THE FUTURE OF IDEAS

THE FATE OF THE COMMONS IN A CONNECTED WORLD

Lawrence Lessig

VINTAGE BOOKS
A DIVISION OF RANDOM HOUSE, INC.
NEW YORK

2001/2002

2

Building Blocks: "Commons" and "Layers"

THIS BOOK is fundamentally about the Internet and its effect on innovation, both commercial and non-. "Internet" and "society" are familiar enough notions. But at the core of my argument are two fairly obscure ideas that we must begin by making a bit more clear. The first of these is the idea of a "commons"; the second is the notion of "layers." The commons is an old idea; layers, in the sense made familiar by network theorists, are relatively new. But the two together organize the argument that follows. They are building blocks to an end that will help reveal the Internet's effect on society.

THE COMMONS

IF YOU'VE used the word *commons* before, you're likely to think of a park, as in the Boston Common. If you've studied economics or political science, your mind will race to tragedy (as in "the tragedy of the commons"). Both senses are related to what I mean, but neither alone is enough.[1]

The Oxford English Dictionary (mankind's first large-scale collaborative open source text project)[2] equates the "commons" to a resource held "in common." That it defines as "in joint use or possession; to be held or enjoyed equally by a number of persons."[3] In this sense, a resource held "in common" is "free" (as I've defined that term) to those "persons." In most cases, the commons is a resource to which anyone within the relevant community has a right without obtaining the permission of anyone else. In some cases, permission is needed but is granted in a neutral way.

Think about some examples:

- The public streets are commons. Anyone is free to access the streets without first getting the permission of someone else. We don't auction rights of access, selling the right to use a particular bit of highway during a particular bit of time. (Of course there are exceptions.) Nor do we insist on particular licenses before we allow people to use the streets or highways. Instead the highways are open and free—in the sense I mean a commons to be free.
- Parks and beaches are increasingly commons. Anyone is free to access these spaces without getting the permission of someone else. Access is not auctioned off to the highest bidder, and the right to control access is not handed off to some private or governmental entity. The resource—as Carol Rose calls it, "the recreational resource"—is made available to anyone.
- Einstein's theory of relativity is a commons. It is a resource—a way of understanding the nature of the universe—that is open and free for anyone to take. Access to this resource is not auctioned off to the highest bidder; the right to use the theory is not allocated to a single organization.
- Writings in the public domain are a commons. They are a resource that is open and free for anyone to take without the permission of anyone else. An 1890 edition of Shakespeare is free for anyone to take and copy. Your right to use and redistribute that 1890 text is without restraint.

Each of these resources is held in common. Each is "free" for others to take. Some are free in the sense that no price is paid (you can use most roads without paying a toll; as we will see, it would be unconstitutional in the United States to require anyone to pay to use Einstein's theory of relativity). Some are free even though a price must be paid (a park is "free" in the sense that I mean even if an access fee is required—as long as the fee is neutrally and consistently applied).[4] In both cases, the essential feature is reasonable, and that access to the resource is not conditioned upon the permission of someone else. The essence, in other words, is that no one exercises the core of a property right with respect to these resources—the exclusive right to choose whether the resource is made available to others.[5]

Economists will object, however, that my list conflates two very different

cases. Einstein's theory of relativity is different from the streets or public beaches. Einstein's theory is fully "nonrivalrous"; the streets and beaches are not. If you use the theory of relativity, there is as much left over afterward as there was before. Your consumption, in other words, does not *rival* my own. But roads and beaches are very different. If everyone tries to use the roads at the very same time (something that apparently happens out here in California often), then their use certainly rivals my own. Traffic jams; public beaches crowd. Your SUV, or your loud radio, reduces my ability to enjoy the roads or beach.

The economists are right. This list of resources held in "the commons" does conflate rivalrous with nonrivalrous resources. But our tradition is not as tidy as the economists' analytics. We have always described as "commons" both rivalrous and nonrivalrous resources. The Boston Common is a commons, though its resource is rivalrous (my use of it competes with your use of it). Language is a commons, though its resource is nonrivalrous (my use of it does not inhibit yours).[6] What has determined "the commons," then, is not the simple test of rivalrousness. What has determined the commons is the *character of the resource* and how it *relates to a community*. In theory, any resource might be held in common (whether it would survive is another question). But in practice, the question a society must ask is which resources *should be*, and for those resources, *how*.

Here the distinction that the economists draw begins to help. Economists distinguish rivalrous and nonrivalrous resources because the issues or problems raised by each kind are different.

If a resource is nonrivalrous, then the problem is whether there is enough incentive to produce it, not whether there is too much demand to consume it. A nonrivalrous resource can't be exhausted. Once it is produced, it can't be undone. Thus the issue for nonrivalrous resources is whether the Edith Whartons of the world have enough incentive to create. The problem with nonrivalrous resources is to assure that I reap enough benefit to induce me to sow.

A rivalrous resource presents more problems. If a resource is rivalrous, then we must worry both about whether there is sufficient incentive to create it (if it is the sort of resource that humans produce) and about whether consumption by some will leave enough to others. With a rivalrous resource, I must still worry that I will reap enough benefit to make it worth it to sow. But I must worry as well that others not deplete the resource that I've produced. If a rivalrous resource is open to all, there is a risk that it will be depleted by the consumption of all.

This depletion of a rivalrous resource is the dynamic that biologist Garrett Hardin famously termed "the tragedy of the commons."[7] "Picture a pasture open to all," Hardin writes, and consider the expected behavior of "herdsmen" who roam that pasture. Each herdsman must decide whether to add one more animal to his herd. In making a decision to do so, Hardin writes, the herdsman reaps a benefit, while everyone else suffers. The herdsman gets the benefit of one more animal, yet everyone suffers the cost, because the pasture has one more consuming cow. And this defines the problem: Whatever costs there are in adding another animal are costs that others bear. The benefits, however, are enjoyed by a single herdsman. Therefore each herdsman has an incentive to add more cattle than the pasture as a whole can bear. As Hardin describes the consequence:

> Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit—in a world that is limited. *Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all.*[8]

This "tragedy" consumes talk about "the commons." "Ruin" is taken for granted as the destiny of those who believe in the "freedom of the commons." Hardheaded sorts thus scorn the rhetoric of undivided resources. Only the romantic wastes time wondering about anything different from the perfect control of property.

