The Wealth of Networks

How Social Production Transforms Markets and Freedom

Yochai Benkler

Yale University Press

New Haven and London

For Deb, Noam & Ari

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Contents

Acknowledgments

Reading this manuscript was an act of heroic generosity. I owe my gratitude to those who did and who therefore helped me to avoid at least some of the errors that I would have made without their assistance. Bruce Ackerman spent countless hours listening, and reading and challenging both this book and its precursor bits and pieces since 2001. I owe much of its present conception and form to his friendship. Jack Balkin not only read the manuscript, but in an act of great generosity taught it to his seminar, imposed it on the fellows of Yale's Information Society Project, and then spent hours with me working through the limitations and pitfalls they found. Marvin Ammori, Ady Barkan, Elazar Barkan, Becky Bolin, Eszter Hargittai, Niva Elkin Koren, Amy Kapczynski, Eddan Katz, Zac Katz, Nimrod Koslovski, Orly Lobel, Katherine McDaniel, and Siva Vaidhyanathan all read the manuscript and provided valuable thoughts and insights. Michael O'Malley from Yale University Press deserves special thanks for helping me decide to write the book that I really wanted to write, not something else, and then stay the course.

This book has been more than a decade in the making. Its roots go back to 1993-1994: long nights of conversations, as only graduate students can have, with Niva Elkin Koren about democracy in cyberspace; a series of formative conversations with Mitch Kapor; a couple of madly imaginative sessions with Charlie Nesson; and a moment of true understanding with Eben Moglen. Equally central from around that time, but at an angle, were a paper under Terry Fisher's guidance on nineteenth-century homesteading and the radical republicans, and a series of classes and papers with Frank Michelman, Duncan Kennedy, Mort Horwitz, Roberto Unger, and the late David Charny, which led me to think quite fundamentally about the role of property and economic organization in the construction of human freedom. It was Frank Michelman who taught me that the hard trick was to do so as a liberal.

Since then, I have been fortunate in many and diverse intellectual friendships and encounters, from people in different fields and foci, who shed light on various aspects of this project. I met Larry Lessig for (almost) the first time in 1998. By the end of a two-hour conversation, we had formed a friendship and intellectual conversation that has been central to my work ever since. He has, over the past few years, played a pivotal role in changing the public understanding of control, freedom, and creativity in the digital environment. Over the course of these years, I spent many hours learning from Jamie Boyle, Terry Fisher, and Eben Moglen. In different ways and styles, each of them has had significant influence on my work. There was a moment, sometime between the conference Boyle organized at Yale in 1999 and the one he organized at Duke in 2001, when a range of people who had been doing similar things, pushing against the wind with varying degrees of interconnection, seemed to cohere into a single intellectual movement, centered on the importance of the commons to information production and creativity generally, and to the digitally networked environment in particular. In various contexts, both before this period and since, I have learned much from Julie Cohen, Becky Eisenberg, Bernt Hugenholtz, David Johnson, David Lange, Jessica Litman, Neil Netanel, Helen Nissenbaum, Peggy Radin, Arti Rai, David Post, Jerry Reichman, Pam Samuelson, Jon Zittrain, and Diane Zimmerman. One of the great pleasures of this

Chapter 1
Introduction: A Moment of Opportunity and Challenge

Information, knowledge, and culture are central to human freedom and human development. How they are produced and exchanged in our society critically affects the way we see the state of the world as it is and might be; who decides these questions; and how we, as societies and polities, come to understand what can and ought to be done. For more than 150 years, modern complex democracies have depended in large measure on an industrial information economy for these basic functions. In the past decade and a half, we have begun to see a radical change in the organization of information production. Enabled by technological change, we are beginning to see a series of economic, social, and cultural adaptations that make possible a radical transformation of how we make the information environment we occupy as autonomous individuals, citizens, and members of cultural and social groups. It seems passé today to speak of "the Internet revolution." In some academic circles, it is positively naïve. But it should not be. The change brought about by the networked information environment is deep. It is structural. It goes to the very foundations of how liberal markets and liberal democracies have coevolved for almost two centuries.

A series of changes in the technologies, economic organization, and social practices of production in this environment has created new opportunities for how we make and exchange information, knowledge, and culture. These changes have increased the role of nonmarket and nonproprietary production, both by individuals alone and by cooperative efforts in a wide range of loosely or tightly woven collaborations. These newly emerging practices have seen remarkable success in areas as diverse as software development and investigative reporting, avant-garde video and multiplayer online games. Together, they hint at the emergence of a new information environment, one in which individuals are free to take a more active role than was possible in the industrial information economy of the twentieth century. This new freedom holds great practical promise: as a dimension of individual freedom; as a platform for better democratic participation; as a medium to foster a more critical and self-reflective culture; and, in an increasingly information-dependent global economy, as a mechanism to achieve improvements in human development everywhere.

