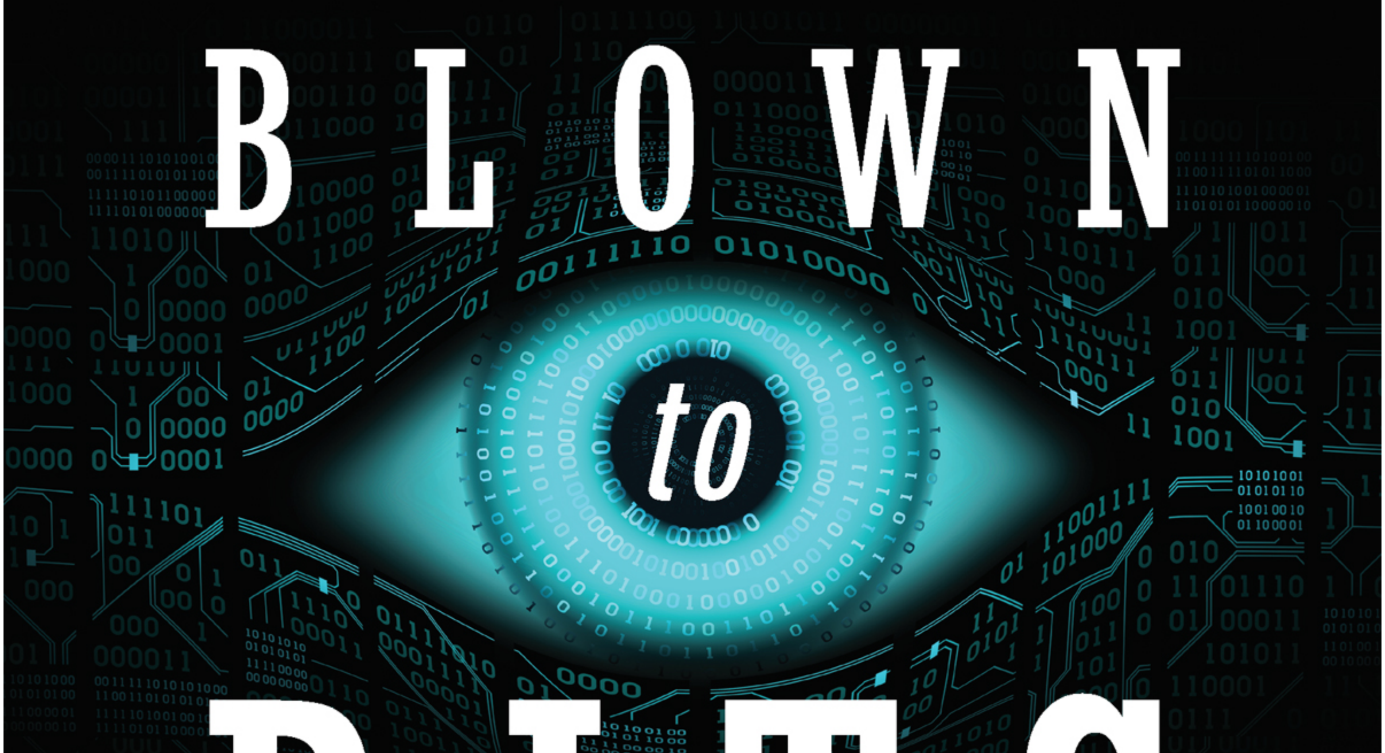
HAL ABELSON • KEN LEDEEN
HARRY LEWIS • WENDY SELTZER



BLOWN to BITS

YOUR Life, Liberty, and Happiness
After the Digital Explosion

SECOND EDITION

Blown To Bits

Your Life, Liberty, and Happiness After the Digital Explosion

Second Edition

Hal Abelson
Ken Ledeen
Harry Lewis
Wendy Seltzer

Pearson

CHAPTER 4

Gatekeepers

Who's in Charge Here?

Who Controls the Flow of Bits?

When the Telecommunications Workers Union went on strike against Telus, the leading telecommunications company in western Canada,[1] a discussion about strike-breaking sprung up on a pro-union website operated by a Telus employee. Then suddenly the site became inaccessible to anyone who was using Telus for Internet service. Telus subscribers could get bits originating in places from Afghanistan to Zimbabwe. They could get bits representing symphony orchestras and pornography. But if they wanted to see the discussion about resisting management's efforts to break the strike, they couldn't. Telus had taken the position that because the cables delivering the bits belonged to the company, it could choose to deliver or not deliver bits.

The union was enraged, and the legal experts were confused. It seemed to be clear enough that Telus couldn't cut off phone service to the union or its supporters if it wanted to, but the laws had been written in pre-Internet days. Was Telus within its rights to cut off Internet service in this case? The company noted that it had also blocked telusscabs.ca, which showed pictures of managers and employees who were going to work in spite of the strike. Telus said that it had a responsibility for their safety and felt compelled to protect them. But it turned out that Telus had blocked many more than these 2 sites. The web server hosting these 2 sites also hosted 766 other sites, including an alternative medicine site and a fundraising site for breast cancer research. In successfully blocking the 2 offending sites, Telus had blocked all the others, too.

Telus backtracked after making sure that no threatening material would be posted. But the incident—and others like it—raised questions that even today have no clear and generally accepted answers. Who controls what people can use the Internet to do?

The Open Internet?

Nobody was supposed to be in charge of the Internet. It wasn't even supposed to be the sort of thing that could be owned or controlled. It was meant to be more like a language that different people could use in any way they could imagine, to talk to each other or recite poetry or sing songs. It was to be like the "luminiferous ether," the invisible space-filling substance that physicists used to think must exist because light could not get from one place to another without it. The Internet was supposed to be a medium that could make communication possible by anyone to anyone and from anywhere to anywhere, but control of communication was assumed to be impossible because there would be no place to throttle it. Anyone who wanted to join in a conversation could—just by speaking the language of Internet protocols.

Ask Alex Jones if it has worked out that way. A leading American conspiracy theorist—or, as he considers himself, a "thought criminal against Big Brother"—Jones developed huge followings on YouTube, Facebook, LinkedIn, and other social networking sites. Millions of people followed his every word—and had every kooky rumor he promoted pushed to their mobile phones so they would see it instantly. And then suddenly many sites banned him. Apple stopped providing Jones's app to users. You can still find his website if you look for it, but Pinterest won't suggest it to you.

If this doesn't convince you that the Internet is not a participatory paradise, use the Internet to ask anyone in China what happened on June 4. Whether you use email, text messaging, or Weibo (China's version of Twitter), it is unlikely that your message will reach anyone because mention of June 4 has been censored thoroughly. Ask anyone in Hong Kong, and you won't even have to say which year; the Tiananmen massacre of June 4, 1989, is still vividly remembered more than 30 years on. But on the Internet of the mainland, the billions of users never talk about June 4. Mention it, and the conversation is snuffed out immediately rather than spreading like wildfire. And the government is not the only gatekeeper controlling the electronic sharing of information in China. When protesters in Hong Kong used an app called HKmap.live to organize themselves, the Chinese government became enraged, and Apple removed the app from its App Store. Google responded similarly to a request from the Hong Kong police to take down a game that enabled users to play the roles of protesters.[2]

Or ask Hasan Minhaj, an American comedian whose "Patriot Act" show is distributed via Netflix. His critical comments about Saudi Crown Prince Mohammed bin Salman can be watched almost everywhere in the world—but not where they would have the most meaning, in Saudi Arabia. The Saudi government demanded that the episode be taken down, citing a law that criminalizes the "production, preparation, transmission, or storage of material impinging on public order, religious values, public morals, and privacy, through the information network or computers." Netflix responded by saying it supported artistic freedom—but then took the video down anyway.[3] Artistic freedom for Minhaj proved to be incompatible with the personal freedom of Netflix employees, who might be subject to ten years' imprisonment or more for the vaguely defined crime of transmitting material impinging on public order.

Or ask anyone who sells anything. Google's search engine is used for more than 90% of Internet searches; Microsoft's Bing is in second place, at 3%. If you want to sell widgets and you're not on the first page of results when people search for "widget," it may be difficult to attract attention. "Google is the gatekeeper for the World Wide Web, for the internet as we know it," as lawyer Gary Reback put it. "It is every bit as important today as petroleum was when John D. Rockefeller was monopolizing that."[4] Google defended itself by responding that it is not a monopoly because Amazon search results also influence shopping outcomes. (Google might have noted that this is especially so because Amazon awards an "Amazon Choice" badge to products it favors.) But Google's defense only drives home the point that that the number of gatekeepers is tiny, even if it is modestly greater than one. There may be 1,000 small businesses making and selling widgets, but only the handful that appear on the first search results page are likely to get any Internet business—and they are competing for visibility against much bigger companies.

Or ask anyone in Browning, Montana, if anyone can use the Internet to communicate with anyone. On the Blackfoot Indian Reservation there, only 0.1% of the population has high-speed Internet. The cheapest Internet connectivity of any kind is at 10 Mbps and costs almost $780 annually[5]—in a place with a 35% poverty rate and a median annual household income of less than $22,000.[6] In most American cities, getting bits over the Internet is like turning on the tap to get water, and the same is true all over Finland and Japan. Internet service is abundant in places where it is profitable for private suppliers or where it is provided as a matter of public policy. Neither is true in Browning, where Internet service is all but unavailable.

Whatever the history, the theory, and the potential, the reality is that a few corporations and a few governments exert enormous control over what most people actually see over the Internet and what they can do with the Internet. If these enterprises and institutions don't provide the infrastructure to deliver your message, or if they otherwise deprioritize delivery of your ad, your news, or your political barbs, you might as well have shouted in the wilderness. Someone might hear what you say, but probably not many people will. That is exactly the opposite of the way the Internet was designed to work. The Internet has evolved from its publicly funded design as an open system with unlimited potential uses to a system in which a few private businesses hold near-monopoly control over each of its major aspects.

This chapter focuses on three kinds of Internet gatekeepers. The first are the controllers of the data pipes through which the bits flow. We'll call these the *links gatekeepers*. The second are the controllers of the tools we use to find things on the Web. We'll call these the *search gatekeepers*. And the third are the controllers of the social connections that are, for many of us, our most important use of the Internet. We'll call these the *social gatekeepers*.

The links gatekeepers control the physical media through which the bits flow, while the search gatekeepers and social gatekeepers control what those bits express—that is, they are content gatekeepers. But such distinctions are not as sharp as they might seem. Links gatekeepers may be able, for example, to censor or to favor certain content over others or certain customers over others. Content gatekeepers may enter the links market if they think it will be to their advantage to resist the near-monopoly control of the links gatekeepers—whether or not consolidating links and content control is in the more broadly construed public interest. Social gatekeepers have added search within their social platforms to undercut the near-monopoly control of the search gatekeepers.

In the United States, all three gatekeeping functions are largely in private hands. In other parts of the world, governments have assumed some of these gatekeeping roles. Familiar debates about private versus public services have played out as part of the Internet almost from its beginning. Competition among private parties drives down costs and improves quality, goes one familiar argument; but consolidation, argue others, results in efficiencies of scale that more than outweigh the negative effects of reduced competition. According to another narrative, the government should provide infrastructure of general benefit to the people, paid for through general taxation rather than private purchase; it should provide the ether, in the same way it provides roadways and mail service, equally to all. But such analogies only invite the question of whether the Internet really is more like the public roadways, on which anyone can drive, or like cable television or movie theaters, which are more accessible in urban areas than rural and not available at all to people who are unwilling to pay the fees.

The results of the various possible answers to such questions have been mixed and depend to some degree on fundamental questions of civic and economic goals. In authoritarian regimes, committed to social "harmony" at the expense of individual liberty, control content may be even more centralized than in the United States. On the other hand, substantial government investment in the grid itself outside the United States has resulted in far better connectivity in certain democratic and undemocratic countries alike. The debates about the right level of government investment and oversight of the Internet are no simpler than the story of government involvement in the delivery of postal mail, electricity, telephone service, education, or medical care. After quickly telling the story of how the open Internet fell under the control of oligopolies of gatekeepers, we'll raise the questions with which society is left about what, if anything, to do.