But obviously Hardin was not describing a law of nature that must apply to every good left in the commons. There is, for example, no tragedy for nonrivalrous goods left in the commons—no matter how many times you read a poem, there's as much left over as there was when you started. Nor is there always a tragedy even for rivalrous goods. As researchers have shown, in many different contexts, norms adequately limit the problem of overconsumption.[9] Communities work out how to regulate overconsumption. *How* and *why* are certainly complex questions. But *that some do* is undeniable.[10]

We therefore can't just jump from the observation that a resource is held "in common" to the conclusion that "freedom in a commons brings ruin to all." Instead, we must think empirically and look at what works. Where there is a benefit from leaving a resource free, we should see whether there is a way to avoid overconsumption, or inadequate incentives, without its falling under either state or private (market) control.

My central claim throughout is that there is a benefit to resources held in common and that the Internet is the best evidence of that benefit. As we will see, the Internet forms an *innovation commons*. It forms this commons not just through norms, but also through a specific technical architecture. The Net of these norms and this architecture is a space where creativity can flourish. Yet so blind are we to the possible value of a commons that we don't even notice the commons that the Internet is. And, in turn, this blindness leads us to ignore changes to the norms and architecture of the Net that weaken this commons. There is a tragedy of the commons that we will identify here; it is the tragedy of losing the innovation commons that the Internet is, through the changes that are being rendered on top.[11]

LAYERS

THE IDEA of the commons may be obscure, but the notion of "layers" is more easily recognized. The layers that I mean here are the different layers within a communications system that together make communications possible. The idea is taken from perhaps the best communications theorist of our generation, NYU law professor Yochai Benkler.[12] As he uses the idea, it helps organize our thought about how any communications system functions. But in organizing our thought, his work helps show something we might otherwise miss.

Following the technique of network architects, Benkler suggests that we understand a communications system by dividing it into three distinct "layers."[13] At the bottom is a "physical" layer, across which communication travels. This is the computer, or wires, that link computers on the Internet. In the middle is a "logical" or "code" layer—the code that makes the hardware run. Here we might include the protocols that define the Internet and the software upon which those protocols run. At the top is a "content" layer—the actual stuff that gets said or transmitted across these wires. Here we include digital images, texts, on-line movies, and the like. These three layers function together to define any particular communications system.

Each of these layers in principle could be controlled or could be free. Each, that is, could be owned or each could be organized in a commons. We could imagine a world where the physical layer was free but the logical and content layers were not. Or we could imagine a world where the physical and code layers were controlled but the content layer was not. And so on.

Consider some examples to make the possibilities real.

Speakers' Corner: Speakers' Corner is a place in London's Hyde Park where people who want to speak publicly gather on Sundays to deliver their speeches. It is a wonderfully English spectacle, ordinarily filled with both orators and loons. But the system of communication is distinctive: the physical layer (the park) is a commons; the code layer (the language used) is a commons, too; and the content layer is ordinarily unowned—what these nuts say is their own creation. All three layers in this context are free; no one can exercise control over the kinds of communications that might happen here.

Madison Square Garden: Madison Square Garden is another place where people give speeches or, more likely, play games. It is a huge stadium/auditorium near the center of Manhattan, owned by Madison Square Garden, L.P. Only those who pay get to use the auditorium; and the Garden is not obligated to take all comers. The physical layer is therefore controlled. But as with Speakers' Corner, both the code layer (the language) and the content layer (what gets uttered) are at least sometimes not controlled. They too can remain free.

The telephone system: The telephone system before its breakup was a single unitary system. The physical infrastructure of this system was owned by AT&T and its affiliates; so too was its logical infrastructure—determining how and who you could connect—controlled by AT&T. But what you said on an AT&T phone (within limits, at least)[14] was free: the content of the telephone conversations was not controlled, even if the physical and code layers underneath were.

Cable TV: Finally, think of cable TV. Here the physical layer is owned—the wires that run the content into your house. The code layer is owned—only the cable companies get to decide what runs into your house. And the content layer is owned—the shows that get broadcast are copyrighted shows. All three layers are within the control of the cable TV company; no communications layer, in Benkler's sense, remains free.

These examples suggest the range of ways of organizing systems of communications. No single mix is best, though the differences among the four are important. To the extent that we want a decentralized system of communications, unowned layers will help. To the extent that we want controlled systems of communications, owned layers will help. But the point of the scheme so far is not to make predictions. The point is simply to make clear the range, and that trade-offs within this range exist.

	Speakers' Corner	*Madison Square Garden*	*Telephone System*	*Cable TV*
Content	Free	Free	Free	Controlled
Code	Free	Free	Controlled	Controlled
Physical	Free	Controlled	Controlled	Controlled

Now, from the language I've used so far, you might think that the Internet is a communications system free all the way down—free, that is, at every one of Benkler's layers. It is not. What is special about the Internet is the way it mixes freedom with control at different layers. The physical layer of the Internet is fundamentally controlled. The wires and the computers across which the network runs are the property of either government or individuals. Similarly, at the content layer, much in the existing Internet is controlled. Not everything served across the Net is free for the taking. Much is properly and importantly protected by property law.

At the code layer, however, in ways that will become clearer below, the Internet was free. So too was much of the content served across the network free. The Internet thus mixed both free and controlled layers, not just layers that were free.

Our aim is to understand how this mix produced the innovation that we have seen so far and why the changes to this mix will kill what we have seen so far.

3

Commons on the Wires

[T]HE INTERNET is a network of networks. In the main, these networks connect over wires. All of these wires, and the machines linked by them, are controlled by someone. The vast majority are owned by private parties—owned, that is, by individuals and corporations that have chosen to link to the Net. Some are owned by the government.

Yet this vast network of privately owned technology has built one of the most important *innovation commons* that we have ever known. Built on a platform that is controlled, the protocols of the Internet have erected a free space of innovation. These private networks have created an open resource that any can draw upon and that many have. Understanding how, and in what sense, is the aim of this chapter.

PAUL BARAN was a researcher at the Rand Corporation from 1959 to 1968. His project in the early 1960s was communications reliability. The fear slowly dawning upon the leaders of the world's largest nuclear arsenal was that the communications system controlling that arsenal was vulnerable to the smallest of attacks. An accident, or a single nuclear explosion, could disable the ability of the commander in chief to command. Chaos—or worse—would be unavoidable.

Baran's task was to explore a more secure telecommunications system. His first step was to understand the system then in place. So he asked the then provider of telecommunications in America, American Telephone &

Telegraph, to see the plans for the AT&T network to determine whether the communications system was secure.

AT&T balked. Though Baran had the proper security clearance, and though the Defense Department supported his request, AT&T refused Baran's inquiry. They had studied the matter, AT&T reported. The system was secure.