The rise of greater scope for individual and cooperative nonmarket production of information and culture, however, threatens the incumbents of the industrial information economy. At the beginning of the twenty-first century, we find ourselves in the midst of a battle over the institutional ecology of the digital environment. A wide range of laws and institutions-from broad areas like telecommunications, copyright, or international trade regulation, to minutiae like the rules for registering domain names or whether digital television receivers will be required by law to recognize a particular code-are being tugged and warped in efforts to tilt the playing field toward one way of doing things or the other. How these battles turn out over the next decade or so will likely have a significant effect on how we come to know what is going on in the world we occupy, and to what extent and in what forms we will be able-as autonomous individuals, as citizens, and as participants in cultures and communities-to affect how we and others see the world as it is and as it might be.

The Emergence of the Networked Information Economy

The most advanced economies in the world today have made two parallel shifts that, paradoxically, make possible a significant attenuation of the limitations that market-based production places on the pursuit of the political values central to liberal societies. The first move, in the making for more than a century, is to an economy centered on information (financial services, accounting, software,

science) and cultural (films, music) production, and the manipulation of symbols (from making sneakers to branding them and manufacturing the cultural significance of the Swoosh). The second is the move to a communications environment built on cheap processors with high computation capabilities, interconnected in a pervasive network-the phenomenon we associate with the Internet. It is this second shift that allows for an increasing role for nonmarket production in the information and cultural production sector, organized in a radically more decentralized pattern than was true of this sector in the twentieth century. The first shift means that these new patterns of production-nonmarket and radically decentralized-will emerge, if permitted, at the core, rather than the periphery of the most advanced economies. It promises to enable social production and exchange to play a much larger role, alongside property- and market-based production, than they ever have in modern democracies.

The first part of this book is dedicated to establishing a number of basic economic observations. Its overarching claim is that we are seeing the emergence of a new stage in the information economy, which I call the "networked information economy." It is displacing the industrial information economy that typified information production from about the second half of the nineteenth century and throughout the twentieth century. What characterizes the networked information economy is that decentralized individual action-specifically, new and important cooperative and coordinate action carried out through radically distributed, nonmarket mechanisms that do not depend on proprietary strategies-plays a much greater role than it did, or could have, in the industrial information economy. The catalyst for this change is the happenstance of the fabrication technology of computation, and its ripple effects throughout the technologies of communication and storage. The declining price of computation, communication, and storage have, as a practical matter, placed the material means of information and cultural production in the hands of a significant fraction of the world's population-on the order of a billion people around the globe. The core distinguishing feature of communications, information, and cultural production since the mid-nineteenth century was that effective communication spanning the ever-larger societies and geographies that came to make up the relevant political and economic units of the day required ever-larger investments of physical capital. Large-circulation mechanical presses, the telegraph system, powerful radio and later television transmitters, cable and satellite, and the mainframe computer became necessary to make information and communicate it on scales that went beyond the very local. Wanting to communicate with others was not a sufficient condition to being able to do so. As a result, information and cultural production took on, over the course of this period, a more industrial model than the economics of information itself would have required. The rise of the networked, computer-mediated communications environment has changed this basic fact. The material requirements for effective information production and communication are now owned by numbers of individuals several orders of magnitude larger than the number of owners of the basic means of information production and exchange a mere two decades ago.

The removal of the physical constraints on effective information production has made human creativity and the economics of information itself the core structuring facts in the new networked information economy. These have quite different characteristics than coal, steel, and manual human labor, which characterized the industrial economy and structured our basic thinking about economic production for the past century. They lead to three observations about the emerging information production system. First, nonproprietary strategies have always been more important in information production than they were in the production of steel or automobiles, even when the economics of communication weighed in favor of industrial models. Education, arts and sciences, political debate, and theological disputation have always been much more importantly infused with nonmarket motivations and actors than, say, the automobile industry. As the material barrier that ultimately nonetheless drove much of our information environment to be funneled through the

proprietary, market-based strategies is removed, these basic nonmarket, nonproprietary, motivations and organizational forms should in principle become even more important to the information production system.

Second, we have in fact seen the rise of nonmarket production to much greater importance. Individuals can reach and inform or edify millions around the world. Such a reach was simply unavailable to diversely motivated individuals before, unless they funneled their efforts through either market organizations or philanthropically or state-funded efforts. The fact that every such effort is available to anyone connected to the network, from anywhere, has led to the emergence of coordinate effects, where the aggregate effect of individual action, even when it is not self-consciously cooperative, produces the coordinate effect of a new and rich information environment. One needs only to run a Google search on any subject of interest to see how the "information good" that is the response to one's query is produced by the coordinate effects of the uncoordinated actions of a wide and diverse range of individuals and organizations acting on a wide range of motivations-both market and nonmarket, state-based and nonstate.

