that the law strikes with holders

of intellectual property as narrowly as we can. We should take a grudging attitude to property rights in intellectual property; we should support them only as much as necessary to build and support information regimes.

But (at least some kinds of) information about individuals should be treated differently. You do not strike a deal with the law about personal or private information. The law does not offer you a monopoly right in exchange for your publication of these facts. That is what is distinct about privacy: Individuals should be able to control information about themselves. We should be eager to help them protect that information by giving them the structures and the rights to do so. We value, or want, our peace. And thus, a regime that allows us such peace by giving us control over private information is a regime consonant with public values. It is a regime that public authorities should support.

There is a second, perhaps more helpful, way of making the same point. Intellectual property, once created, is non-diminishable. The more people who use it, the more society benefits. The bias in intellectual property is thus, properly, towards sharing and freedom. Privacy, on the other hand, is diminishable. The more people who are given license to tread on a person's privacy, the less that privacy exists. In this way, privacy is more like real property than it is like intellectual property. No single person's trespass may destroy it, but each incremental trespass diminishes its value by some amount.

This conclusion is subject to important qualifications, only two of which I will describe here.

The first is that nothing in my regime would give individuals final or complete control over the kinds of data they can sell, or the kinds of privacy they can buy. The P3P regime would in principle enable upstream control of privacy rights as well as individual control. If we lived, for example, in a regime that identified individuals based on jurisdiction, then transactions with the P3P regime could be limited based on the rules for particular jurisdictions.

Second, there is no reason such a regime would have to protect all kinds of private data, and nothing in the scheme so far tells us what should and should not be considered "private" information. There may be facts about yourself that you are not permitted to hide; more important, there may be claims about yourself that you are not permitted to make ("I am a lawyer," or, "Call me, I'm a doctor"). You should not be permitted to engage in fraud or to do harm to others. This limitation is an analog to fair use in intellectual property—a limit to the space that privacy may protect.

I started this chapter by claiming that with privacy the cat is already out of the bag. We already have architectures that deny individuals control over what others know about them; the question is what we can do in response.

My response has been: Look to the code, Luke. We must build into the architecture a capacity to enable choice—not choice by humans but by machines. The architecture must enable machine-to-machine negotiations about privacy so that individuals can instruct their machines about the privacy they want to protect.

But how will we get there? How can this architecture be erected? Individuals may want cyberspace to protect their privacy, but what would push cyberspace to build in the necessary architectures?

Not the market. The power of commerce is not behind any such change. Here, the invisible hand would really be invisible. Collective action must be taken to bend the architectures toward this goal, and collective action is just what politics is for. *Laissez-faire* will not cut it.