This was "the Bell system." It is hard for us today to appreciate the power of such a company. This was not just a large company, or even a large company with a very large market share. This was a partner with the government, ruling telecommunications in America. It was therefore, in its own view of itself, the governor of communications. States and the Federal Communications Commission (FCC) might regulate it, but the information and cooperation to make that regulation possible came from AT&T. It had been, since interconnection began in earnest in 1912, America's telecommunications master.[1]

Things with telephones were not always this way. Indeed, the early history of telecommunications is essentially unrecognizable to us. Though the Bell companies held the first patents on telephone technology, once those patents expired, a vigorous competition emerged to bring telephone service to Americans. AT&T concentrated on businesses. "Independents" focused on residences. The competition produced a rapid expansion of coverage. "From 1900 to 1915, at least 45% of the U.S. cities with populations over 5,000 had competing, non-interconnected telephone exchanges. During the peak of the independent movement's strength, between 1902 and 1910, that percentage was more than 55%."[2]

Today we would not recognize the phone system that this early competition produced. Though the reach of the telephone network was great—in 1920, 38.7 percent of farms and 30 percent of residences had a telephone[3]—the networks did not interconnect. There was no guarantee that if your grandmother across town had a telephone, you, using yours, could call her. Thus when you purchased telephone service, your decision in part depended upon whom you wanted to call and what service they were likely to have.

The world was then with telephones as the world was with personal computers ten years ago,[4] or as the world with instant messaging is today. Though there was a dominant system (AT&T for phones; Microsoft/Intel for computers; AOL's AIM for instant messaging), there was vigorous competition among other systems (the "independents" for phones; Apple's Macintosh or IBM's OS/2 for computers; Yahoo! or MSN for messaging). This competition effectively pushed the dominant system to become better and different. Just as the windows of Macintosh pushed Microsoft to Windows, so too the rural service of the "independents" pushed AT&T to extend its reach to farmers.

After a while, however, AT&T grew weary of this competition. The view grew within the company that security would come only by merging with the competitors. From 1908 to 1913, the Bell system adopted a number of strategies to destroy the independents, including selective interconnection and acquisition of competitors. If it could not gain customers through direct competition, it would gain customers by purchasing competitors.[5]

Initially, this consolidation inspired skepticism among regulators and the public. AT&T was attacked as a monster seeking monopoly. But by the early 1920s, antitrust enforcement in the United States was waning. The spirit of the time favored consolidation and rationalization; competition was viewed as "ruinous." Thus AT&T was slowly able to secure agreements with the government that essentially permitted it to extend its reach while protecting it against antitrust review.

Paradoxically, AT&T's most effective weapon in this expansion was to offer competitors the ability to interconnect. Though our intuition is likely to tell us that it was the failure to interconnect that hampered competition, in fact, as economist Milton Mueller has effectively argued, it was a lack of interconnection that spurred competition.[6] As each independent interconnected to the AT&T system, any distinctive advantage it could offer as an independent disappeared. Consumers had no further interest in subscribing to it over AT&T; hence the drive to AT&T as universal provider was only increased. The network advantage of AT&T would grow relative to other independents; hence the power of AT&T's increasing monopoly was enhanced.

Independents at the time understood this dynamic. Associations of independents vigorously attacked the "traitors" among them that chose to interconnect with AT&T.[7] But these competitors were increasingly seen as inconveniences, by both regulators and the public. The idea of a world of "universal service"—meaning not a telephone in every house, but a system where every phone could reach every other phone—was too seductive.[8] So in 1913 the government entered into an agreement with AT&T that would secure its monopoly in telecommunications in America, even though it was sold as a solution to telecommunications monopoly.

Named after Bell vice president Nicholas C. Kingsbury, the "Kingsbury Commitment" required that Bell "stop acquiring independent phone companies and [. . .] connect the remaining independents to Bell's long-distance network."[9] Bell also had to divest its telegraph arm, Western Union. This stopped the company from its increasingly ravenous practice of acquisition, but it "did nothing to promote competition in either telephony or telegraphy."[10] The commitment did not force local exchanges to be more competitive. It did not require interconnection with other long-distance carriers. The solution, "in short, was not the steamy unsettling cohabitation that marks competition, but rather a sort of competitive apartheid, characterized by segregation and quarantine."[11] As a major treatise on telecommunications describes it:

> The Kingsbury Commitment could be viewed as a solution only by a government bookkeeper who counted several separate monopolies as an advance over a single monopoly, even absent any trace of competition among them.[12]

Monopolies are not all bad, and no doubt this monopoly did lots of good. AT&T produced an extraordinary telephone system, linking 85 percent of American homes at the peak of its monopoly power in 1965.[13] It spent billions of dollars to support telecommunications research. Bell Labs invented fiber optic technology, the transistor, and scads of other major technological advances. Its scientists earned at least half a dozen Nobel Prizes in physics.

And it attracted a certain kind of person. As Paul Baran described it:

> They were not motivated by making a lot of money. They were in the business to provide a service: Loyalty to the organization and help to the country providing the world's best communication. And that was their motivation and their belief. It was a religion, a pure religion. . . . In their mind, they were doing the right thing.[14]

These were not fat monopolists seeking to rob the nation of a quick buck. These were "soldiers of communications," for whom control and hierarchy were key. As one publication in 1941 put it:

> Because each of them has a part in this speeding of the spoken word, the thousands of men and women who are engaged in the telephone service in America are ever conscious of the fact that theirs is a high calling.[15]

AT&T in turn succeeded during its monopoly reign in attracting the very best telecommunications researchers. Baran attributes its success to "an absolutely brilliant compensation system,"[16] but the reason may well be that AT&T was the only show in town. As Baran describes, "[F]or years and years, that was the only place in the country that was doing work in telecommunications."[17] One could research different telecommunications systems, and one could in principle even develop other telecommunications systems. But there was nothing one could do with one's innovation unless AT&T bought it.

For much of the twentieth century, it was essentially illegal even to experiment with the telephone system. It was a crime to attach a device to the telephone system that AT&T didn't build or expressly authorize. In 1956, for example, a company built a device called a "Hush-a-Phone." The Hush-a-Phone was a simple piece of plastic that attached to the mouthpiece of a telephone. Its design was to block noise in a room so that someone on the other end of the line could better hear what was being said. The device had no connection to the technology of the phone, save the technology of the plastic receiver. All it did was block noise, the way a user might block noise by cupping his hand over the phone.[18]

When the Hush-a-Phone was released on the market, AT&T objected. This was a "foreign attachment." Regulations forbade any foreign attachments without AT&T's permission. AT&T had not given Hush-a-Phone any such permission. The FCC agreed with AT&T. Hush-a-Phone was history.

Hush-a-Phone is an extreme case.[19] The real purpose of the foreign attachments rule was, at least as AT&T saw it, to protect the system from dirty technology. A bad telephone or a misbehaving computer attached to the telephone system could, AT&T warned, bring down the system for the whole region. Telephones were lifelines, and they had to be protected from the experiments of an inquisitive nation. Rules such as the foreign attachments rules were intended to achieve this protection.