Let's start with how it all works.

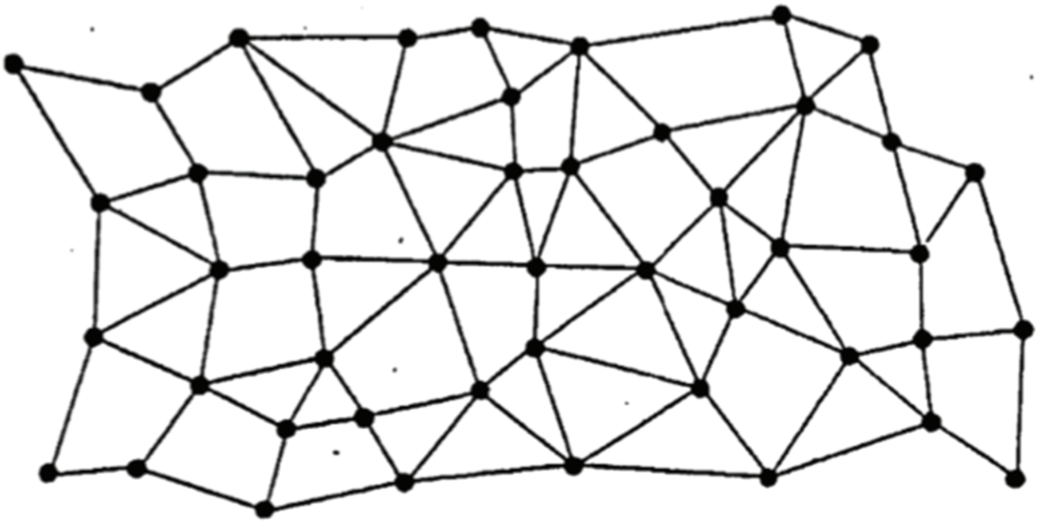
Connecting the Dots: Designed for Sharing and Survival

The Internet grew out of the ARPANET, a U.S. Department of Defense (DoD) computer networking project from the 1970s. Through its Advanced Research Projects Agency (ARPA, later DARPA), the DoD was directly or indirectly paying for state-of-the-art machines in many academic and national research laboratories. ARPA had two worries.

One worry was pedestrian: The agency was paying for big, expensive computers across the country, but there was no way for underutilized machines at one location to be put to work on problems researchers needed solved at other locations. So every researcher wanted the biggest possible computer, and a lot of computer time was going unused. Scientists could put their data on tapes and send it across the country by air freight, but there was no way to ship the bits without shipping atoms. So ARPA wanted improve utilization by networking the research computers it was funding.

ARPA's other concern cut to the heart of the military mission. The DoD had for some time been worried that its far-flung bases and ships might not be able to communicate if critical sites were destroyed in a nuclear war. In the early 1960s, the worry was whether the telephone network would survive an attack that knocked out a few key switching centers, where many long-distance telephone lines interconnected.

At that time, the researcher Paul Baran studied the properties of a decentralized network, one in which there were many junction points, each connected to only a few others. (The telephone network, by contrast, consisted of a small number of central switching stations connected to customers like spokes joining the hub of a wheel to its rim.) In Baran's proposed mesh-like network, there would be many paths between any two points, so knocking out any one of the junction points would not prevent other points from communicating. Baran imagined an irregular connection of switching points, like the one shown below in an illustration from his 1962 paper.[7]



Baran contributed a second important idea: If a switchpoint went down, another route could be found that did not go through it, as long as the switchpoint itself was neither end of the communication. By setting the switches the right way, communication between two points could be established along a particular path. But knocking out any of the points along the path would then interrupt that communication. So would an ordinary hardware failure at any of those intermediate points. It was important to protect the integrity of individual communications even as the network components failed in unpredictable ways.

Baran proposed to chop communications into small chunks of bits, what we today refer to as "packets." In addition to the "payload," a fragment of the communication itself, the packets would contain information identifying the source and destination (much like the address information on a postal letter), and also a serial number so that the destination node could reassemble them in the right order if they happened to arrive out of sequence. With this much information on the "envelope," the packets comprising a single communication did not need to follow the same path. If a portion of the network was unavailable, the network nodes could direct packets along a different path. Making this all work was not simple—how would the network nodes know in what direction to forward a packet?—but in principle, Baran's idea of a mesh-like interconnection and packetized communication would meet the military requirements for survivability.

Protocols: How to Shake Hands with Strangers

Once ARPANET was operational and connected a few dozen computers, it started to become clear that what needed to be connected were not individual computers but existing computer networks. Different ways of networking computers together could coexist, as long as the networks used some common language for communicating with each other. And in the 1970s and 1980s, different kinds of computer networks did exist, each using the standards of a different computer company. IBM had its SNA (Systems Network Architecture). Digital Equipment Corporation had DECnet. Apollo Computer connected its machines in a ring rather than a branching tree or mesh. Each company touted the advantages of its networking scheme, and some of the claims were valid for particular use cases. But none of the manufacturers had any incentive to make their machines interoperable with those of other manufacturers—until ARPA declared that it would pay for no more computers unless they could be interconnected. Starting from the success of the ARPANET, Vinton Cerf and Robert Kahn designed the protocols for interconnecting computer networks.[8] That is, they designed the Internet.

The Internet *is* its protocols. The Internet is not a machine or even a collection of machines. It's not some piece of software. It is a set of rules. Any person or organization can build hardware or write software that abides by those rules and become a functioning part of the Internet.

Protocols are communication conventions, like the convention that people shake their right hands. Having everyone greet each other by shaking their left hands would work equally well, but the established convention of right-hand shaking makes it possible for strangers to greet each other with no prior mediation. Internet protocols are the conventions by which different networks shake hands in order to pass information from one network to the other. Each network can operate as it wishes internally; only at the points where networks are connected together do the Internet protocols become relevant.

The decision to make ARPANET a packet-switched network simplified the Internet design considerably. Networks were connected to the Internet via connection points called *gateways*. If a gateway behaved as it should, information would flow through it. If it didn't behave properly, that caused no harm except to cut off that network from the rest of the Internet. No computer or network of computers needed permission to join the Internet. If it adhered to Internet standards, it could be understood by others and could interpret messages directed to it.

As we look at the Internet today, it seems varied and complicated: so many different kinds of content, so many different kinds of devices, and so many different kinds of connections. But it's all built on top of a single protocol, known simply as Internet Protocol (IP). It's the job of IP to get a single packet of around a thousand bits from one end of a communications network link to the other. The bits, as delivered, may contain errors; nothing physical ever works perfectly all the time. But errors can be recognized and, if necessary, dealt with. To get packets across the network, IP is used repeatedly, bucket brigade style, with each switching point receiving packets, checking them, and then dispatching them toward their intended destination.

The simplicity of the Internet design made it possible to build protocols on top of protocols to expand the Internet's utility. The earliest uses of the Internet were for logging in to time-shared computers remotely, for moving files from place to place, and for electronic mail. All these services required the data to arrive error free but not necessarily instantaneously. No one would notice if a file transfer or an email delivery took an extra fraction of a second, but having a single bit turn from a 0 to a 1 in transit could have catastrophic consequences. For such transfers, a protocol was developed to make sure that packets sent by the source were received correctly and reassembled in the correct order. Given the unreliability of the intermediate nodes of the network, this requires some bookkeeping at both source and destination. A packet, once received, is acknowledged by sending a special packet back from the destination to the source. The source runs a timer; if a packet sent is not acknowledged before the timer runs down to zero, the source figures that the packet has gotten lost somehow and retransmits it.

The details are tricky, but they are not important to the big picture. The result is that as long as the switches are making their best effort to pass packets along toward the destination, any message sent will be received in perfect order. The protocol that ensures such perfect transmissions is called Transmission Control Protocol (TCP). Because the underlying protocol for moving packets along single links of the network is IP, TCP/IP is the everyday name for the pair of conventions that make reliable communications possible across an unreliable network.

Since there are no rules for joining the Internet, it is fair to wonder about the "best effort" assumption. Couldn't a rogue actor try to sabotage the network by adding switches that would discard or misdirect packets rather than send them toward their intended destination? Indeed, that could happen, but neighboring switch points would eventually realize that the packets were not being delivered and would start avoiding the rogues. Internet routing heals itself by learning to avoid trouble spots—not just in case of hardware failures but also in case of malice. The Internet becomes more reliable the larger it becomes and the more interconnected it becomes.

The Internet worked because once a large enough number of parties agreed to use it in the intended way, bad actors could in effect be frozen out since they were few in number.

In addition to routing information and payload, packets also include some redundant bits to aid error detection. For example, a single extra bit might be added to every packet so that all transmitted packets have an odd number of 1 bits. If a packet arrives with an even number of 1 bits, it can be recognized as having been corrupted in transit and discarded so that the sender will retransmit it. Such extra bits can't guarantee that every packet received is correct. But they do guarantee correct transmission with overwhelming likelihood, and from a practical standpoint, this process suffices to make the likelihood of an undetected error less than the likelihood of a catastrophe, such as a meteor strike, at the source of the transmission.

IP, the best-effort packet forwarding protocol, can also be used for delivering messages imperfectly but quickly. For example, think about how the Internet might be used to convey voice communications, such as telephone calls. The voice signal can be chopped up into small time slices, each digitized and sent over the Internet. But instead of using TCP, which guarantees delivery but not timeliness, a different protocol, called UDP, is used. UDP accepts some packet loss in exchange for speedy delivery. Voice tones change slowly enough from one instant to the next that packets of a telephone conversation can be scrambled a bit, and some could be omitted entirely, without causing the conversation to become hard to understand—as long as the packets that make it arrive at about the right rate.

Many other protocols have been designed for other purposes and to layer on top of these, using TCP and UDP to carry out more complex communications tasks. For example, Hypertext Transfer Protocol (HTTP) is designed for communication between a web browser on a user's computer and a web server anywhere else in the world. HTTP relies on TCP to retrieve web pages on the basis of location information such as lewis.seas.harvard.edu. So without knowing the details of how TCP operates, anyone could set up a web server that would deliver web pages in response to incoming requests.

Who's in Charge?

There are no Internet cops to force anyone to format their packets as TCP, IP, UDP, or other protocols stipulated. No one will throw you off the Internet if you put your source address where the destination belongs and vice versa. If your packets don't conform to the standards, they just won't be delivered, or they will be ignored if they are delivered.

The Internet does, however, have some governing authorities. One is the Internet Engineering Task Force (IETF), which establishes the standards for Internet protocols. The IETF is a remarkable organization. It is open to anyone who wants to join, and it makes decisions on the basis of "rough consensus and running code." In earlier years, the IETF would meet in a room and determine "rough consensus" by having members hum. Substantial majorities were evident to everyone, and individual preferences enjoyed a level of anonymity because in a large group it is hard to tell who is humming and who isn't. Because most changes to the Internet protocols are enhancements and additions that do not change anything that is already working, there is rarely a need to make a positive decision under time pressure; the IETF can defer decisions, let people talk more while tweaking their proposals, and wait for true consensus to develop.