Third, and likely most radical, new, and difficult for observers to believe, is the rise of effective, large-scale cooperative efforts-peer production of information, knowledge, and culture. These are typified by the emergence of free and open-source software. We are beginning to see the expansion of this model not only to our core software platforms, but beyond them into every domain of information and cultural production-and this book visits these in many different domains-from peer production of encyclopedias, to news and commentary, to immersive entertainment.

It is easy to miss these changes. They run against the grain of some of our most basic Economics 101 intuitions, intuitions honed in the industrial economy at a time when the only serious alternative seen was state Communism-an alternative almost universally considered unattractive today. The undeniable economic success of free software has prompted some leading-edge economists to try to understand why many thousands of loosely networked free software developers can compete with Microsoft at its own game and produce a massive operating system-GNU/Linux. That growing literature, consistent with its own goals, has focused on software and the particulars of the free and open-source software development communities, although Eric von Hippel's notion of "user-driven innovation" has begun to expand that focus to thinking about how individual need and creativity drive innovation at the individual level, and its diffusion through networks of like-minded individuals. The political implications of free software have been central to the free software movement and its founder, Richard Stallman, and were developed provocatively and with great insight by Eben Moglen. Free software is but one salient example of a much broader phenomenon. Why can fifty thousand volunteers successfully coauthor *Wikipedia,* the most serious online alternative to the *Encyclopedia Britannica,* and then turn around and give it away for free? Why do 4.5 million volunteers contribute their leftover computer cycles to create the most powerful supercomputer on Earth, SETI@Home? Without a broadly accepted analytic model to explain these phenomena, we tend to treat them as curiosities, perhaps transient fads, possibly of significance in one market segment or another. We should try instead to see them for what they are: a new mode of production emerging in the middle of the most advanced economies in the world-those that are the most fully computer networked and for which information goods and services have come to occupy the highest-valued roles.

Human beings are, and always have been, diversely motivated beings. We act instrumentally, but also noninstrumentally. We act for material gain, but also for psychological well-being and gratification, and for social connectedness. There is nothing new or earth-shattering about this, except perhaps to some economists. In the industrial economy in general, and the industrial

information economy as well, most opportunities to make things that were valuable and important to many people were constrained by the physical capital requirements of making them. From the steam engine to the assembly line, from the double-rotary printing press to the communications satellite, the capital constraints on action were such that simply wanting to do something was rarely a sufficient condition to enable one to do it. Financing the necessary physical capital, in turn, oriented the necessarily capital-intensive projects toward a production and organizational strategy that could justify the investments. In market economies, that meant orienting toward market production. In state-run economies, that meant orienting production toward the goals of the state bureaucracy. In either case, the practical individual freedom to cooperate with others in making things of value was limited by the extent of the capital requirements of production.

In the networked information economy, the physical capital required for production is broadly distributed throughout society. Personal computers and network connections are ubiquitous. This does not mean that they cannot be used for markets, or that individuals cease to seek market opportunities. It does mean, however, that whenever someone, somewhere, among the billion connected human beings, and ultimately among all those who will be connected, wants to make something that requires human creativity, a computer, and a network connection, he or she can do so-alone, or in cooperation with others. He or she already has the capital capacity necessary to do so; if not alone, then at least in cooperation with other individuals acting for complementary reasons. The result is that a good deal more that human beings value can now be done by individuals, who interact with each other socially, as human beings and as social beings, rather than as market actors through the price system. Sometimes, under conditions I specify in some detail, these nonmarket collaborations can be better at motivating effort and can allow creative people to work on information projects more efficiently than would traditional market mechanisms and corporations. The result is a flourishing nonmarket sector of information, knowledge, and cultural production, based in the networked environment, and applied to anything that the many individuals connected to it can imagine. Its outputs, in turn, are not treated as exclusive property. They are instead subject to an increasingly robust ethic of open sharing, open for all others to build on, extend, and make their own.