Whatever their intent, however, these rules had an effect on innovation in telecommunications. Their effect was to channel innovation through Bell Labs. Progress would be as Bell Labs determined it. Experiments would be pursued as Bell Labs thought best. Thus telecommunications would evolve as Bell Labs thought best.

BARAN UNDERSTOOD this. As a researcher at a Defense Department–supported lab, he knew how the "military" thought, and AT&T was military. Thus he had reason to be skeptical about the claims that the existing system would withstand a nuclear attack. He didn't believe AT&T really understood the threat. And if it did, he believed it simply didn't want anyone else understanding its weakness.

So he pushed AT&T to let him examine the system. It pushed back. And so, from sources unnamed, Baran secured a copy of AT&T's plans—the blueprints for the telecommunications system of the United States.

When he saw the plans, Baran knew AT&T was wrong. He was certain that the system it had built would not withstand a nuclear attack. The network was too concentrated; it had no effective redundancy. So he continued to press his idea for a different telecommunications system. He had a different design for telecommunications, and he wanted AT&T to help him build it.

This different model was not the Internet, but it was close to the Internet. Baran proposed a kind of packet-switching technology to replace the persistent circuits around which the telephone system was built. Under AT&T's design, when you called someone in Paris, a circuit was opened between you and Paris. In principle, you could trace the line of copper that linked you to Paris; along that line of copper, all your conversation would travel.

Baran's idea was fundamentally different. If you digitized a conversation—translating it from waves to bits—and then chopped the resulting stream into packets, these packets could flow independently across a network and create the impression of a real-time connection on the other end. As long as they flowed fast enough, and the computers at both ends were quick, the conversation encoded in this packet form would seem just like a conversation along a single virtual wire across the ocean.

Baran was probably not the first person to come up with this idea—MIT loyalists insist that that was Leonard Kleinrock.[20] And he was also not the only person working on the idea in the early 1960s. Independently, in England, Donald Davies was developing something very similar.[21] But whether the first, or the only, doesn't really matter for our purposes here. What is important is that Baran outlined a telecommunications system fundamentally different from the dominant design, and that different telecommunications system would have effected a radically different evolution of telecommunications.

BARAN PUSHED to get AT&T to help build this alternative design. AT&T said he didn't understand telephones. Over the course of many months, he attended classes sponsored by AT&T so that he would get with its program. But the more Baran saw, the more convinced he was. And in a final push, the Defense Department offered simply to pay AT&T to build the system. The government promised no risk; it wanted only cooperation. But even here, AT&T balked. As recounted in John Naughton's *A Brief History of the Future*:[22]

> [AT&T's] views were once memorably summarised in an exasperated outburst from AT&T's Jack Osterman after a long discussion with Baran. 'First,' he said, 'it can't possibly work, and if it did, damned if we are going to allow the creation of a competitor to ourselves.'[23]

"Allow." Here is the essence of the AT&T design, supported by the state-sanctioned monopoly. In "defend[ing] the monopoly,"[24] it reserved to itself the right to decide what telecommunications would be "allowed." As Baran put it, AT&T "didn't want anybody in their vicarage."[25] It controlled the wires; nothing but its technology could be attached, and no other system of telecommunications would be permitted. One company, through one research lab, with its vision of how communications should occur, decided. Innovation here, for this crucial aspect of modern economic life, was as this single organization would decide.

Now again, the point is not that AT&T was evil. Indeed, quite the contrary. We get nowhere in understanding how systems of innovation work when we personify organizations and imagine them responsible for social goals. AT&T had an obligation to its stockholders; it had an obligation to the government to assure consistent quality service. It was simply acting to assure that it met both of these obligations—maximizing its profits for its shareholders while meeting its obligations to the government.

But what's good for AT&T is not necessarily good for America. What AT&T was doing may well have made sense for it; its vision of telecommunications may well have made sense for the interests it understood itself to be serving. But AT&T's vision of what a telecommunications service should be is not necessarily what a telecommunications service should be. There is a possible—and in this case actual—conflict between the interests of a centralized controller of innovation and the interest in innovation generally.

Here the conflict was plain. If the Defense Department built a telecommunications system based on packets rather than circuits, then the efficiency of that system could in theory be much greater. When you're on a circuit-switched system, listening to your lover in Paris tell you about someone new, there's lots of downtime on the line—silence—that is just wasted bandwidth. If instead the system were packets, then the data from the downtime would be silence; it's easier to send the information necessary to reproduce silence than it is to hold open a line while silence happens. The system could better utilize the wires if the architecture enabled the sharing of the wires.

The owner of a legacy system built on a different model could well decide that this challenge was too dangerous. If a more efficient system came on-line, there would be strong pressure from the government to allow the exception; that exception would not be easy to limit; the corrosion of the existing model could be great. Monopoly control would be lost.

Thus it is completely understandable that a company like AT&T would not want to give birth to this new competitor, even if this new competitor would be better for communications as a whole. The natural desire of any company is to find ways to protect its market. And the chosen desire of a competitive market is to limit the ways in which a company can protect its market—but for most of the century, this chosen desire was not telecommunications policy. For most of the century, in this context and others that we will consider later on, the chosen desire of policy makers was to back up the desire of companies to architect and support systems that protected them against competition in the market. Competition was a bother; the vision of a telecommunications system was limited; and our telecommunications architecture—including, as we will see, broadcasting and radio—was architected to maximize the power and control of the few.[26]

AT A certain point, Baran understood. When the project was pushed into the Defense Communications Agency (DCA), Baran realized the project would be bungled. As he told author John Naughton:

> I felt that [DCA] could be almost guaranteed to botch the job since they had no understanding of digital technology, nor for leading edge high technology development. Further, they lacked enthusiasm. Sometimes, if a manager doesn't have the staff but has the drive and smarts to assemble the right team, one could justify taking a chance. But lacking skills, competence and motivation meant backing a sure loser.[27]

So Baran had the project pulled. There were not "the people at the time who could successfully undertake this project, [and they] would likely screw up the program. An expensive failure would make it difficult for a more competent agency to later undertake the project."[28] Thus, this architecture of control—centralizing innovation and protecting an existing model of doing business—would not be questioned by Baran's work. At least not then.

THE INTERNET is not the telephone network. It is a network of networks that sometimes run on the telephone lines. These networks and the wires that link them are privately owned, like the wires of the old AT&T. Yet at the core of this network is a different principle from the principle that guided AT&T. Like the principle Baran confronted, this principle affects what is allowed and what is not. And like the principle that Baran confronted, this principle has an effect on innovation.