So the Internet is *open* by design. Anyone can join the decision-making process. You would not be wrong to be reminded of the communal utopianism of the 1960s. Early IETF member David Clark famously said, “We reject kings, presidents, and voting. We believe in rough consensus and running code”—the last phrase indicating the engineer’s preference for proofs of concept over concepts alone.[9] Of course, once the Internet became widely adopted, you would need to do a lot of persuading to develop a consensus to change anything that had become important to lots of people. But if you and I were halfway around the world from each other and decided to develop our own secret protocol for (say) trans-Pacific xylophone duets, we could happily program our computers to exchange IP packets that no one else would know what to do with. The IETF explains its role this way in its mission statement:

> When the IETF takes ownership of a protocol or function, it accepts the responsibility for all aspects of the protocol, even though some aspects may rarely or never be seen on the Internet. Conversely, when the IETF is not responsible for a protocol or function, it does not attempt to exert control over it, even though it may at times touch or affect the Internet.[10]

This is a remarkable statement, and it shows how badly the “Information Superhighway” metaphor breaks down when applied to the Internet. If the Internet is a highway, it is one in which motor vehicles voluntarily adhere to certain conventions so they can share the highway safely, but bicyclists and skateboarders are welcome to use the roadways, too—though at their own risk.

The Internet is open in another direction as well. Just as IP serves as the base layer for a hierarchy of protocols, IP itself is a logical, not physical, protocol. Internet packets can be transmitted on copper wire, through fiber-optic cables, or by radio waves. If you are an ordinary personal computer user buying something on Amazon, it is likely that the packets going back and forth between you and Amazon pass through all three and more, as they move from your computer to your wireless router, to your ISP, through the Internet, into Amazon’s corporate network, and to one of Amazon’s computers. Whenever engineers come up with a new way to move bits through physical media, they can also develop an implementation of IP that runs on that physical medium. There is even a carrier-pigeon protocol that could, in principle, be used to implement IP.

IP, the format in which all packets pass through the Internet, plays a role like the design of the ubiquitous 120V electric outlet, with three holes of specified shape and dimensions. The electric source on one side of the outlet may ultimately be a hydroelectric dam hundreds of miles away, solar panels only a few feet away, or battery storage. As long as the electricity conforms to the standards, the outlet is doing its job. The devices that get plugged into the outlet can be refrigerators, toothbrushes, vacuum cleaners, or dental drills.

As long as a device is fitted with the right plug and is designed to run on standard alternating current, it will work. In the same way, Internet Protocol acts as a universal mediator between applications and physical media.

In fact, the standardization of IP is the reason the Internet has so many uses that were not initially anticipated. Zoom and Facetime—Internet applications for connecting people via live audio and video links—were built on IP, even though there was absolutely nothing about such services in the original Internet design. The inventors of the Internet telephone system Skype—a small group of Scandinavian and Estonian engineers—just needed to adapt the Internet protocols to their purposes. And they did not need to ask the permission of the IETF or any other authority to start using Skype or to encourage others to start paying them to use it.

The Internet Has No Gatekeepers?

Now it has always been an overstatement to say that the Internet has no gatekeepers, but it is less true now than it used to be. As we will soon see, in some countries, governments are the primary gatekeepers, and in others, such as the United States, private corporations assume gatekeeper roles. Let's start with the forms of gatekeeping that have long existed.

Names to Numbers: What's Your Address?

The first fact of Internet life is that it does no good to be "on" the Internet if no one can find you. Packets flowing through the Internet have numeric addresses. Some entity has to translate the symbolic names—like cornell.edu and Skype.com—into numbers and keep track of which connecting points have which numbers.

The Internet Corporation for Assigned Names and Numbers (ICANN) is the body that decrees which numeric addresses are assigned to Cornell University or the nation of Australia. It oversees publication of electronic directories in such a way that anyone sending an email to the address president@cornell.edu or retrieving a web page from an address such http://anu.edu.au (the home page of the Australian National University) is directed to the correct place on the Internet. The translation tables, from letters to numbers, are held on Domain Name System (DNS) servers, which other computers consult in order to look up the numeric addresses to insert in the "destination" field of IP packets before they are launched into the Internet. If the Internet has a single vulnerability, it is control over the DNS servers. Does the island nation of Tuvalu get its own Internet top-level domain, like .au for Australia? (It does—and a very valuable one at that. It's .tv, and the nation, which used to make money from selling postage stamps, now gets some revenue by letting video sites like twitch.tv use the .tv extension.) Who decides whether Coca-Cola is entitled to cocacola.com or, for that matter, cocacola.sucks? It should be no surprise that such high-stakes questions trigger strong interest from governments and multinational corporations. Such questions generally cannot be settled with anonymous humming or any similarly inclusive process.

Nonetheless, such territorial disputes get resolved without force and without fracturing the network. Indeed, from the first few machines connected as ARPANET in 1969, the number of connected devices has grown into the billions today. Any of them can, in principle, connect to any of the others.[11] The serious gatekeeping problems of the Internet lie elsewhere.

Links Gatekeepers: Getting Connected

The Internet isn't very useful if you can't get connected.

If you drive west from Boston across the northern United States, the offerings at supermarket deli counters change around the time you hit Iowa. Suddenly, they feature gelatin salads in great variety, incorporating various chopped fruits, colorful layers, and creamed toppings. Sometimes the gelatin is molded around fish or meat. The fashion persists across the Great Plains and up into the Rockies, but it disappears on the downward slopes. By the time you reach the Pacific Ocean, Jell-O is again mostly for children and hospital patients. The love of Jell-O concoctions in the rural heartland is so prevalent that these dishes are commonly featured on the covers of food magazines available at the checkout counters of coastal and urban supermarkets even though no one shopping there would dream of serving such a thing. Among the coastal and urban elites, gelatin salads are considered unsophisticated.

The midwestern fondness for gelatin salads is fading now, as food culture, like the rest of American culture, is becoming geographically homogenized. In some areas, grandma's gelatin salad recipes are remembered like handmade straw hats and calico dresses—as artifacts of a rural past out of place in more advanced times. But gelatin salads were not taken West in covered wagons; that would have been impossible since creating them requires refrigeration. They are instead a byproduct of a twentieth century technological revolution: rural electrification. Gelatin salads were a delicacy on farms because you couldn't make them if your farm wasn't electrified. People who served gelatin salads also had electric well pumps and electric lights. Serving Jell-O proved you were technologically advanced.

The economics of diffusing electricity and of diffusing bits through wires or cables are similar. It is expensive to lay cables over long distances, and if there are few customers at the end of the line, it's not cost-effective to lay the cable. It’s also expensive to pull wires to many individual residences if they are far apart. The profits are much higher for wiring up a city because once the line is brought down a street, every housing unit on the street becomes a customer, and the distance from the main line to the customers tends to be short. Indeed, cities were electrified first, and primarily for street lighting, which did not require installation of wiring inside private buildings. All the other uses of electricity—for interior lighting, refrigerators, washing machines, dishwashers, and radios—were derivative of the use for street lighting. To this day, a stroll down Beacon Street near Fenway Park in Boston takes you past a structure that reads “The Edison Electric Illuminating Company.”

It would have made no sense to diffuse electricity nationally just so people could have Jell-O. In fact, there was no such thing as an electric home refrigerator when the first electric streetlights lit up. Household electricity was what Jonathan Zittrain calls a *generative* technology.[12] Once the infrastructure was in place, creative people began dreaming up uses for it, and whole new technologies developed that could not have existed without it. In the process, some old industries died, at enormous costs to those who had profited from them. “Ice box” is at best a nostalgic phrase for a refrigerator today, but a century ago there was an enormous industry devoted to cutting the frozen surface of lakes into blocks, shipping them long distances, and distributing them to American homes. Generative technologies are also destructive technologies.

The diffusion of the Internet has followed much the same stages as the diffusion of electricity once did. Lighting was the killer app for electricity; gelatin salads were the cat videos of the electric grid. And yet the U.S. experience with the Internet has been very different than it was with the electric grid.

The United States was electrified quickly and ubiquitously, but the fast build-out would not have happened without an impetus from the federal government. Wiring remote areas was unprofitable, and it would have been even more unprofitable for a competitor to lay a second cable to serve the same customers. So electricity was readily available in cities but extremely costly in remote areas, if it was available at all.

Franklin Roosevelt could see the difference in the price of electricity when he retreated from New York and Washington, DC, to Warm Springs, Georgia, where he sought respite and spa treatments for his paralysis. He conceived the Rural Electrification Act in Warm Springs and signed it into law in 1936. The initiative stimulated not only the diffusion of electricity but the invention of new ways for ordinary consumers to use it.

In the early 1920s, fewer than 1% of U.S. households had electricity. Six years after the signing of the Rural Electrification Act, on the heels of a terrible economic recession, half of U.S. households were electrified. By 1960, virtually the whole country had electricity.

To make a sensible comparison between diffusion of the Internet and diffusion of electricity, we need to define our terms. Electricity as delivered to the home became standardized: In the United States, electricity is alternating current, 60 Hz frequency and 120V. These standards are analogous to IP for the Internet. They guarantee that the same appliances can be plugged into outlets anywhere in the country.

But there is another important parameter: the *amount* of electric energy used by an appliance or a household. The rate at which electricity is used is called *power* and is measured in watts or kilowatts (for thousands of watts). The amount of energy used is measured in kilowatt-hours; a kilowatt-hour (kwh) is the amount of energy consumed by using it for an hour at a rate of 1,000 watts. Electric codes were standardized so that household circuits can handle around 2,000 watts; if you use much more than that, a circuit breaker will trip, and the wire might melt if there were no fuse or breaker. The wiring in an old house may have to be upgraded when an owner wants to use more electric equipment or more powerful electric equipment, such as air conditioning or a hot tub. On the other hand, new equipment tends to use less power than old, so average consumption per household has increased only slowly over time. The electric utilities that actually supply the power may have to upgrade their distribution grid to keep up with demand, but a combination of consumer pressure and federal standards generally make it rare in the United States to have "brownouts," when a whole city or neighborhood has insufficient power. In the United States, electricity is, for the most part, a successfully regulated industry.

The analog of power for the Internet is bit rate, and here the experience of the Internet and the electric grid have diverged significantly. Provisioning of Internet connectivity has been left almost entirely to the private sector, with minimal government regulation and minimal government support. Almost nowhere does serious competition exist, so consumers cannot switch to better providers. The monopoly provider of Internet services may offer a choice of speeds, but the higher speeds are likely to be exorbitantly expensive. In a word, rather than providing high-speed Internet, most suppliers of Internet services try to convince us that we have it already, and the government is assisting in their deceit rather than prodding them to improve their services.

Where Are the Bottlenecks, and What Counts as "High Speed"?

The rate at which bits complete their transit—for example, from a web server somewhere to the browser running in your home computer or from your office computer to the video chat room at your London headquarters—is the slowest of the rates of any of the links along the way. The rate at which bits flow through a link is affected by some physical parameters—the electrical or electromagnetic properties of the copper or glass of which the link is composed—and how heavy the traffic is. If your communication has the link to itself, it can utilize all the bits-per-second of which the communication medium is capable, but if it has to share that capacity with a million other transmissions, yours may get only a millionth.

Think again of your home computer retrieving a web page from a server belonging to some big company. Your request has to get to the corporate server, and the packets comprising the web page have to get back to your home computer. Simplifying greatly, you can think of the bits in your request making three hops. They have to get from your computer to the outside wall of your house; then from your house to the "backbone," the long-haul cables that crisscross the country; and then across the hundreds or thousands of miles of the backbone. The connection from your house to the backbone is commonly referred to as the "last mile." The same hierarchy is, in principle, traversed at the other end, except that Amazon and Google are connected directly to the backbone because of their enormous capacity needs. If you were communicating with someone else sitting at a home computer, the bits would have to traverse the "last mile" to that person's house.

Inside a house, most people use Wi-Fi, a kind of short-distance radio communication. Newer Wi-Fi technologies can reach gigabit speeds, but in practice connections are likely to be slower because of interference or obstructions. Power users may still have their houses wired so they can connect their computers using Ethernet cable rather than wireless.

But slow wireless may be fast enough if the last mile is slow anyway. And in the United States, it almost certainly is slow. By global standards, what is described in the United States as "high-speed Internet" just isn't.