Because the presence and importance of nonmarket production has become so counterintuitive to people living in market-based economies at the end of the twentieth century, part I of this volume is fairly detailed and technical; overcoming what we intuitively "know" requires disciplined analysis. Readers who are not inclined toward economic analysis should at least read the introduction to part I, the segments entitled "When Information Production Meets the Computer Network" and "Diversity of Strategies in our Current Production System" in chapter 2, and the case studies in chapter 3. These should provide enough of an intuitive feel for what I mean by the diversity of production strategies for information and the emergence of nonmarket individual and cooperative production, to serve as the basis for the more normatively oriented parts of the book. Readers who are genuinely skeptical of the possibility that nonmarket production is sustainable and effective, and in many cases is an efficient strategy for information, knowledge, and cultural production, should take the time to read part I in its entirety. The emergence of precisely this possibility and practice lies at the very heart of my claims about the ways in which liberal commitments are translated into lived experiences in the networked environment, and forms the factual foundation of the political-theoretical and the institutional-legal discussion that occupies the remainder of the book.

Networked Information Economy and Liberal, Democratic Societies

How we make information, how we get it, how we speak to others, and how others speak to us are core components of the shape of freedom in any society. Part II of this book provides a detailed

look at how the changes in the technological, economic, and social affordances of the networked information environment affect a series of core commitments of a wide range of liberal democracies. The basic claim is that the diversity of ways of organizing information production and use opens a range of possibilities for pursuing the core political values of liberal societies-individual freedom, a more genuinely participatory political system, a critical culture, and social justice. These values provide the vectors of political morality along which the shape and dimensions of any liberal society can be plotted. Because their practical policy implications are often contradictory, rather than complementary, the pursuit of each places certain limits on how we pursue the others, leading different liberal societies to respect them in different patterns. How much a society constrains the democratic decision-making powers of the majority in favor of individual freedom, or to what extent it pursues social justice, have always been attributes that define the political contours and nature of that society. But the economics of industrial production, and our pursuit of productivity and growth, have imposed a limit on how we can pursue any mix of arrangements to implement our commitments to freedom and justice. Singapore is commonly trotted out as an extreme example of the trade-off of freedom for welfare, but all democracies with advanced capitalist economies have made some such trade-off. Predictions of how well we will be able to feed ourselves are always an important consideration in thinking about whether, for example, to democratize wheat production or make it more egalitarian. Efforts to push workplace democracy have also often foundered on the shoals-real or imagined-of these limits, as have many plans for redistribution in the name of social justice. Market-based, proprietary production has often seemed simply too productive to tinker with. The emergence of the networked information economy promises to expand the horizons of the feasible in political imagination. Different liberal polities can pursue different mixtures of respect for different liberal commitments. However, the overarching constraint represented by the seeming necessity of the industrial model of information and cultural production has significantly shifted as an effective constraint on the pursuit of liberal commitments.

Enhanced Autonomy

The networked information economy improves the practical capacities of individuals along three dimensions: (1) it improves their capacity to do more for and by themselves; (2) it enhances their capacity to do more in loose commonality with others, without being constrained to organize their relationship through a price system or in traditional hierarchical models of social and economic organization; and (3) it improves the capacity of individuals to do more in formal organizations that operate outside the market sphere. This enhanced autonomy is at the core of all the other improvements I describe. Individuals are using their newly expanded practical freedom to act and cooperate with others in ways that improve the practiced experience of democracy, justice and development, a critical culture, and community.

I begin, therefore, with an analysis of the effects of networked information economy on individual autonomy. First, individuals can do more for themselves independently of the permission or cooperation of others. They can create their own expressions, and they can seek out the information they need, with substantially less dependence on the commercial mass media of the twentieth century. Second, and no less importantly, individuals can do more in loose affiliation with others, rather than requiring stable, long-term relations, like coworker relations or participation in formal organizations, to underwrite effective cooperation. Very few individuals living in the industrial information economy could, in any realistic sense, decide to build a new Library of Alexandria of global reach, or to start an encyclopedia. As collaboration among far-flung individuals becomes more common, the idea of doing things that require cooperation with others becomes much more attainable, and the range of projects individuals can choose as their own therefore

qualitatively increases. The very fluidity and low commitment required of any given cooperative relationship increases the range and diversity of cooperative relations people can enter, and therefore of collaborative projects they can conceive of as open to them.

These ways in which autonomy is enhanced require a fairly substantive and rich conception of autonomy as a practical lived experience, rather than the formal conception preferred by many who think of autonomy as a philosophical concept. But even from a narrower perspective, which spans a broader range of conceptions of autonomy, at a minimum we can say that individuals are less susceptible to manipulation by a legally defined class of others-the owners of communications infrastructure and media. The networked information economy provides varied alternative platforms for communication, so that it moderates the power of the traditional mass-media model, where ownership of the means of communication enables an owner to select what others view, and thereby to affect their perceptions of what they can and cannot do. Moreover, the diversity of perspectives on the way the world is and the way it could be for any given individual is qualitatively increased. This gives individuals a significantly greater role in authoring their own lives, by enabling them to perceive a broader range of possibilities, and by providing them a richer baseline against which to measure the choices they in fact make.