First described by network architects Jerome Saltzer, David Clark, and David P. Reed in 1981, this principle—called the "end-to-end argument" (e2e)—guides network designers in developing protocols and applications for the network.[29] End-to-end says to keep intelligence in a network at the ends, or in the applications, leaving the network itself to be relatively simple.

There are many principles in the Internet's design. This one is key. But it will take some explaining to show why.

Network designers commonly distinguish computers at the "end" or "edge" of a network from computers within that network. The computers at the end of a network are the machines you use to access the network. (The machine you use to dial into the Internet, or your cell phone connecting to a wireless Web, is a computer at the edge of the network.) The computers "within" the network are the machines that establish the links to other computers—and thereby form the network itself. (The machines run by your Internet service provider, for example, could be computers within the network.)

The end-to-end argument says that rather than locating intelligence within the network, intelligence should be placed at the ends: computers within the network should perform only very simple functions that are needed by lots of different applications, while functions that are needed by only some applications should be performed at the edge. Thus, complexity and intelligence in the network are pushed away from the network itself. Simple networks, smart applications. As a recent National Research Council (NRC) report describes it:

> Aimed at simplicity and flexibility, [the end-to-end] argument says that the network should provide a very basic level of service—data transport—and that the intelligence—the information processing needed to provide applications—should be located in or close to the devices attached to the edge [or ends] of the network.[30]

The reason for this design was flexibility, inspired by a certain humility. As Reed describes it, "we wanted to make sure that we didn't somehow build in a feature of the underlying network technology . . . that would restrict our using some new underlying transport technology that turned out to be good in the future. . . . That was really the key to why we picked this very, very simple thing called the Internet protocol."[31]

It might be a bit hard to see how a principle of network design could matter much to issues of public policy. Lawyers and policy types don't spend much time understanding such principles; network architects don't waste their time thinking about the confusions of public policy.

But architecture matters.[32] And arguably no principle of network architecture has been more important to the success of the Internet than this single principle of network design—e2e. How a system is designed will affect the freedoms and control the system enables. And how the Internet was designed intimately affected the freedoms and controls that it has enabled. The *code* of cyberspace—its architecture and the software and hardware that implement that architecture—regulates life in cyberspace generally. Its code is its law. Or, in the words of Electronic Frontier Foundation (EFF) cofounder Mitch Kapor, "Architecture is politics."[33]

To the extent that people have thought about Kapor's slogan, they've done so in the context of individual rights and network architecture. Most think about how "architecture" or "software" or, more simply, "code" enables or restricts the things we think of as human rights—speech, or privacy, or the rights of access.

That was my purpose in *Code and Other Laws of Cyberspace.* There I argued that it was the architecture of cyberspace that constituted its freedom, and that, as this architecture was changed, that freedom was erased. *Code,* in other words, is a *law* of cyberspace and, as the title suggests, in my view, its most significant law.

But in this book, my focus is different. The question I want to press here is the relationship between architecture and innovation—both commercial innovation and cultural innovation. My claim is that here, too, code matters. That to understand the source of the flourishing of innovation on the

Internet, one must understand something about its original design. And then, even more important, to understand as well that changes to this original architecture are likely to affect the reach of innovation here.

SO WHICH code matters? Which parts of the architecture?[34]

The Internet is not a novel or a symphony. No one authored a beginning, middle, and end. At any particular point in its history, it certainly has a structure, or architecture, that is implemented through a set of protocols and conventions. But this architecture was never fully planned; no one designed it from the bottom up. It is more like the architecture of an old European city, with a central section that is clear and well worn, but with additions that are many and sometimes confused.

At various points in the history of the Net's development, there have been efforts at restating its principles. Something called "RFC 1958," published in 1996, is perhaps the best formal effort. The Internet was built upon "requests for comments," or RFCs. Researchers—essentially grad students—charged with the task of developing the protocols that would eventually build the Internet developed these protocols through these humble requests for comments. RFC 1 was written by Steve Crocker and outlined an understanding about the protocols for host ("IMP") software. Some RFCs specify particular Internet protocols; some wax philosophical. RFC 1958 is clearly in the latter camp—an "informational" document about the "Architectural Principles of the Internet."[35]

According to RFC 1958, though "[m]any members of the Internet community would argue that there is no architecture," this document reports that "the community" generally "believes" this about the Internet: "that the goal is connectivity, the tool is the Internet protocol and the intelligence is end-to-end rather than hidden in the network."[36] "The network's job is to transmit datagrams as efficiently and flexibly as possible. Everything else should be done at the fringes."[37]

This design has important consequences for innovation—indeed, we can count three:

- First, because applications run on computers at the edge of the network, innovators with new applications need only connect their computers to the network to let their applications run. No change to the computers within the network is required. If you are a developer, for example, who wants to use the Internet to make telephone calls, you need only develop that application and get users to adopt it for the Internet to be capable of making "telephone" calls. You can write the application and send it to the person on the other end of the network. Both of you install it and start talking. That's it.
- Second, because the design is not optimized for any particular existing application, the network is open to innovation not originally imagined. All the Internet protocol (IP) does is figure a way to package and route data; it doesn't route or process certain kinds of data better than others. That creates a problem for some applications (as we'll see below), but it creates an opportunity for a wide range of other applications too. It means that the network is open to adopting applications not originally foreseen by the designers.
- Third, because the design effects a neutral platform—neutral in the sense that the network owner can't discriminate against some packets while favoring others—the network *can't* discriminate against a new innovator's design. If a new application threatens a dominant application, there's nothing the network can do about that. The network will remain neutral regardless of the application.

The significance of each of these consequences to innovation generally will become apparent as we work through the particulars that follow. For now, all that's important is that you see this design as a *choice*. Whether or not the framers of the network understood what would grow from what they built, they built it with a certain philosophy in mind. The network itself would not control how it would grow. Applications would. That was the key to end-to-end design. As the inventor of the World Wide Web, Tim Berners-Lee, describes it:

> Philosophically, if the Web was to be a universal resource, it had to be able to grow in an unlimited way. Technically, if there was any centralized point of control, it would rapidly become a bottleneck that restricted the Web's growth, and the Web would never scale up. Its being "out of control" was very important.[38]

NETWORK ARCHITECTS Saltzer, Clark, and Reed were not the only people to notice the value of an end-to-end design. Quite independently, if later, the idea became apparent within AT&T itself. In the early 1990s, while trying to implement an improvement in the voice quality of the AT&T network (competition was beginning to have an effect: the effort was in response to the claim by Sprint that on its network you could hear a pin drop), Bell Labs researcher David Isenberg became increasingly frustrated with the "smart" network that AT&T was: at every layer in the distributional chain, the AT&T network had been optimized for voice telephony. But this optimization meant that any effort to change a layer in the AT&T distributional chain would disable other layers. Tweaking one part threw other parts into disarray. The system was in no sense "modularized," so change became impossibly difficult.