The backbone of the Internet is fiber-optic cable. Fiber is amazing stuff; the glass itself has almost limitless information-carrying capacity. Its actual capacity is limited not by the glass but by the electronics that connect the network at the switching points. The electronics are constantly being improved; fiber, once installed, is never replaced (unless it breaks—for example, because a fishing trawler snags it).[13]

In some parts of the world, the last mile is also fiber, so those amazing information capacities go right to the doors of homes and businesses. In Singapore and Sweden, virtually everyone has access to Internet speeds in the billions of bits per second. By contrast, perhaps 15% of Americans have fiber connections to their homes, and the percentage is not increasing. Most of us are connected by legacy telephone wiring, using so-called DSL service, or by the coaxial cable that was installed to bring cable TV to our houses. Even DSL service is being phased out in some places as unprofitable. And the cable and telecommunications providers have effectively divided up the map between them; in few places can a consumer choose between cable and DSL service, much less between multiple cable services.

The Internet itself does not connect to your home or office or cell phone. Ordinary people connect their devices not to the Internet but to an Internet service provider (ISP). Since Internet packets can travel over a variety of physical media, there is in principle no limit to the number of ISPs that might carry packets into and out of your house, for a price. The reality is very different. In the United States, it is likely that your home ISP is AT&T, Time Warner, Comcast/Xfinity, Verizon, or Charter. The reason there are so few is that each of these is either a telecommunications or a cable television company, and they are using the wiring or fiber-optic cable that they have already brought to your house to provide telephone or television service. Wireless Internet access is also possible—by which is meant not the Wi-Fi connection between your computer and a router, which is then connected to your ISP, but a wireless connection directly to the ISP. Connection via satellite is possible in rural areas where no form of wired connection has been installed, but satellite Internet is both slow and expensive. Cell phones connect to the Internet via the cellular telephone network, but the cellular network is not a viable option for home use. And so called 5G radio signals, which are being billed as the future of Internet connectivity, travel only short distances. So a 5G infrastructure is realistic only in densely populated areas, where it is possible to install many hubs economically.

Americans are bombarded with advertisements for “high-speed Internet,” but in reality, even the government definition of “high speed” is deceptive. The last time it was updated was in 2015, when—over the objections of Internet service providers—the FCC raised the standard from 4 Mbps to 25 Mbps. That is one-fortieth of the gigabit speeds that are now standard in Japan and Sweden and which even China is diffusing in rural areas. And the U.S. 25 Mbps standard applies only to download speeds, as though the Internet were basically a broadcast medium for consumers to receive Netflix movies. A great many applications, from video chats to transfer of medical imagery, require high speeds in both directions. Businesses of every kind are dependent on Internet connectivity for both uploads and downloads; they use the Internet to get information about their products and services, and indeed the products and services themselves, to their customers. So it is very difficult to start a business in an area where connectivity is poor or limited to fast downloads and slow uploads. And yet in the United States, the Internet has been optimized as a replacement for television—as a way for the few to supply content to the many.

A second form of creative semantics distorts government statistics about the availability of “high-speed Internet” in the United States. The government considers a census tract—one of the some 75,000 geographic areas into which the United States is divided for census purposes—to have high-speed Internet if even a single household has access to it, regardless of the price and regardless of whether any household has actually signed up for a connection. Thus government estimates of the diffusion of high-speed Internet are wildly inflated.

And price matters. In large parts of the world, gigabit speeds are available for less than $50 per month. In the Boston area as we write, it's $70 per month in the restricted areas where it is available at all—and requires a 24-month contract.

The typical American has only one or two realistic ISP choices. More than 30% of U.S. households have no providers of Internet service at 25 Mbps or better, and fewer than a quarter of U.S. households have more than one choice.[14]

For the most part, diffusion of Internet connectivity has been left to the private sector in the United States. In fact, rules in 26 states hinder or prohibit local governments from offering Internet connectivity, forcing individuals to whom no affordable household service is available to connect at public libraries or fast food restaurants.[15] Montana's code[16] is typical: "Except as provided in subsection (2)(a) or (2)(b), an agency or political subdivision of the state may not directly or through another agency or political subdivision be an internet services provider." So unless all the private ISPs pull out, the people of Browning, Montana, are stuck with the poor-quality but expensive private offerings. No amount of municipal entrepreneurship can help its people out. The lobbyists from the private telecommunications firms got to the state legislature first.

Now, of course, there are reasons to keep governments from competing with private companies. The arguments are familiar. Competition drives down prices and improves quality. Taxpayer dollars should not be used to undercut private vendors. Governments should not interfere with free markets.

But there just aren't enough people at the other end to pay for the long-haul connection. This is no different from the situation of rural free delivery of postal mail—which became the law in 1893 and reached Billings in 1902—or the electric grid in the 1930s and 1940s. Connecting the country requires viewing electronic communications like electricity or postal mail: It would have to be available to everyone at an affordable price. That principle is in fact not generally accepted. Instead, the operative metaphor for the Internet is television or a multi-screen movie theater. The dominant Internet service providers see the Internet as a way of connecting active content providers to passive content consumers. That is why ISPs offer packages that encourage downloads and limit uploads.

In the absence of the kind of push from the government that has resulted in far better Internet service in South Korea, Switzerland, and even largely rural Finland than in the United States, why hasn't competition driven down prices and quality up?

Some would argue that corporate greed, collusion, and corruption are to blame, and while that perspective may have some validity in some cases, the reality is that communications networks grow and consolidate almost organically for reasons of efficiency. Paul Baran himself anticipated that this phenomenon would affect computer networks several years before he designed one of the earliest networks. In testimony before Congress on electronic privacy in 1966, he said,[17]

> Our first railroads in the 1830's were short routes connecting local population centers. No one sat down and laid out a master plan for a network of railroad rails. With time, an increasing number of such separate local systems were built. A network gradually grew as economic pressure caused the new links to be built to span the gaps between the individual routes.
>
> We didn't start to build a nationwide telegraph network in the late 1840's; only independent telegraph links. But it was not long before we had an integrated nationwide network. Even the name, Western Union, recalls the pattern of independent links joined together to provide a more useful system.
>
> We didn't start to build a nationwide telephone system in the early days of the telephone in the 1890's. Yet, today we have a highly integrated telephone network.
>
> Such patterns of growth are not accidents. Communications and transportation are services that historically tend to form "natural monopolies." The reason is well understood. It's cheaper to share use of a large entity than to build your own facilities. Hence, if you were to look at the earth, say, from the far-off vantage point of the moon, it would appear that the growth of these integrated networks out of individual pieces is almost biological.

So it's not very complicated. It is more valuable to be part of a big network than a small one, and the bigger the network, the more valuable it is to be part of it. In the absence of pushback from some social structure with the authority to resist mergers, consolidations, buyouts, and strategic corporate decisions to gain control over network traffic, communication networks will grow larger and fewer with time. Such monopolization is not necessarily against the public interest—as long as the public interest is at the table when distribution and pricing decisions are being made. Today, they rarely are.

Can the Letter Carrier Decide What Mail to Deliver?

The story of Telus and its striking workers with which this chapter began demonstrates that the dichotomy between links and content is unhelpful when the links gatekeeper takes on the role of content gatekeeper. The notion that Internet service providers should not be deciding which bits to deliver is known as *net neutrality*. In principle, it can seem simple and inarguable; after all, we don't want the phone company deciding which conversations it is going to allow to happen over its voice lines. True, when customers don't pay their bills, their phone service can be cut off, but even that is rare because society generally recognizes—or used to recognize—that phone service is important to daily life. But the Internet is not exactly like the telephone network.

About the same time as Telus was shutting down pro-union websites in Canada, a small North Carolina ISP by the name of Madison River Communications shut down Vonage, which offered Voice over IP (VoIP) service. Using the Internet to deliver live voice conversations would have seemed crazy at the birth of the Internet because the network was too slow and the computers connected to it couldn't keep up with the flood of packets in order to reassemble them into comprehensible speech. But times change. As link speeds improved and new protocols—based on IP—were optimized for voice communications, a systematic difference between telephone and Internet services intervened. Telephone companies charged extra for long-distance service; Internet service providers didn't care where packets were coming from or going to. They might charge more for higher data rates but not for more distant destinations. Inevitably, VoIP software—Skype was the earliest commercial success in this space—was developed to replace telephony by Internet communications and make long-distance "calling" virtually free for anyone with an Internet connection. Vonage was using Madison River's data service to undercut Madison River's phone service.

When Vonage was blocked, the company complained to the Federal Communications Commission, which has jurisdiction over telephone services. The case was resolved when Madison River agreed to pay a fine and not block VoIP for three years, but this resolution left more questions than answers in its wake. What if Madison River had been a cable company offering Internet services rather than a phone company? On the other hand, what if Madison River had been big enough to fight the FCC in court? It was not at all clear that the FCC had the congressional authority to back up its strong-arm tactics on the way even telephone ISPs were picking and choosing what bits to deliver.[18]

Matters came to a head in 2008, when the FCC ordered the ISP Comcast to stop "throttling"—that is, slowing down—BitTorrent, a peer-to-peer file-sharing service heavily used for delivering movies to the home. Comcast was profitably delivering movies over the same cable it was using to deliver Internet service, so BitTorrent was undercutting its video delivery business. Comcast successfully sued the FCC, establishing that indeed the FCC lacked the authority to regulate its Internet service business. This decision kicked off a net neutrality debate that has raged for more than a decade.

The details are complex. In a nutshell, pro-neutrality voices have argued for consumer choice and freedom; opponents have argued that market forces would resolve any tensions, an argument greeted skeptically by those noting how little competition exists in the ISP space. In the United States, net neutrality rules were instituted during the Obama administration and repealed during the Trump administration. Many other nations have adopted net neutrality in principle, but some of them allow usage-based billing, which may have the effect of rendering certain applications, such as watching movies, unacceptably expensive, thus achieving the same result—prioritizing other means of delivering movies to the home—that Comcast had achieved by throttling peer-to-peer services in 2008.

Search Gatekeepers: If You Can't Find It, Does It Exist?

Prescient as Baran was, he could not have anticipated the extent to which, as communication networks became accessible to everyone, control over the information they carry would also tend to fall into a small number of private hands. Search technology was a surprising development of the 1990s; it is now hard to imagine a world without it. And yet it is *not* hard to imagine a world in which Google does not control most of the searches in the Western world. It has just turned out that way, with troubling consequences.

Found After 70 Years

Rosalie Polotsky was 10 years old when she waved goodbye to her cousins, Sophia and Ossie, at the Moscow train station in 1937. The two sisters were fleeing the oppression of Soviet Russia to start a new life. Rosalie's family stayed behind. She grew up in Moscow, taught French, married Nariman Berkovich, and raised a family. In 1990, she emigrated to the United States and settled near her son, Sasha, in Massachusetts. Rosalie, Nariman, and Sasha always wondered about the fate of Sophia and Ossie. The Iron Curtain had utterly severed communication among Jewish relatives. By the time Rosalie left for the United States, her ties to Sophia and Ossie had been broken for so long that she had little hope of reconnecting with them—and, as the years wore on, she had less reason for optimism that her cousins were still alive. Although his grandfather dreamed of finding them, Sasha's search of immigrant records at Ellis Island and the International Red Cross provided no clues. Perhaps, traveling across wartime Europe, the little girls had never even made it to the United States.

Then one day, Sasha's cousin typed "Polotsky" into Google's search window and found a clue. An entry on a genealogical website mentioned "Minacker," the name of Sophia's and Ossie's father. In short order, Rosalie, Sophia, and Ossie were reunited in Florida, after 70 years apart. "All the time when he was alive, he asked me to do something to find them," said Sasha, recalling his grandfather's wish. "It's something magic."[19]

The World Wide Web has put vast amounts of information within reach of millions of ordinary people. But you can't reach for something if you don't know where it is. Most of that vast store of digital information might as well not exist without a way to find it. Indeed, the "dark web" exists as a kind of parallel universe, with troves of information invisible to search engines and to users who don't know where to look for it.