Democracy: The Networked Public Sphere

The second major implication of the networked information economy is the shift it enables from the mass-mediated public sphere to a networked public sphere. This shift is also based on the increasing freedom individuals enjoy to participate in creating information and knowledge, and the possibilities it presents for a new public sphere to emerge alongside the commercial, mass-media markets. The idea that the Internet democratizes is hardly new. It has been a staple of writing about the Internet since the early 1990s. The relatively simple first-generation claims about the liberating effects of the Internet, summarized in the U.S. Supreme Court's celebration of its potential to make everyone a pamphleteer, came under a variety of criticisms and attacks over the course of the past half decade or so. Here, I offer a detailed analysis of how the emergence of a networked information economy in particular, as an alternative to mass media, improves the political public sphere. The first-generation critique of the democratizing effect of the Internet was based on various implications of the problem of information overload, or the Babel objection. According to the Babel objection, when everyone can speak, no one can be heard, and we devolve either to a cacophony or to the reemergence of money as the distinguishing factor between statements that are heard and those that wallow in obscurity. The second-generation critique was that the Internet is not as decentralized as we thought in the 1990s. The emerging patterns of Internet use show that very few sites capture an exceedingly large amount of attention, and millions of sites go unnoticed. In this world, the Babel objection is perhaps avoided, but only at the expense of the very promise of the Internet as a democratic medium.

In chapters 6 and 7, I offer a detailed and updated analysis of this, perhaps the best-known and most contentious claim about the Internet's liberalizing effects. First, it is important to understand that any consideration of the democratizing effects of the Internet must measure its effects as compared to the commercial, mass-media-based public sphere, not as compared to an idealized utopia that we embraced a decade ago of how the Internet might be. Commercial mass media that have dominated the public spheres of all modern democracies have been studied extensively. They have been shown in extensive literature to exhibit a series of failures as platforms for public discourse. First, they provide a relatively limited intake basin-that is, too many observations and concerns of too many people in complex modern societies are left unobserved and unattended to by the small cadre of commercial journalists charged with perceiving the range of issues of public

concern in any given society. Second, particularly where the market is concentrated, they give their owners inordinate power to shape opinion and information. This power they can either use themselves or sell to the highest bidder. And third, whenever the owners of commercial media choose not to exercise their power in this way, they then tend to program toward the inane and soothing, rather than toward that which will be politically engaging, and they tend to oversimplify complex public discussions. On the background of these limitations of the mass media, I suggest that the networked public sphere enables many more individuals to communicate their observations and their viewpoints to many others, and to do so in a way that cannot be controlled by media owners and is not as easily corruptible by money as were the mass media.

The empirical and theoretical literature about network topology and use provides answers to all the major critiques of the claim that the Internet improves the structure of the public sphere. In particular, I show how a wide range of mechanisms-starting from the simple mailing list, through static Web pages, the emergence of writable Web capabilities, and mobility-are being embedded in a social system for the collection of politically salient information, observations, and comments, and provide a platform for discourse. These platforms solve some of the basic limitations of the commercial, concentrated mass media as the core platform of the public sphere in contemporary complex democracies. They enable anyone, anywhere, to go through his or her practical life, observing the social environment through new eyes-the eyes of someone who could actually inject a thought, a criticism, or a concern into the public debate. Individuals become less passive, and thus more engaged observers of social spaces that could potentially become subjects for political conversation; they become more engaged participants in the debates about their observations. The various formats of the networked public sphere provide anyone with an outlet to speak, to inquire, to investigate, without need to access the resources of a major media organization. We are seeing the emergence of new, decentralized approaches to fulfilling the watchdog function and to engaging in political debate and organization. These are being undertaken in a distinctly nonmarket form, in ways that would have been much more difficult to pursue effectively, as a standard part of the construction of the public sphere, before the networked information environment. Working through detailed examples, I try to render the optimism about the democratic advantages of the networked public sphere a fully specified argument.

The networked public sphere has also begun to respond to the information overload problem, but without re-creating the power of mass media at the points of filtering and accreditation. There are two core elements to these developments: First, we are beginning to see the emergence of nonmarket, peer-produced alternative sources of filtration and accreditation in place of the market-based alternatives. Relevance and accreditation are themselves information goods, just like software or an encyclopedia. What we are seeing on the network is that filtering for both relevance and accreditation has become the object of widespread practices of mutual pointing, of peer review, of pointing to original sources of claims, and its complement, the social practice that those who have some ability to evaluate the claims in fact do comment on them. The second element is a contingent but empirically confirmed observation of how users actually use the network. As a descriptive matter, information flow in the network is much more ordered than a simple random walk in the cacophony of information flow would suggest, and significantly less centralized than the mass media environment was. Some sites are much more visible and widely read than others. This is true both when one looks at the Web as a whole, and when one looks at smaller clusters of similar sites or users who tend to cluster. Most commentators who have looked at this pattern have interpreted it as a reemergence of mass media-the dominance of the few visible sites. But a full consideration of the various elements of the network topology literature supports a very different interpretation, in which order emerges in the networked environment without re-creating the failures of the mass-media-dominated public sphere. Sites cluster around communities of interest:

Australian fire brigades tend to link to other Australian fire brigades, conservative political blogs (Web logs or online journals) in the United States to other conservative political blogs in the United States, and to a lesser but still significant extent, to liberal political blogs. In each of these clusters, the pattern of some high visibility nodes continues, but as the clusters become small enough, many more of the sites are moderately linked to each other in the cluster. Through this pattern, the network seems to be forming into an attention backbone. "Local" clusters-communities of interest-can provide initial vetting and "peer-review-like" qualities to individual contributions made within an interest cluster. Observations that are seen as significant within a community of interest make their way to the relatively visible sites in that cluster, from where they become visible to people in larger ("regional") clusters. This continues until an observation makes its way to the "superstar" sites that hundreds of thousands of people might read and use. This path is complemented by the practice of relatively easy commenting and posting directly to many of the superstar sites, which creates shortcuts to wide attention. It is fairly simple to grasp intuitively why these patterns might emerge. Users tend to treat other people's choices about what to link to and to read as good indicators of what is worthwhile for them. They are not slavish in this, though; they apply some judgment of their own as to whether certain types of users-say, political junkies of a particular stripe, or fans of a specific television program-are the best predictors of what will be interesting for them. The result is that attention in the networked environment is more dependent on being interesting to an engaged group of people than it is in the mass-media environment, where moderate interest to large numbers of weakly engaged viewers is preferable. Because of the redundancy of clusters and links, and because many clusters are based on mutual interest, not on capital investment, it is more difficult to buy attention on the Internet than it is in mass media outlets, and harder still to use money to squelch an opposing view. These characteristics save the networked environment from the Babel objection without reintroducing excessive power in any single party or small cluster of them, and without causing a resurgence in the role of money as a precondition to the ability to speak publicly.

Justice and Human Development

Information, knowledge, and information-rich goods and tools play a significant role in economic opportunity and human development. While the networked information economy cannot solve global hunger and disease, its emergence does open reasonably well-defined new avenues for addressing and constructing some of the basic requirements of justice and human development. Because the outputs of the networked information economy are usually nonproprietary, it provides free access to a set of the basic instrumentalities of economic opportunity and the basic outputs of the information economy. From a liberal perspective concerned with justice, at a minimum, these outputs become more readily available as "finished goods" to those who are least well off. More importantly, the availability of free information resources makes participating in the economy less dependent on surmounting access barriers to financing and social-transactional networks that made working out of poverty difficult in industrial economies. These resources and tools thus improve equality of opportunity.

From a more substantive and global perspective focused on human development, the freedom to use basic resources and capabilities allows improved participation in the production of information and information-dependent components of human development. First, and currently most advanced, the emergence of a broad range of free software utilities makes it easier for poor and middle-income countries to obtain core software capabilities. More importantly, free software enables the emergence of local capabilities to provide software services, both for national uses and as a basis for participating in a global software services industry, without need to rely on permission from multinational software companies. Scientific publication is beginning to use commons-based

strategies to publish important sources of information in a way that makes the outputs freely available in poorer countries. More ambitiously, we begin to see in agricultural research a combined effort of public, nonprofit, and open-source-like efforts being developed and applied to problems of agricultural innovation. The ultimate purpose is to develop a set of basic capabilities that would allow collaboration among farmers and scientists, in both poor countries and around the globe, to develop better, more nutritious crops to improve food security throughout the poorer regions of the world. Equally ambitious, but less operationally advanced, we are beginning to see early efforts to translate this system of innovation to health-related products.

All these efforts are aimed at solving one of the most glaring problems of poverty and poor human development in the global information economy: Even as opulence increases in the wealthier economies-as information and innovation offer longer and healthier lives that are enriched by better access to information, knowledge, and culture-in many places, life expectancy is decreasing, morbidity is increasing, and illiteracy remains rampant. Some, although by no means all, of this global injustice is due to the fact that we have come to rely ever-more exclusively on proprietary business models of the industrial economy to provide some of the most basic information components of human development. As the networked information economy develops new ways of producing information, whose outputs are not treated as proprietary and exclusive but can be made available freely to everyone, it offers modest but meaningful opportunities for improving human development everywhere. We are seeing early signs of the emergence of an innovation ecosystem made of public funding, traditional nonprofits, and the newly emerging sector of peer production that is making it possible to advance human development through cooperative efforts in both rich countries and poor.