This led Isenberg to a treasonous thought: what if the problem was in the fundamental design of the network itself? What if the whole idea of a smart network was a mistake? What if a better design would be a "stupid network," with intelligence built into the devices, and the network itself kept as simple as possible?[39]

Isenberg had arrived through frustration at Saltzer, Clark, and Reed's fundamental insight: A simple, or, as Isenberg described it, stupid network would facilitate the greatest degree of innovation. A smart, or intelligent, network would perhaps be optimized for certain users, but its own sophistication would inhibit different or new uses not initially understood. By "build[ing] in assumptions about what the business proposition of the network is, you constrain what's possible."[40] The AT&T network was burdened by the intelligence built into it. A simpler design could beat the sophisticated design, at least along the dimension of innovation and change.

When Isenberg started to discuss his seditious thoughts, his employer, AT&T, was not happy. In the early summer of 1997, he was permitted to post a reply to an article that sang the virtues of smart networks. But soon after his article was posted, it was republished in many different places on the Net. Finally, in August 1997, Harry Newton published the article in his *Computer Telephony* magazine—without AT&T's permission. Isenberg became the enemy from within the AT&T network. He was told not to accept invitations from others to discuss his ideas. This control, understandably, became intolerable. As he told me, "[T]he AT&T pension became portable on January 1, 1998. I quit on January 2, 1998."

However disliked high up within the fortress, Isenberg's ideas began to catch on both outside and inside. The virtues of "stupid networks" became increasingly obvious, as the power of this simple network, the Internet, became undeniable. Isenberg's idea echoed the end-to-end principle: the two were the same, and both showed why the Internet would flourish.

* * *

THE INTERNET isn't the only network to follow an end-to-end design, though it is the first large-scale computer network to choose that principle at its birth. The electricity grid is an end-to-end grid; as long as my equipment complies with the rules for the grid, I get to plug it in.[41] Conceivably, things could be different. In principle, we might imagine that every device you plug into a grid would register itself with the network before it would run. Before you connected, you would have to get permission for that device. The owner of the network could then choose which devices to prohibit.

Likewise, the roads are end-to-end systems. Any car gets to enter the highway grid (put tolls to one side). As long as the car is properly inspected, and the driver properly licensed, whether and when to use the highway is no business of the highway. Again, we could imagine a different architecture: each car might first register with the grid before it got on the highway (the way airlines file flight plans before they fly).

But these systems don't require this sort of registration, likely because, when they were built, such registration was simply impracticable. The electronics of a power grid couldn't handle the registration of different devices; roads were built stupid because smart roads were impossible. Things are different now; smart grids, and smart roads, are certainly possible. Control is now feasible. So we should ask, would control be better?

In at least some cases, it certainly would be better. But from the perspective of innovation, in some cases it would not. In particular, when the future is uncertain—or more precisely, when future uses of a technology cannot be predicted—then leaving the technology uncontrolled is a better way of helping it find the right sort of innovation. Plasticity—the ability of a system to evolve easily in a number of ways—is optimal in a world of uncertainty.

This strategy is an attitude. It says to the world, I don't know what functions this system, or network, will perform. It is based in the idea of uncertainty. When we don't know which way a system will develop, we build the system to allow the broadest range of development.

This was a key motivation of the original Internet architects. They were extremely talented; no one was more expert. But with talent comes humility. And the original network architects knew more than anything that they didn't know what this network would be used for.

As David Reed describes, "[T]here were a lot of experiments in those days," and "we . . . realized that [there] was very little in common [other] than the way they used the network. There were sort of interesting ways that they used the network differently from application to application. So we felt that we couldn't presume anything about how networks would be used by applications. Or we wanted to presume as little as possible. . . . We basically said, 'Stop. You're all right' as opposed to running a bake-off."[42] These designers knew only that they wanted to assure that it could develop however users wanted.

Thus, end-to-end disables central control over how the network develops. As Berners-Lee puts it, "There's a freedom about the Internet: as long as we accept the rules of sending packets around, we can send packets containing anything to anywhere."[43] New applications "can be brought to the Internet without the need for any changes to the underlying network."[44] The "architecture" of the network is designed to be "neutral with respect to applications and content."[45] By placing intelligence in the ends, the network has no intelligence to tell which functions or content are permitted or not. As RFC 1958 puts it, the job of the network is simply to "transmit datagrams." As the NRC has recently concluded:

> Underlying the end-to-end argument is the idea that it is the system or application, not the network itself, that is in the best position to implement appropriate protection.[46]

In chapter 2, I introduced the idea of a commons. We can now see how the end-to-end principle renders the Internet an *innovation commons*, where innovators can develop and deploy new applications or content *without the permission of anyone else*. Because of e2e, no one need register an application with "the Internet" before it will run; no permission to use the bandwidth is required. Instead, e2e means the network is designed to assure that the network cannot decide which innovations will run. The system is built—constituted—to remain open to whatever innovation comes along.

This design has a critical effect on innovation. It has been, in the words of the NRC, a "key to the explosion of new services and software applications" on the Net.[47] Because of e2e, innovators know that they need not get the permission of anyone—neither AT&T nor the Internet itself—before they build a new application for the Internet. If an innovator has what he or she believes is a great idea for an application, he or she can build it without authorization from the network itself and with the assurance that the network can't discriminate against it.

At this point, you may be wondering, So what? It may be interesting (at least I hope you think this) to learn that the Internet has this feature; it is at least plausible that this feature induces a certain kind of innovation. But

why do we need to worry about this feature of the Internet? If this is what makes the Internet run, then as long as we have the Internet, won't we have this feature? If e2e is in the Internet's nature, why do we need to worry about e2e?

But this raises the fundamental point: The design the Internet has *now* need not be its design *tomorrow*. Or more precisely, any design it has just now can be supplemented with other controls or other technology. And if that is true, then this feature of e2e that I am suggesting is central to the network now can be removed from the network as the network is changed. The code that defines the network at one time need not be the code that defines it later on. And as that code changes, the values the network protects will change as well.