Search both fulfills dreams and shapes human knowledge. The search tools that help us find needles in the digital haystack are the lenses through which we view the digital landscape. But the "lenses" are not passive. They actively color what we see by their selection of what to show us on the first page of results and by the order in which the results are presented to us. Whoever controls the search engine shapes—and distorts—the reality we see through it. Google, which is used for more than 90% of the world's searches,[20] is supported by advertising, so questions inevitably arise about whether results optimize Google's profits or users' satisfaction. Microsoft's Bing is no less good at producing results, but it has less than 5% of the market. DuckDuckGo, which offers much stronger privacy protections than Google or Bing but produces less targeted results, has a negligible share of the market.[21] Baidu is the dominant search engine in China but for a reason: It censors heavily, as any search engine in the Chinese market must do. How did these lopsided statistics arise, and what are their consequences?

The Fall of Hierarchy

From the dawn of writing until about 1994, there were only two ways to organize information so it could be retrieved quickly. You could put it in a hierarchy, or you could create an index.

A hierarchy enables you to put things into categories and divide those categories into subcategories. Aristotle tried to classify everything. Living things, for example, were either plants or animals. Animals either had red blood or did not; red-blooded animals were either live-bearers or egg-bearers; live-bearers were either humans or other mammals; egg-bearers either swam or flew; and so on. Sponges, bats, and whales all presented classification enigmas, on which Aristotle did not think he had the last word. At the dawn of the Enlightenment, Linnaeus provided a more useful way of classifying living things, using an approach that gained intrinsic scientific validity once it reflected evolutionary lines of descent.

Our traditions of hierarchical classification are evident everywhere. We just love outline structures. The law against cracking copyright protection is Title 17, Section 1201, paragraph (a), part (1), subpart (A). In the Library of Congress system, every book is in one of 26 major categories, designated by an uppercase letter, and these major categories are internally divided in a similar way; for example, in category B, philosophy, you find BQ, Buddhism.

If the categories are clear, it may be possible to use an organizing hierarchy to locate what you are looking for. This requires that the person doing the searching not only know the classification system but be skilled at making all the necessary decisions. For example, if knowledge about living things was organized as Aristotle had it, anyone wanting to know about whales would have to know already whether a whale is a fish or a mammal in order to go down the proper branch of the classification tree. As more and more knowledge has to be stuffed into the tree, the tree grows and sprouts twigs, which over time become branches sprouting more twigs. The classification problem becomes unwieldy, and the retrieval problem becomes practically impossible.

In 1991, when the Internet was barely known outside academic and government circles, some academic researchers offered a program called Gopher. This program provided a hierarchical directory of many websites, by organizing the directories provided by the individual sites into one big outline. Finding things using Gopher was tedious by today's standards, and it was dependent on the organizational skills of the contributors. Yahoo! was founded in 1994 as an online Internet directory, with human editors placing products and services in categories, making recommendations, and generally trying to make the Internet accessible to non-techies. Although it is today a search and news site, the name "Yahoo" was originally said to be an acronym for "Yet Another Hierarchical Organized Oracle."

The practical limitations of hierarchical organization trees were foreseen 60 years ago, long before the explosive growth of the World Wide Web and the countless daily changes to it. During World War II, President Franklin Roosevelt appointed Vannevar Bush of MIT to serve as director of the Office of Strategic Research and Development (OSRD). The OSRD coordinated scientific research in support of the war effort. It was a large effort, with 30,000 people and hundreds of projects covering the spectrum of science and engineering. The Manhattan Project, which produced the atomic bomb, was just a small piece of it.

From his vantage point, Bush saw a major obstacle to continued scientific progress. We were producing information faster than it could be consumed—or even classified. Decades before computers became commonplace, he wrote about this problem in a visionary article, "As We May Think."[22] It appeared in the *Atlantic Monthly*—a popular magazine, not a technical journal. As Bush saw it,

> The difficulty seems to be, not so much that we publish unduly...but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships....Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing.

The dawn of the digital era was at this time barely a glimmer on the horizon. But Bush imagined a machine, which he called a "memex," that would augment human memory by storing and retrieving all the information needed. It would be an "enlarged intimate supplement" to human memory, which could be "consulted with exceeding speed and flexibility."

Bush clearly perceived the problem, but the technologies available at the time—microfilm and vacuum tubes—could not solve it. He understood that the problem of finding information would eventually overwhelm the progress of science in creating and recording knowledge, and he anticipated that it would be possible to search using multiple terms to isolate special kinds of information:

> Wholly new forms of encyclopedias will appear, ready made with a mesh of associative trails running through them, ready to be dropped into the memex and there amplified....
>
> The historian, with a vast chronological account of a people, parallels it with a skip trail which stops only on the salient items, and can follow at any time contemporary trails which lead him all over civilization at a particular epoch. There is a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record. The inheritance from the master becomes, not only his additions to the world's record, but for his disciples the entire scaffolding by which they were erected.

Bush was intensely aware that civilization itself had been imperiled in the war, but he thought we must proceed with optimism about what the record of our vast knowledge might bring us:

> Presumably man's spirit should be elevated if he can better review his shady past and analyze more completely and objectively his present problems. He has built a civilization so complex that he needs to mechanize his records more fully if he is to push his experiment to its logical conclusion and not merely become bogged down part way

> there by overtaxing his limited memory. His excursions may be more enjoyable if he can reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand, with some assurance that he can find them again if they prove important.
>
> ...He may perish in conflict before he learns to wield that record for his true good. Yet, in the application of science to the needs and desires of man, it would seem to be a singularly unfortunate stage at which to terminate the process, or to lose hope as to the outcome.

A Futurist Precedent

In 1937, H. G. Wells anticipated Vannevar Bush's 1945 vision of a memex. Wells wrote even more clearly about the possibility of indexing everything and what that would mean for civilization:

> There is no practical obstacle whatever now to the creation of an efficient index to all human knowledge, ideas and achievements, to the creation, that is, of a complete planetary memory for all mankind. And not simply an index; the direct reproduction of the thing itself can be summoned to any properly prepared spot.... This in itself is a fact of tremendous significance. It foreshadows a real intellectual unification of our race. The whole human memory can be, and probably in a short time will be, made accessible to every individual....This is no remote dream, no fantasy.[23]

Capabilities Bush could not have seen clearly are commonplace now. Digital computers, vast storage, and high-speed networks make information search and retrieval necessary. They also make it possible. The Web is a realization of Bush's memex, and search is key to making it useful.

Search Histories

Bush did not imagine that everyone would have a memex, but he did foresee that "associative trails" would endure. It's worth looking a little more closely at what that implies about the way search engines work. What Bush saw as an important new knowledge structure has turned out to be something more like digital exhaust—a mostly harmless side effect of the fancy digital engines we use to get things done.

"Search" is something of a misnomer for what Google and other search engines actually do. When you type something into a search engine, the engine does not go check the entire World Wide Web, looking for it. It looks up your search term in an index that has already been created. It is a very large index and very cleverly organized so that it can be updated constantly and so that searches can focus on multiple terms, but it is fundamentally no different from the index in the back of a book...except that when you look up something in the index of a book, nobody but you knows that you did so. If you ask Google to look something up in its index, it remembers that you did so.

There are good reasons for Google to remember what you searched for. The information might be useful in helping Google respond to future searches more appropriately. It would certainly be useful in helping Google target advertising to you. You can use a privacy-preserving search engine (such as DuckDuckGo, mentioned earlier), but you may not be as happy with the quality of the results. The overwhelming dominance of Google suggests that people are happy to trade their privacy for quality—or are just going with the household name and don't realize what exchange they are making.

Casey Anthony may not have been thinking about the endurance of search histories when she or someone using her computer Googled "neck breaking" and "how to make chloroform" before the mysterious death of her daughter Caylee in 2012.[24] The disclosure of this search history in her trial didn't result in her conviction; later it came out that the same computer had been used to search for "foolproof suffocation" on the very day the girl went missing.[25] (The detectives had missed this because the search was done using the Firefox browser rather than the Internet Explorer browser that gave up the information about the other searches.) A few years before, James Petrick had been convicted of killing his wife in part on the basis of searches he had done for terms such as "neck" and "snap break," and for topographic information about the lake where her body was found.[26] An appeals court in Illinois upheld the murder and solicitation of murder convictions of Steven Louis Zirko[27] in part on the basis of searches on Zirko's computer for terms like "mercenary for hire" and the hours when the child of one of his victims would be at school.

These cases all involved police searches of someone's home computer. But there is another way to get information about search history: by asking Google. Google won't just give that information out to anyone who asks, but you can see for yourself what Google is remembering about your searches and other activity that was conducted while you were logged into Google. (Or at least what Google tells you it is remembering; it probably is remembering a great deal more.) Under your Google account page, there is a "Data and personalization" screen, where you can turn off the recording of your search history, for example. You can even edit the history without getting rid of it entirely, if you want to—as Dr. Brent Dennis apparently did in order to mislead law enforcement. He told the police that his wife had died from drinking antifreeze, but it was he himself who searched for "antifreeze" and then hired someone to clean up the search history.[28]

Whenever you go to a new website that asks you to create an account or to "sign in using Google" or "sign in using Facebook," you might be pleased to save time and have one fewer password to remember. But what you are really doing if you choose to use your existing Google or Facebook login credentials is giving Google or Facebook permission to add to the enormous store of information it already holds about you the new information it gleans from your activity on the new site.

And the government can force Google to turn over what it knows about you—your search history, for example. It's not even that complicated. In 2018 Google received about 130,000 requests from courts and other government agencies, and complied, at least in part, with about two-thirds of them.[29] Google says it will inform you when an agency seeks your records, but it is under no obligation to do so or to comply with your wishes if you object.

To get that information from your laptop, law enforcement would need a search warrant; that is, it would

Can the Police Search Your Searches?

Can law enforcement get information about your searches even if you have done nothing wrong? It seems to have happened in Edina, Minnesota, in early 2017.[30] Someone impersonating a man named Douglas called Spire Credit Union and persuaded the clerk to transfer $28,500 of Douglas's money to another bank. The fake Douglas duly supplied Douglas's name, birth date, Social Security number, and a faxed copy of his passport—or at least a passport that had Douglas's photo on it, which matched the bank's records and completed its authentication.

When Douglas realized that his money was gone, he contacted law enforcement. Detective David Lindman used Google's image search to find a matching photo of Douglas online. Lindman asked the Hennepin County judge to issue a search warrant to Google for records of anyone in Edina who had searched for Douglas's image during a five-week period prior to the incident. So at least in some cases, a search warrant can now be issued not against a particular individual but for the set of individuals who have performed a certain kind of search.

Google the gatekeeper turns out to have gates swinging both ways. As you search, it is determining what information to show you, and it is also collecting information about you at the same time. It can use that information for its own advertising purposes, and under court order, it can open the gate to outsiders.

have to make a case to a judge that your Fourth Amendment rights against unreasonable searches were not being violated. Why can the police more easily get the same information directly from Google?

The underlying legal principle is simple: In the absence of any more specific legislation, if law enforcement asks Google what you searched for, it may spill the beans on you because of the "Third Party Doctrine." You can tell me a secret, and the government can't make either of us disclose it. But if you use Gmail—essentially asking Google to pass your secret to me—then Google is a third party and is not bound under the Fourth Amendment to respect your desire and mine to keep the information secret, any more than if you shared your secret with a stranger on the street. The same applies to other information you have entrusted to Google—for example, the terms you've asked it to search for. Those searches are Google's property, not yours. It can use them to generate targeted advertising and for other purposes.