A Critical Culture and Networked Social Relations

The networked information economy also allows for the emergence of a more critical and self-reflective culture. In the past decade, a number of legal scholars-Niva Elkin Koren, Terry Fisher, Larry Lessig, and Jack Balkin-have begun to examine how the Internet democratizes culture. Following this work and rooted in the deliberative strand of democratic theory, I suggest that the networked information environment offers us a more attractive cultural production system in two distinct ways: (1) it makes culture more transparent, and (2) it makes culture more malleable. Together, these mean that we are seeing the emergence of a new folk culture-a practice that has been largely suppressed in the industrial era of cultural production-where many more of us participate actively in making cultural moves and finding meaning in the world around us. These practices make their practitioners better "readers" of their own culture and more self-reflective and critical of the culture they occupy, thereby enabling them to become more self-reflective participants in conversations within that culture. This also allows individuals much greater freedom to participate in tugging and pulling at the cultural creations of others, "glomming on" to them, as Balkin puts it, and making the culture they occupy more their own than was possible with mass-media culture. In these senses, we can say that culture is becoming more democratic: self-reflective and participatory.

Throughout much of this book, I underscore the increased capabilities of individuals as the core driving social force behind the networked information economy. This heightened individual capacity has raised concerns by many that the Internet further fragments community, continuing the long trend of industrialization. A substantial body of empirical literature suggests, however, that we are in fact using the Internet largely at the expense of television, and that this exchange is a good one from the perspective of social ties. We use the Internet to keep in touch with family and intimate friends, both geographically proximate and distant. To the extent we do see a shift in social

ties, it is because, in addition to strengthening our strong bonds, we are also increasing the range and diversity of weaker connections. Following Manuel Castells and Barry Wellman, I suggest that we have become more adept at filling some of the same emotional and context-generating functions that have traditionally been associated with the importance of community with a network of overlapping social ties that are limited in duration or intensity.

Four Methodological Comments

There are four methodological choices represented by the thesis that I have outlined up to this point, and therefore in this book as a whole, which require explication and defense. The first is that I assign a very significant role to technology. The second is that I offer an explanation centered on social relations, but operating in the domain of economics, rather than sociology. The third and fourth are more internal to liberal political theory. The third is that I am offering a liberal political theory, but taking a path that has usually been resisted in that literature-considering economic structure and the limits of the market and its supporting institutions from the perspective of freedom, rather than accepting the market as it is, and defending or criticizing adjustments through the lens of distributive justice. Fourth, my approach heavily emphasizes individual action in nonmarket relations. Much of the discussion revolves around the choice between markets and nonmarket social behavior. In much of it, the state plays no role, or is perceived as playing a primarily negative role, in a way that is alien to the progressive branches of liberal political thought. In this, it seems more of a libertarian or an anarchistic thesis than a liberal one. I do not completely discount the state, as I will explain. But I do suggest that what is special about our moment is the rising efficacy of individuals and loose, nonmarket affiliations as agents of political economy. Just like the market, the state will have to adjust to this new emerging modality of human action. Liberal political theory must first recognize and understand it before it can begin to renegotiate its agenda for the liberal state, progressive or otherwise.

The Role of Technology in Human Affairs

The first methodological choice concerns how one should treat the role of technology in the development of human affairs. The kind of technological determinism that typified Lewis Mumford, or, specifically in the area of communications, Marshall McLuhan, is widely perceived in academia today as being too deterministic, though perhaps not so in popular culture. The contemporary effort to offer more nuanced, institution-based, and political-choice-based explanations is perhaps best typified by Paul Starr's recent and excellent work on the creation of the media. While these contemporary efforts are indeed powerful, one should not confuse a work like Elizabeth Eisenstein's carefully argued and detailed *The Printing Press as an Agent of Change*, with McLuhan's determinism. Assuming that technologies are just tools that happen, more or less, to be there, and are employed in any given society in a pattern that depends only on what that society and culture makes of them is too constrained. A society that has no wheel and no writing has certain limits on what it can do. Barry Wellman has imported into sociology a term borrowed from engineering-affordances.[1] Langdon Winner called these the "political properties" of technologies.[2] An earlier version of this idea is Harold Innis's concept of "the bias of communications."[3] In Internet law and policy debates this approach has become widely adopted through the influential work of Lawrence Lessig, who characterized it as "code is law."[4]