THE CONSEQUENCES of this commitment to e2e are many. The birth of the World Wide Web is just one. If you're free from geekhood, you are likely not to distinguish the WWW from the Internet. But in fact, they are quite distinct. The World Wide Web is a set of protocols for displaying hyperlinked documents linked across the Internet. These protocols were developed in the late 1980s by researchers at the European particle physics lab CERN—in particular by Tim Berners-Lee. These protocols specify how a "Web server" serves content on the WWW. They also specify how "browsers"—such as Netscape Navigator or Microsoft's Internet Explorer—retrieve content on the World Wide Web. But these protocols themselves simply run on top of the protocols that define the Internet. These Internet protocols, referred to as TCP/IP, are the foundation upon which the protocols that make the World Wide Web function—HTTP (hypertext transfer protocol) and HTML (hypertext markup language)—run.[48]

The emergence of the World Wide Web is a perfect illustration of how innovation works on the Internet and of how important a neutral network is to that innovation. Tim Berners-Lee came up with the idea of the World Wide Web after increasing frustration over the fact that computers at CERN couldn't easily talk to each other. Documents built on one system were not easily shared with other systems; content stored on individual computers was not easily published to the networks generally. As Berners-Lee writes:

> Incompatibility between computers had always been a huge pain in everyone's side, at CERN and anywhere else. . . . The real world of high-energy physics was one of incompatible networks, disk formats, and character-

encoding schemes, which made any attempt to transfer information between computers generally impossible. The computers simply could not communicate with each other.[49]

Berners-Lee thus began to think about a system to enable linking among documents—through a process called "hypertext"—and to build this linking on top of the protocols of the Internet. His ideal was a space where any document in principle could be linked to any other and where any document published was available to anyone.

The components of this vision were nothing new. Hypertext—links from one document to another—had been born with Vannevar Bush,[50] and made famous by Bill Atkinson's HyperCard on the Apple Macintosh. The world where documents could all link to each other was the vision of Robert Fano in an early article in the *Proceedings of the IEEE*.[51] But Berners-Lee put these ideas together using the underlying protocol of the Internet. Hyperlinked documents would thus be available to anyone with access to the Internet, and any document published according to the protocols of the World Wide Web would be available to all.

The idea strikes us today as genius. Its success makes us believe the idea must have been obvious. But what is amazing about the story of the birth of the World Wide Web is how hard it was for Tim Berners-Lee to convince anyone of the merit in the plan. When Berners-Lee tried to sell the plan at CERN, management was unimpressed. As Berners-Lee writes:

> What we hoped for was that someone would say, "Wow! This is going to be the cornerstone of high-energy physics communications! It will bind the entire community together in the next ten years. Here are four programmers to work on the project and here's your liaison with Management Information Systems. Anything else you need, you just tell us." But it didn't happen.[52]

When he went to a meeting of hypertext fans, he could get few to understand the "ah-ha" of hypertext on the Net. For years he wandered from expert to expert, finding none who understood the potential here. And it was only after he started building the Web out, and started informing ordinary people on a hypertext mailing list about the protocols he was developing, that the Net started to grow.

The experts didn't get it. Someone should put that on a bumper sticker and spread it around. Those controlling the resources of the CERN com-

puter lab wouldn't support the technology that would give the world the Web. Only those innovators outside of the control of these managers saw something of the potential for the Web's growth.

Berners-Lee feared that competing protocols for using the Internet would wipe away interest in the WWW. One protocol built about the same time was called Gopher. Gopher enabled the easy display of a menu of options from a site. When you went to a Gopher-enabled site, you would see a list of links that you could then click on to perform some function. Gopher was extremely popular as an Internet application—running on the Internet protocols—and use of Gopher took off in the early 1990s.[53]

But for the purposes that Berners-Lee imagined, Gopher was extremely limited. It would not enable the easy construction of interlinked documents. It was closer to a universal menuing system than a system for linking ideas. Berners-Lee was afraid that this inferior standard would nonetheless stick before the new and better WWW became well known.

His fear, however, was not realized, both because of something Berners-Lee did and because of something the creators of Gopher did—and both are lessons for us.

Berners-Lee was no bully. He was not building a protocol that everyone had to follow. He had a protocol for displaying content on the World Wide Web—the HTML language that Web pages are built in. But he decided not to limit the content that one could get through a WWW browser to just Web pages. Instead he designed the transfer protocol—HTTP—so that a wide range of protocols could be accessed through the WWW—including the Gopher protocol, a protocol for transferring files (FTP), and a protocol for accessing newsgroups on the Internet (NNTP). The Web would be neutral among these different protocols—it would in this sense interconnect.[54]

That made it easy to use the Web, even if one wanted to get access to Gopher content. But the second doing was much more important to the death of Gopher as a standard.

As Berners-Lee describes it, high off its success in populating the world with Gopher, the University of Minnesota—owner of the right to Gopher—suggested it might exercise its rights to charge for the use of the Gopher protocol.[55] Even the suggestion of this terrified developers across the world. (It was, Berners-Lee writes, "an act of treason."[56]) Would developers be hijacked by the university once they depended upon their system? How much would they lose if the platform eventually turned against the developers?

Berners-Lee responded to this by convincing CERN to release the right to the Web to the public. At first he wanted to release the protocol under the GPL, or General Public License (the "GNU General Public License," which we will see much more of in chapter 4). But when negotiations over that bogged down, he convinced CERN simply to release the rights into the public domain. Anyone had the right to take and use the protocols of the WWW and build anything upon them that they wanted.[57]

The birth of the Web is an example of the innovation that the end-to-end architecture of the original Internet enabled. Though no one quite got it—this the most dramatic aspect of the Internet's power—a few people were able to develop and deploy the protocols of the World Wide Web. They could deploy it because they didn't need to convince the owners of the network that this was a good idea or the owners of computer operating systems that this was a good idea. As Berners-Lee put it, "I had designed the Web so there should be no centralized place where someone would have to 'register' a new server, or get approval of its contents."[58] It would be a "good idea" if people used it, and people were free to use it because the Internet's design made it free.

THUS TWO networks—the network built by AT&T and the network we call the Internet—create two different environments for innovation. One network centralizes creativity; the other decentralizes it. One network is built to keep control of innovation; the other constitutionally renounces the right to control. One network closes itself except where permission is granted; the other dedicates itself to a commons.

How did we get from the one to the other? What moved the world governing our telecommunications system from the centralized to the decentralized?

This is one of the great forgotten stories of the Internet's birth. Everyone knows that the government funded the research that led to the protocols that govern the Internet.[59] It is part of the Internet's lore that it was the government that pushed network designers to design machines that could talk to each other.[60] The government in general, and the Defense Department in particular, had grown tired of spending millions for "autistic computing machines."[61] It therefore wanted some system for linking the systems.

Yet we are practically trained to ignore another form of governmental intervention that also made the Internet possible. This is the regulation that assured that the platform upon which the Internet was built would not turn against it.