In 2017 Google announced that it would not scan users' email in order to improve its advertisement targeting, but that was a policy decision, not a response to any U.S. law. In fact, the United States lacks comprehensive privacy laws, and corporate policies need not be unchangeable or consistent with users' expectations. When Gmail was launched in 2004, Google explained its practice of scanning users' email as being helpful in targeting advertisements to offset the cost of offering a free service. A decade of mounting criticism and some litigation resulted in Google's decision to stop scanning email. What Google did not explain at the time was that it was enabling certain corporate partners to scan emails—and sometimes have humans read them.

Navideh Forghani of Phoenix, Arizona, seems to have been unaware of such practices when she signed up for Earny, a money-saving service.[31] Earny checks the customer's inbox for receipts of items she has purchased, searches the Web to see if it can locate the same items at cheaper prices, files with her credit card company for a refund of the price difference, and then splits the proceeds with the customer, all quietly in the background. Once Navideh had signed up, the only thing she had to do was to watch the credits appear on her credit card bills.

Earny is a private company separate from Google, and Google is not scanning her email. But when she clicked a button to give Earny access to her inbox, she was authorizing Earny to do exactly that. Earny, in turn, was sharing her email with Return Path, a company with which Earny had partnered to do the actual scanning.

Google, Earny, and Return Path all explained that they had done nothing wrong because these practices were authorized under the companies' privacy policies. Navideh acknowledges that she did not read Earny's privacy policy[32]. No surprise: It was almost 3,000 words long when I checked it in 2019. Earny's policy is sufficiently dense as to be difficult for many readers; and, to make matters worse, it links to a host of other privacy policies making it nearly impossible to understand what you are relinquishing when you sign up for their service. Return Path's privacy policy is almost 6,000 words.

The bottom line is that your data is valuable, and every convenience you accept comes at a price. As Marc Rotenberg, president of EPIC, a major privacy advocacy organization, stated, "The privacy policy model is simply broken beyond repair. There is simply no way that Gmail users could imagine that their personal data would be transferred to third parties."[33] When a product has so little competition and is so useful to everyday life and to the conduct of business, the "notice and consent" protocol does not realistically protect users from having their data used in unexpected ways.

How Did Google Get So Big?

As the Web grew in the early 1990s, hierarchical structures, never satisfactory for finding unclassifiable information, quickly failed to keep up with the size of the Web. Several search engines based on an automatically constructed index began to appear, and some were modestly successful. But very shortly Google became dominant, to the extent that the name of the search engine, and the company, has become a verb synonymous with "web search."

In 1996, the Google founders, Larry Brin and Sergey Brin, had a good idea while they were in graduate school. An important web page is one referenced—that is, linked to—by a lot of important pages. That sounds like a circular definition, but if the entire structure of the Web can be captured and analyzed, some fairly simple mathematics can be used to get a consistent measure of the importance of every web page. That mathematics, plus some solid engineering to get all the data organized and processed in the limited storage available at the time, got the company off the ground. Its dramatically simple interface—just type in something and get answers back, no options, bells, or whistles—comforted even the most naïve users and lured them into using it more.

Google's search engine was good, but it was not ten times better than others available when the company was founded in 1998. For example, by this time AltaVista had been operational for three years[34] and was processing hundreds of millions of search queries as a free service to the public.

Digital Equipment Corporation, which developed AltaVista, never figured out how to make it profitable. Digital was primarily a hardware company, and it sold AltaVista to another company. (Digital itself was bought by Compaq soon after.) AltaVista changed hands again and was finally quietly shut down. Microsoft didn't launch its Bing search engine until 2009. By then Google had a seemingly unsurmountable head start, in spite of the fact that users can switch search engines from Google to Bing with minimal effort.

Google gained its advantage by carrying advertising from the very beginning. The ads are generated in response to search terms; search for "cell phones," and you are likely to see ads for products and services related to cell phones. Which ads come up, among all the advertisers who might want to tout their goods to people interested in cell phones, is determined by a continuous auction. Advertisers willing to pay more for their ads are more likely to have their ads appear. The auctions run automatically and invisibly, and the result is a system of unprecedented efficiency. An advertiser in a newspaper, magazine, or radio station has to hope that among the undifferentiated mass of people exposed to the ad, a few will be interested in the product advertised. Advertisers can try to tilt the odds in their favor by, for example, putting sports-related advertisements on radio stations that cater to sports fans. But associating advertising with search directs the ads only to those individuals who have shown at least enough interest in a topic to search for it.

The Google founders themselves recognized[35] the downside of mixing advertising with search, which was already being done with a few of the other search engines then in use. It would lower confidence in search results, for example, if users suspected that the search results themselves were biased to favor the advertisers:

> For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention," a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving. This search result came up first because of its high importance as judged by the PageRank algorithm, an approximation of citation importance on the web. It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers.

After mentioning a few other examples of conflicts of interest between returning useful search results and gaining advertising revenue, Page and Brin concluded, "we believe the issue of advertising causes enough mixed incentives that it is crucial to have a competitive search engine that is transparent and in the academic realm." Be that as it may, no such search engine is widely used today. Google's enormous revenues are largely derived from exactly this kind of advertising, and in 2017 the European Union fined Google 2.4 billion euros for biasing its search results in favor of its advertisers.[36] And without knowing exactly what is going on inside Google's code, it is hard to know whether results are being biased. Brin and Page anticipated this, too: "For example, a search engine could add a small factor to search results from 'friendly' companies and subtract a factor from results from competitors. This type of bias is very difficult to detect but could still have a significant effect on the market." How the conflicting interests and lack of transparency will be resolved remains unknown, but the stakes are extremely high.

Social Gatekeepers: Known by the Company You Keep

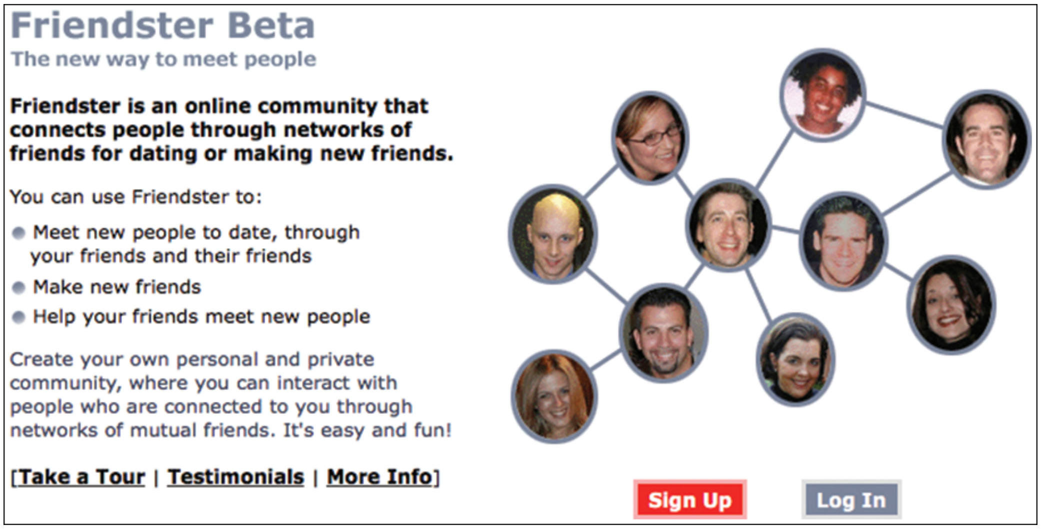
When it was created, the Internet was a means of connecting one computer to another and, ultimately, one network of computers to another (hence the name "Internet"). It expanded from connecting machines to connecting users to information. The complexity of these connections led to the role of the search gatekeeper. In its latest phase of connectivity, the Internet has facilitated the connection of people to each other at a level, and with implications, that were not imagined even by the creators of the dominant solutions.

The Social Network: Facebook and More

The digital explosion has never been so powerful as in the growth of Facebook. As glamorized in the movie *The Social Network*, Facebook's success would seem to be the result of adolescent dumb luck and capitalist ruthlessness. The full story is more interesting—and more telling in terms of the ways people share bits.

There were online social networks before Facebook. The earliest was Sixdegrees.com in 1997. The name was derived from the early 1990s play and movie *Six Degrees of Separation*. The site grew to include millions of users, but it eventually stalled and died in 2000 for lack of a sustainable business model and because there wasn't much for people to do on it once they had connected to each other.[37]

Friendster launched in 2002 and quickly grew to be among the most popular sites on the Web. Originally it billed itself as a place where users could meet new people to date, make new friends, or help friends meet new people.[38] At its peak, it allowed easy search through the entire membership database of tens of millions of people.



Friendster Beta

The new way to meet people

Friendster is an online community that connects people through networks of friends for dating or making new friends.

You can use Friendster to:

- Meet new people to date, through your friends and their friends
- Make new friends
- Help your friends meet new people

Create your own personal and private community, where you can interact with people who are connected to you through networks of mutual friends. It's easy and fun!

[Take a Tour | Testimonials | More Info]

Friendster collapsed in 2006, a victim of its own success. Growth was so explosive that the site was plagued with technical problems. People have very limited tolerance for delay while waiting for pages to load; they will abandon a site completely if it doesn't work and they have no strong reason to keep trying. Friendster also made too much of its main merit, which was making it easy to connect to people you didn't already know but might be interested in meeting. Moreover, unanticipated social problems arose because it was so easy to view the profiles of other users. For example, it turned out that not everything users put in their profiles to stimulate social connections was something they wanted their bosses to know.

Myspace was started in 2003 as a competitor to Friendster. It drew Friendster users who were either dissatisfied or had been thrown off the site. Myspace had a loyal following among indie rock bands and their followers who had refused to play by Friendster's rules. Soon Myspace had more visitors than Google or any other website, though culturally it retained the naughty and creative feel of its rebellious origins. But within a few years, users were abandoning it in droves after a number of highly publicized meetups between adults and children who found each other online. The moral panic caused the U.S. government to consider legislation to control online social networking (see box).[39]

Do You Know Where Your Child Is on the Web Tonight?

It was every parent's worst nightmare. Katherine Lester, a 16-year-old honors student from Fairgrove, Michigan, went missing in June 2006. Her parents had no idea what had happened to her; she had never given them a moment's worry. They called the police. Then federal authorities got involved.

After three days of terrifying absence, she was found, safe—in Amman, Jordan.

Fairgrove is too small to have a post office, and the Lesters lived in the last house on a dead-end street. In another time, Katherine's school, 6 miles away, might have been the outer limit of her universe. But through the Internet, her universe was the whole world. Katherine met a Palestinian man, Abdullah Jinzawi, from Jericho on the West Bank. She found his profile on the social networking website Myspace and sent him a message: "u r cute." They quickly learned everything about each other through online messages. Lester tricked her mother into getting her a passport and then took off for the Middle East. When U.S. authorities met her plane in Amman, she agreed to return home and apologized to her parents for the distress she had caused them.

A month later, Representative Judy Biggert of Illinois rose in the House to co-sponsor the Deleting Online Predators Act (DOPA). "MySpace.com and other networking websites have become new hunting grounds for child predators," she said, noting that "we were all horrified" by the story of Katherine Lester. "At least let's give parents some comfort that their children won't fall prey while using the Internet at schools and libraries that receive federal funding for Internet services." The law would require those institutions to prevent children from using on-location computers to access chat rooms and social networking websites without adult supervision.

Speaker after speaker rose in the House to stress the importance of protecting children from online predators, but not all supported the bill. The language was "overbroad and ambiguous," said one. As originally drafted, it seemed to cover not just Myspace but sites such as Amazon and Wikipedia. These sites possess some of the same characteristics as Myspace: Users can create personal profiles and continually share information with each other by using the Web. Although the law might block children in schools and libraries from "places" where they meet friends (and sometimes predators), it would also prevent access to online encyclopedias and bookstores, which rely on content posted by users.

Instead of taking the time to develop a sharper definition of what exactly was to be prohibited, DOPA's sponsors hastily redrafted the law to omit the definition, leaving it to the Federal Communications Commission to decide later just what the law would cover. Some murmured that the upcoming midterm elections were motivating the sponsors to put forward an ill-considered and showy effort to protect children—an effort that would likely be ineffective and so vague as to be unconstitutional.

Children use computers in lots of places; restricting what happens in schools and libraries would hardly discourage determined teenagers from sneaking onto Myspace. Only the most overbearing parents could honestly answer the question *USA Today* asked in its article about "cyber-predators": "It's 11 p.m. Do you know where your child is on the Web tonight?"

The statistics about what can go wrong were surely terrifying. The Justice Department has made thousands of arrests for "cyber enticement"—almost always older men using social networking websites to lure teenagers into meetings, some of which end very badly. Yet, as the American Library Association stated in opposition to DOPA, education, not prohibition, is the "key to safe use of the Internet." Students have to learn to cooperate online because network use and all the human interactions it enables are basic tools of the new, globally interconnected world of business, education, and citizenship—and perhaps even the globally interconnected world of true love.

The tale of Katherine Lester took an unexpected turn. From the moment she was found in Jordan, Lester steadily insisted that she intended to marry Jinzawi. Jinzawi, who was 20 when he and Lester first made contact, claimed to be in love with her—and his mother loved her, too. Jinzawi begged Lester to tell her parents the truth before she headed off to meet him, but she refused. Upon her return, authorities charged Lester as a runaway child and took her passport away from her. But on September 12, 2007, having attained legal independence by turning 18, she again boarded a plane to the Middle East, finally to meet her beloved face to face. The affair finally ended a few weeks later in an exchange of accusations and denials, as well as a hint that a third party had attracted Lester's attentions. There was no high-tech drama to the breakup—except that it was televised on *Dr. Phil*.

This was the environment in which Facebook was launched in a Harvard dorm room in 2004. Zuckerberg had been studying sociology and psychology and computer networking, and he hacked together a simple website that he proposed to call "Six Degrees to Harry Lewis." As he wrote to Lewis,

> Professor, I've been interested in graph theory and its applications to social networks for a while now, so I did some research...that has to do with linking people through articles they appear in from the Crimson. I thought people would find this interesting, so I've set up a preliminary site that allows people to find the connection (through people and articles) from any person to the most frequently mentioned person in the time frame I looked at. This person is you. I wanted to ask your permission to put this site up though, since it has your name in its title.

Lewis briefly demurred. "Can I see it before I say yes? It's all public information, but there is somehow a point at which aggregation of public information feels like an invasion of privacy." Shortly thereafter, thinking that the project sounded educational, Lewis replied, "Sure, what the hell. Seems harmless." thefacebook, as it was originally known, launched a week later.

Within two years, Facebook had overtaken Myspace. It has succeeded through a combination of good decisions:

- Good engineering (it was generally reliable even in its most explosive growth phase)
- Design that balanced, more successfully than its competitors, the opposing imperatives to connect the world and to provide spaces for more intimate conversation between birds of a feather

- An interface that was calmer and more standardized than that of Myspace, perhaps reflecting that Facebook had originated as a community of college students rather than indie rock music lovers
- An advertising model that was largely palatable to its users (in part because they were unaware of the extent to which their data was repurposed)
- A large number of strategic acquisitions that had the combined effect of making Facebook a single-stop platform not just for social networking but for text messaging, video search, photo storing and viewing, shopping, and gaming, among other things

The result was growth at an astonishing rate.[40] Facebook was launched on February 4, 2004, as a site just for Harvard students, a replacement for the "face books" printed by the Harvard student residences to familiarize students with each other. A month later the network was extended to Stanford, Yale, and Columbia, and by the end of the year it had more than 1 million users. The site may have become more popular because it began with an aura of exclusivity. During 2005 it added hundreds more colleges as well as high schools, and by the end of 2006 anyone could join, and the user base was over 12 million. A year later it was up to 60 million, and it hit 500 million by mid-2010. At the time of this writing, Facebook puts the number of users at 1.59 billion daily users, with 2.41 billion who use it at least once a month. That's about a third of the population of the earth, including infants. The number is still rising, in spite of adverse publicity due to misuse of the company's data.

Indeed, Facebook has a lot of data about its users, and it has done a lot of "learning on the job" about how to handle it. It was aware, early on, that privacy would be important to users and stated unequivocally in 2007, "We do not and will not use cookies to collect private information from any user."[41] But only a few months later, Facebook user Sean Lane bought a diamond eternity ring online—and his wife learned about it instantly, from Facebook. Facebook had recently launched a new feature called Beacon. In an attempt to keep Facebook friends up to date about what users were doing—and also to expand the opportunities for advertising on Facebook—Beacon posted information about what users were buying from non-Facebook sites on friends' news feeds. Lane's wife not only learned about the ring he had purchased but that he had gotten a 51% discount on it. "Who is the ring for?" she asked.[42] It was for her, so only the surprise was spoiled, and not the marriage!

Facebook had partnered with other sites in an information-sharing scheme. When users made a purchase on a partner site, Facebook was informed, and it would sometimes insert the information in friends' news feeds. Users could opt out—if they noticed the tiny box on the partner site and understood what they were being invited to opt out of. Tens of thousands of users were furious and petitioned the company to remove the feature. Matters got worse when a researcher discovered that under certain circumstances, the information was sent to Facebook even when the user was logged out and no opt-out box was shown. Denials by company spokespeople proved to be inaccurate. Lawsuits ensued. Zuckerberg first apologized,[43] had to pay millions of dollars to settle, and then shut down the Beacon feature entirely.

But the financial penalty didn't prevent further privacy flubs. In late 2009, when the network had grown to 350 million users, its privacy policy was updated without prior notice. Wrapped in an announcement[44] touting that users would now be expected to "personalize their privacy," the company noted that the default privacy settings had changed. It had been the case that only a user's name and "network" were visible to the outside world. ("Networks" were a vestige of the days when members of only certain groups could join Facebook—a user's network was his or her college or high school, for example.) According to the new policy,[45]

> Certain categories of information such as your name, profile photo, list of friends and pages you are a fan of, gender, geographic region, and networks you belong to are considered publicly available to everyone, including Facebook-enhanced applications, and therefore do not have privacy settings. You can, however, limit the ability of others to find this information through search using your search privacy settings.

The announcement noted that most people make this information public anyway. Perhaps so, but there was a big difference between what some people chose to do and what others were expecting. In short order it became apparent how revealing friends lists and fan pages might be. MIT researchers, for example, found that it was not difficult to figure out, with high accuracy, who was gay, even in the absence of any explicit information about sexual orientation:

> Public information about one's coworkers, friends, family, and acquaintances, as well as one's associations with them, implicitly reveals private information....Our research demonstrates a method for accurately predicting the sexual orientation of Facebook users by analyzing friendship associations. After analyzing 4,080 Facebook profiles from the MIT network, we determined that the percentage of a given user's friends who self-identify as gay male is strongly correlated with the sexual orientation of that user, and we developed a logistic regression classifier with strong predictive power.[46]

The aggregation of enough public information does indeed constitute an invasion of privacy—something Zuckerberg seems not to have thought through. After the Facebook defaults changed, he discovered that his own photographs had become public, and he quickly reset his privacy settings.[47] Another Facebook executive suggested that users who did not want their hometown to be made public should lie about it—apparently forgetting that Facebook requires such information to be truthful.[48]

The reaction was vocal, not just from users but from government officials and privacy organizations. On April 27, 2010, Zuckerberg received a polite but ominous letter from four U.S. senators that concluded,

> We look forward to the FTC examining this issue, but in the meantime we believe Facebook can take swift and productive steps to alleviate the concerns of its users. Providing opt-in mechanisms for information sharing instead of expecting users to go through long and complicated opt-out processes is a critical step towards maintaining clarity and transparency.[49]

In May of 2010, Facebook reversed the defaults, so that only name, photo, gender, and networks would automatically be public.[50]

In spite of the outcry from users and the hint of forthcoming involvement by a federal agency, in the few months between the ill-considered change of privacy defaults and the decision to revert to the previous assumptions, Facebook had added 50 million users. People were complaining but were finding Facebook too useful to give up. Every new member of the network made it that much more valuable to join. People went to Facebook not to make new friends but because all their friends were already there. This phenomenon is known as a *network* effect: As Paul Baran had anticipated (see page 92), the value to an individual of being in the network increases as the size of the network as a whole increases. After 2010, Facebook had no serious social network competition in the United States, although in some parts of the world, Facebook was not even allowed to sign up users.

As most of the United States and a large number of non-U.S. individuals joined, the network effect was boosted by product diversification. Facebook added a text messaging service in 2008 and acquired Instagram in 2012 and WhatsApp in 2014. Facebook had become a full-service platform for all kinds of communication—good, bad, and fraudulent. In 2019 it was reported that Facebook was used for 90% of reported instances of sharing of child pornography.[51] The bulk of that communication was via Facebook's Messenger app, and when Facebook announced that it would be adding end-to-end encryption to Messenger, which would make it impossible for anyone but the recipient to decipher Messenger communications en route, the U.S. attorney general and his counterparts in other countries firmly asked Zuckerberg not to follow through.[52] The letter cited one of thousands of examples of how law enforcement had used electronic surveillance to catch a criminal:

> To take one example, Facebook sent a priority report to NCMEC, having identified a child who had sent self-produced child sexual abuse material to an adult male. Facebook located multiple chats between the two that indicated historical and ongoing sexual abuse. When investigators were able to locate and interview the child, she reported that the adult had sexually abused her hundreds of times over the course of four years, starting when she was 11. He also regularly demanded that she send him sexually explicit imagery of herself. The offender, who had held a position of trust with the child, was sentenced to 18 years in prison. Without the information from Facebook, abuse of this girl might be continuing to this day.

Encryption would make it impossible to catch such offenders. "We therefore call on Facebook and other companies," the letter continued, "whatever form of encryption they use, to enable law enforcement to obtain lawful access to content in a readable and usable format."

Chapter 5, "Secret Bits," traces the history of encryption. Most privacy and security experts agree that allowing law enforcement access to encrypted communications greatly increases the risk that others will be able to gain access to them as well. But an even larger issue is at stake: Facebook and Google have more information about most human beings and every form of human activity than any government. The mistakes they make—data leaks and service interruptions, for example—can affect significant portions of the population of the earth. In the case of Facebook, Mark Zuckerberg alone can, in principle, make the decisions by himself.

As a vehicle for many forms of communication, Facebook is under no obligation to allow every lawful form of speech—and given the protections that Internet companies enjoy under Section 230 (see Chapter 7, "You Can't Say That on the Internet") of the Communications Decency Act, their responsibility even for unlawful forms of speech is limited in the United States. They recognize, however, both a business interest and a political interest in limiting some forms of speech even in the United States, and abroad their obligation to censor is much more explicit. Fraudulent political ads may have affected voters in the 2016 U.S. presidential election. Several mass shooters were steeped in violent social media content.

Yet these companies can largely set policies—what and whether to censor and whether and how to encrypt messages, for example—as they see fit. They will doubtless tend to make those decisions with their shareholders' interests in mind; it would actually be improper for them to do otherwise routinely. Of course, those interests are best served by following policies of which the public will generally approve. But does the public itself have an interest in intervening, as Attorney General Barr suggests, and as Senator Schumer had suggested earlier?

And having a policy is not the same as executing it perfectly. Human review of every comment or advertisement or video posted to Facebook is impossible; even responding in a timely manner to those that users complain about would be incredibly difficult. Inevitably the gatekeepers have turned to artificial intelligence (AI) software to do some of the screening for them. But where intent and context are important, AI is not yet a match for human readers.