The physical platform on which the Internet took off came prewired. It was the telephone wires that linked homes to homes. But the legal right to use the telephone wires to link to the Internet did not come preordained. That right had to be earned, and it was regulation that earned it. Nothing guaranteed that modems would be permitted on telephone lines. Even today, countries in Asia regulate the use of modems on telephone lines.[62] What was needed before the revolution could begin was permission to connect the Net to this net.

And what made that permission possible? What made it possible for a different use to be made of the telephone wires from that which AT&T had originally imagined?

Here a second kind of regulation enters the story. Beginning in force in 1968, when it permitted foreign attachments to telephone wires, continuing through the 1970s, when it increasingly forced the Bells to lease lines to competitors, regardless of their purpose, and ending in the early 1980s with the breakup of AT&T, the government increasingly intervened to assure that this most powerful telecommunications company would not interfere with the emergence of competing data-communications companies.

This intervention took many forms. In part it was a set of restrictions on AT&T's permissible businesses.[63] In part it was a requirement that it keep its lines open to competitors.[64] In part it was the general fear that any effort to bias communications more in its favor would result in a strong reaction from the government.[65]

But whatever the mix, and whichever factor was most significant, the consequence of this strategy was to leave open the field for innovation in telecommunications. AT&T did not control how its wires would be used, because the government restricted that control. By restricting that control, the government in effect created a commons on AT&T's wires.

In a way analogous to the technical requirements of end-to-end, then, these regulations had the effect of leaving the network open and hence of keeping the use of the network neutral. Once the telephone system was used to establish a circuit, the system was kept free for that circuit to send whatever data across it the user wished. The network thus functioned as a resource left open for others to use.

This is end-to-end operating at a different layer in the network design. It is end-to-end not at the layer determining the connection between two phones on the telephone system. That connection may well be formed by a system that does not comply with the end-to-end rule.

But once the circuit is connected, then the environment created by the mix of technical principles and legal rules operating upon the telecommunications system paralleled an end-to-end design at the network layer. This mix of design and control kept the telephone system open for innovation; that innovation enabled the Internet.

ARE THERE costs to the e2e design? Do we lose something by failing to control access to the resources—the bandwidth—of the network?

Certainly the Internet is not without its weaknesses. The capacity of the Net at any one moment is not infinite, and though it grows more quickly than the demand, it does at times get congested. It deals with this congestion equally—packets get transported on a first-come, first-served basis. Once packets leave one end, the network relays them on a best-efforts basis. If nodes on the network become overwhelmed, then packets passing across those nodes slow down.[66]

For certain applications, "best efforts" is not enough. Internet telephony, for example, doesn't do well when packets carrying voice get delayed. Any delay greater than 250 milliseconds essentially makes the system unusable.[67] And as content on the Net moves to real-time, bandwidth-demanding technology, this inability to guarantee quality of service becomes increasingly costly.

To deal with this problem, technologists have begun to propose changes to the architecture of the Net that might better enable some form of guaranteed service. These solutions generally pass under the title "Quality of Service" (QoS) solutions. These modifications would enable the network to treat different "classes" of data differently—video, for example, would get different treatment from e-mail; voice would get different treatment from the Web.

To enable this capacity to discriminate, the network would require more functionality than the original design allowed. At a minimum, the network would need to be able to decide what class of service a particular application should get and then treat the service accordingly. This in turn would make developing a new application more complex, as the programmer would need to consider the behavior of the network and enable the application to deal with that behavior.

The real danger, however, comes from the unintended consequences of these additional features—the ability of the network to then sell the feature that it will discriminate in favor of (and hence also against) certain kinds of

content. As the marketing documents from major router manufacturers evince, a critical feature of QoS solutions will be their ability to enable the network owner to slow down a competitor's offerings while speeding up its own—like a television set with built-in static for ABC but a clear channel for CBS.

These dangers could be minimized depending upon the particular QoS technology chosen. Some QoS technologies, in other words, are more consistent with the principle of end-to-end than are others.[68] But proponents of these changes often overlook another relatively obvious solution—increasing capacity.[69] That is, while these technologies will certainly add QoS to the Internet, if QoS technologies like the "RSVP" technology do so only at a significant cost, then perhaps increased capacity would be a cheaper social cost solution.[70]

Put differently, a pricing system for allocating bandwidth solves certain problems, but if it is implemented contrary to end-to-end, it may well do more harm than good.

That is not to argue that it *will* do more harm than good. We don't know enough yet to know that. But it raises a fundamental issue that the scarcity mentality is likely to overlook: The best response to scarcity may not be a system of control. The best response may simply be to remove the scarcity.

This is the promise that conservative commentator George Gilder reports. The future, Gilder argues, is a world with "infinite" bandwidth.[71] Our picture of the Net now—of slow connections and fast machines—will soon flip. As copper is replaced with glass (as in fiber optics) and, more important, as electronic switches are replaced by optical switches, the speed of the network will approach the speed of light. The constraints that we know from the wires we now use will end, Gilder argues. And the end of scarcity, he argues, will transform all that we do.[72]

There is skepticism about Gilder's claims about technology.[73] So, too, about his economics. The economist in all of us can't quite believe that any resource would fail to be constrained; the realist in all of us refuses to believe in Eden. But I'm willing to believe in the potential of essentially infinite bandwidth. And I am happy to imagine the scarcity-centric economist proven wrong.

The part I'm skeptical about is the happy progress toward a world where network owners simply provide neutral fat (or glass) pipe. This is not the trend now, and there is little to suggest it will be the trend later. As law professor Tim Wu wrote to me about Gilder's book:

> I think it is a "delta dollar sign" problem as we used to say in chemistry (to describe reactions that were possible, but not profitable). Private actors seem to only make money from infrastructure projects if built with the ability to exclude. . . . [H]ere in the industry, all the projects that are "hot" are networks with built-in techniques of exclusion and prioritization.[74]

Here *is* a tragedy of the commons. If the commons is the *innovation* commons that the protocols of the Net embrace, e2e most important among them, then the tragedy of that commons is the tendency of industry to add technologies to the network that undermine it. But this is an issue for the dark part of this book. For now, my aim is only brightness: to get you to see the commons that has been built through a set of protocols that defined the Internet that was.

THE INTERNET was born on a controlled physical layer; the code layer, constituted by the TCP/IP, was nonetheless free. These protocols expressed an end-to-end principle, and that principle effectively opened the space created by the computers linked to the Net for innovation and change. This open space was an important freedom, built upon a platform that was controlled. The freedom built an innovation commons. That commons, as do other commons, makes the controlled space more valuable.[75]

Freedom thus enhanced the social value of the controlled: this is a lesson that will recur.