What if a word or phrase in a post trips Facebook's "hate speech" prohibition, but it is taken out of context? An algorithm cannot make that judgement. Reluctant to be accused of political bias, Facebook has announced that it won't generally remove political advertising, even when it is known to be blatantly false. Its decision, it says, "is grounded in Facebook's fundamental belief in free expression, respect for the democratic process, and the belief that, in mature democracies with a free press, political speech is arguably the most scrutinized speech there is."[53] Is this what will best serve the public interest? It is not surprising that not everyone agreed. What to do instead is far less clear.

Or are these companies just too big? That notion became popular among some of the presidential candidates during the 2020 election cycle. Elizabeth Warren thinks Facebook should be broken up—perhaps by unwinding some of its acquisitions. Whether that would be either legal or helpful will excite intense discussion. Another idea would be to leave them intact but to regulate them more tightly—though the devil would be in the details. It would not be a small step on the part of the government to treat a private company whose product is bits as though it were a public utility whose product is water.

One part of the government has particular concerns about a specific aspect of the services offered by technology companies. As a result of one of the most remarkable discoveries of the twentieth century—just a little bit of arithmetic on bits—private citizens can and do now exchange over the public Internet encrypted messages that law enforcement can intercept but cannot decode. How this happened and what it portends is the subject of the next chapter.

Endnotes

1 Ian Austen, "A Canadian Telecom's Labor Dispute Leads to Blocked Web Sites and Questions of Censorship," *The New York Times*, August 1, 2005, https://www.nytimes.com/2005/08/01/business/worldbusiness/a-canadian-telecoms-labor-dispute-leads-to-blocked.html.

2 Tripp Mickle et al., "Apple, Google Pull Hong Kong Protest Apps Amid China Uproar," *Wall Street Journal*, October 10, 2019, https://www.wsj.com/articles/apple-pulls-hong-kong-cop-tracking-map-app-after-china-uproar-11570681464.

3 Jim Rutenberg, “Netflix’s Bow to Saudi Censors Comes at a Cost to Free Speech,” *The New York Times*, January 6, 2019, https://www.nytimes.com/2019/01/06/business/media/netflix-saudi-arabia-censorship-hasan-minhaj.html.

4 Steve Kroft, “How Did Google Get so Big?” CBS News, May 21, 2018, https://www.cbsnews.com/news/how-did-google-get-so-big/.

5 “Internet Providers in Browning, Montana,” Broadband Now, accessed April 27, 2020, https://broadbandnow.com/Montana/Browning.

6 “Browning, MT,” Data USA, accessed April 27, 2020, https://datausa.io/profile/geo/browning-mt/.

7 Paul Baran, “On Distributed Communications Networks,” (RAND Corporation, Santa Monica, CA, September 1962), Reprinted with permission. https://www.rand.org/pubs/papers/P2626.html.

8 Vinton G. Cerf and Robert E Kahn, “A Protocol for Packet Network Intercommunication,” *IEEE Transactions on Communications*, no. 5 (1974): 13.

9 Pete Resnick, “On Consensus and Humming in the IETF,” Internet Engineering Task Force, June 2014, https://tools.ietf.org/html/rfc7282.

10 Harald Tveit Alvestrand, “A Mission Statement for the IETF,” Internet Engineering Task Force, October 2004, https://tools.ietf.org/html/rfc3935.

11 The Internet protocol packet design was set before the days of massive miniaturization and at a time when computer memory was limited and costly. Nobody then imagined the need to connect more than 4 billion computers to the Internet, so only 32 bits were reserved for the address fields. But now wristwatches and refrigerators have their own IP addresses, and the total number of connected computers is more than can be distinguished using 32 bits. Various workarounds have been developed, and a new protocol, IPv6, which has 128-bit addresses, is slowly being rolled out.

12 Jonathan Zittrain, *The Future of the Internet—and How to Stop It* (Yale University Press, 2008).

13 For a full account of the story of fiber, see Susan Crawford, Fiber: *The Coming Tech Revolution—and Why America Might Miss It*, Yale University Press, 2018.

14 Jon Brodkin, “US Broadband: Still No ISP Choice for Many, Especially at Higher Speeds,” *Ars Technica*, August 10, 2016, https://arstechnica.com/information-technology/2016/08/us-broadband-still-no-isp-choice-for-many-especially-at-higher-speeds/.

15 Kendra Chamberlain, “Municipal Broadband Is Roadblocked or Outlawed in 25 States,” *Broadband Now*, May 13, 2020, https://broadbandnow.com/report/municipal-broadband-roadblocks/.

16 “Government Competition with Private Internet Services Providers Prohibited—Exceptions,” Montana Code Annotated 2019, https://leg.mt.gov/bills/mca/title_0020/chapter_0170/part_0060/section_0030/0020-0170-0060-0030.html.

17 Paul Baran, “Full Text of ‘The Computer and Invasion of Privacy,’” July 26, 1966, https://archive.org/stream/U.S.House1966TheComputerAndInvasionOfPrivacy/U.S.%20House%20%281966%29%20-%20The%20Computer%20and%20Invasion%20of%20Privacy_djvu.txt.

18 Scott Bradner, “The Internet: Unblocking Pipes,” Network World, March 14, 2005, https://www.networkworld.com/article/2319666/the-internet--unblocking-pipes.html.

19 Eva Wolchover, “Web Reconnects Cousins Cut off by Iron Curtain,” *Boston Herald*, December 18, 2007.

20 “Search Engine Market Share Worldwide 2019,” Statista, accessed April 27, 2020, https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/.

21 Nathaniel Popper, “A Feisty Google Adversary Tests How Much People Care About Privacy,” *The New York Times*, July 15, 2019, https://www.nytimes.com/2019/07/15/technology/duckduckgo-private-search. html.

22 Vannevar Bush, “As We May Think,” *The Atlantic*, July 1, 1945, https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

23 H. G. Wells, *World Brain* (Methuen, 1938), pp. 60–61.

24 “Shady Web Searches in Missing Girl Case,” CBS News, November 26, 2008, https://www.cbsnews.com/news/shady-web-searches-in-missing-girl-case/.

25 Tony Pipitone, “Cops, Prosecutors Botched Casey Anthony Evidence,” WKMG, November 28, 2012, https://www.clickorlando.com/news/2012/11/28/cops-prosecutors-botched-casey-anthony-evidence/.

26 K. C. Jones, “Ex-Computer Consultant Convicted in ‘Google Murder’ Trial,” *InformationWeek*, November 30, 2005, https://www.informationweek.com/ex-computer-consultant-convicted-ingoog/174403074.

27 *People v. Zirko*, 2012 IL App (1st) 092158.

28 George Knapp and Matt Adams, “I-Team: Details the Night Attorney Susan Winters Died,” 8NewsNow, February 10, 2017, https://www.8newsnow.com/news/i-team-details-the-night-attorney-susan-winters-died/.

29 “Requests for User Information,” Google Transparency Report, accessed April 27, 2020, https://transparencyreport.google.com/user-data/overview.

30 “US Judge Asks Reports of Google Searches,” SEL, accessed April 27, 2020, https://searchenginelaw.net/security/103-us-judge-asks-reports-of-google-searches.

31 Douglas MacMillan, “Tech’s ‘Dirty Secret’: The App Developers Sifting Through Your Gmail,” *Wall Street Journal*, July 2, 2018, https://www.wsj.com/articles/techs-dirty-secret-the-app-developers-sifting-through-your-gmail-1530544442.

32 https://support.earny.co/hc/en-us/articles/218609757-Privacy-Policy#:~:text=We%2C%20at%20Earny%20Inc.%2C,save%20money%20in%20multiple%20ways.

33 John D. McKinnon and Douglas MacMillan, “Google Says It Continues to Allow Apps to Scan Data from Gmail Accounts,” *Wall Street Journal*, September 20, 2018, https://www.wsj.com/articles/google-says-it-continues-to-allow-apps-to-scan-data-from-gmail-accounts-1537459989.

34 Peter H. Lewis, “Digital Equipment Offers Web Browsers Its ‘Super Spider,’” *The New York Times*, December 18, 1995, https://www.nytimes.com/1995/12/18/business/digital-equipment-offers-web-browsers-its-super-spider.html.

35 Sergey Brin and Lawrence Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Networks and ISDN Systems* 30, no. 1–7 (April 1998): 107–117, https://snap.stanford.edu/class/cs224w-readings/Brin98Anatomy.pdf.

36 Mark Scott, "Google Fined Record $2.7 Billion in E.U. Antitrust Ruling," *The New York Times*, June 27, 2017, https://www.nytimes.com/2017/06/27/technology/eu-google-fine.html.

37 Danah M Boyd and Nicole B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication* 13, no. 1 (October 1, 2007): 210–30, https://doi.org/10.1111/j.1083-6101.2007.00393.x.

38 "Friendster," June 11, 2004, https://web.archive.org/web/20040611192459/http://www.friendster.com/index.jsp.

39 Pete Cashmore, "MySpace, America's Number One," Mashable, July 11, 2006, https://mashable.com/2006/07/11/myspace-americas-number-one/.

40 "Company Info," *About Facebook*, accessed April 28, 2020, https://about.fb.com/company-info/.

41 thefacebook, "Privacy Policy," January 7, 2005, https://web.archive.org/web/20050107221705/http:/www.thefacebook.com/policy.php.

42 Bill Goodwin and Sebastian Klovig Skelton, "Facebook's Privacy Game—How Zuckerberg Backtracked on Promises to Protect Personal Data," ComputerWeekly.com, July 1, 2019, https://www.computerweekly.com/feature/Facebooks-privacy-U-turn-how-Zuckerberg-backtracked-on-promises-to-protect-personal-data.

43 Facebook, "Thoughts on Beacon," December 5, 2007, https://www.facebook.com/notes/facebook/thoughts-on-beacon/7584397130/.

44 "Facebook Asks More Than 350 Million Users Around the World to Personalize Their Privacy," *About Facebook*, December 10, 2009, https://about.fb.com/news/2009/12/facebook-asks-more-than-350-million-users-around-the-world-to-personalize-their-privacy/.

45 Kurt Opsahl, "Facebook's Eroding Privacy Policy: A Timeline," Electronic Frontier Foundation, April 28, 2010, https://www.eff.org/deeplinks/2010/04/facebook-timeline.

46 Carter Jernigan and Behram F. T. Mistree, "Gaydar: Facebook Friendships Expose Sexual Orientation," *First Monday*, September 25, 2009, https://doi.org/10.5210/fm.v14i10.2611.

47 Kashmir Hill, "Either Mark Zuckerberg Got a Whole Lot Less Private or Facebook's CEO Doesn't Understand the Company's New Privacy Settings," *Forbes*, December 10, 2009, https://www.forbes.com/sites/kashmirhill/2009/12/10/either-mark-zuckerberg-got-a-whole-lot-less-private-or-facebooks-ceo-doesnt-understand-the-companys-new-privacy-settings/.

48 Julia Angwin, "How Facebook Is Making Friending Obsolete," *The Wall Street Journal*, December 15, 2009, https://www.wsj.com/articles/SB126084637203791583.

49 Politico Staff, "Senators' Letter to Facebook," Politico, April 27, 2010, https://www.politico.com/news/stories/0410/36406.html.

50 “Facebook Redesigns Privacy,” *About Facebook*, May 26, 2010, https://about.fb.com/news/2010/05/facebook-redesigns-privacy/.

51 Jennifer Valentino-DeVries and Gabriel J. X. Dance, “Facebook Encryption Eyed in Fight Against Online Child Sex Abuse,” *The New York Times*, October 2, 2019, https://www.nytimes.com/2019/10/02/technology/encryption-online-child-sex-abuse.html.

52 Priti Patel et al., “Open Letter to Facebook,” October 4, 2019, https://www.justice.gov/opa/press-release/file/1207081/download.

53 Cecilia Kang, “Facebook’s Hands-off Approach to Political Speech Gets Impeachment Test,” *The New York Times*, October 8, 2019, https://www.nytimes.com/2019/10/08/technology/facebook-trump-biden-ad.html.