The idea is simple to explain, and distinct from a naïve determinism. Different technologies make different kinds of human action and interaction easier or harder to perform. All other things being equal, things that are easier to do are more likely to be done, and things that are harder to do are less likely to be done. All other things are never equal. That is why technological determinism in the

strict sense-if you have technology "t," you should expect social structure or relation "s" to emerge-is false. Ocean navigation had a different adoption and use when introduced in states whose land empire ambitions were effectively countered by strong neighbors-like Spain and Portugal-than in nations that were focused on building a vast inland empire, like China. Print had different effects on literacy in countries where religion encouraged individual reading-like Prussia, Scotland, England, and New England-than where religion discouraged individual, unmediated interaction with texts, like France and Spain. This form of understanding the role of technology is adopted here. Neither deterministic nor wholly malleable, technology sets some parameters of individual and social action. It can make some actions, relationships, organizations, and institutions easier to pursue, and others harder. In a challenging environment-be the challenges natural or human-it can make some behaviors obsolete by increasing the efficacy of directly competitive strategies. However, within the realm of the feasible-uses not rendered impossible by the adoption or rejection of a technology-different patterns of adoption and use can result in very different social relations that emerge around a technology. Unless these patterns are in competition, or unless even in competition they are not catastrophically less effective at meeting the challenges, different societies can persist with different patterns of use over long periods. It is the feasibility of long-term sustainability of different patterns of use that makes this book relevant to policy, not purely to theory. The same technologies of networked computers can be adopted in very different patterns. There is no guarantee that networked information technology will lead to the improvements in innovation, freedom, and justice that I suggest are possible. That is a choice we face as a society. The way we develop will, in significant measure, depend on choices we make in the next decade or so.

The Role of Economic Analysis and Methodological Individualism

It should be emphasized, as the second point, that this book has a descriptive methodology that is distinctly individualist and economic in orientation, which is hardly the only way to approach this problem. Manuel Castells's magisterial treatment of the networked society[5]locates its central characteristic in the shift from groups and hierarchies to networks as social and organizational models-looser, flexible arrangements of human affairs. Castells develops this theory as he describes a wide range of changes, from transportation networks to globalization and industrialization. In his work, the Internet fits into this trend, enabling better coordination and cooperation in these sorts of loosely affiliated networks. My own emphasis is on the specific relative roles of market and nonmarket sectors, and how that change anchors the radical decentralization that he too observes, as a matter of sociological observation. I place at the core of the shift the technical and economic characteristics of computer networks and information. These provide the pivot for the shift toward radical decentralization of production. They underlie the shift from an information environment dominated by proprietary, market-oriented action, to a world in which nonproprietary, nonmarket transactional frameworks play a large role alongside market production. This newly emerging, nonproprietary sector affects to a substantial degree the entire information environment in which individuals and societies live their lives. If there is one lesson we can learn from globalization and the ever-increasing reach of the market, it is that the logic of the market exerts enormous pressure on existing social structures. If we are indeed seeing the emergence of a substantial component of nonmarket production at the very core of our economic engine-the production and exchange of information, and through it of information-based goods, tools, services, and capabilities-then this change suggests a genuine limit on the extent of the market. Such a limit, growing from within the very market that it limits, in its most advanced loci, would represent a genuine shift in direction for what appeared to be the ever-increasing global reach of the market economy and society in the past half century.

Economic Structure in Liberal Political Theory

The third point has to do with the role of economic structure in liberal political theory. My analysis in this regard is practical and human centric. By this, I mean to say two things: First, I am concerned with human beings, with individuals as the bearers of moral claims regarding the structure of the political and economic systems they inhabit. Within the liberal tradition, the position I take is humanistic and general, as opposed to political and particular. It is concerned first and foremost with the claims of human beings as human beings, rather than with the requirements of democracy or the entitlements of citizenship or membership in a legitimate or meaningfully self-governed political community. There are diverse ways of respecting the basic claims of human freedom, dignity, and well-being. Different liberal polities do so with different mixes of constitutional and policy practices. The rise of global information economic structures and relationships affects human beings everywhere. In some places, it complements democratic traditions. In others, it destabilizes constraints on liberty. An understanding of how we can think of this moment in terms of human freedom and development must transcend the particular traditions, both liberal and illiberal, of any single nation. The actual practice of freedom that we see emerging from the networked environment allows people to reach across national or social boundaries, across space and political division. It allows people to solve problems together in new associations that are outside the boundaries of formal, legal-political association. In this fluid social economic environment, the individual's claims provide a moral anchor for considering the structures of power and opportunity, of freedom and well-being. Furthermore, while it is often convenient and widely accepted to treat organizations or communities as legal entities, as "persons," they are not moral agents. Their role in an analysis of freedom and justice is derivative from their role-both enabling and constraining-as structuring context in which human beings, the actual moral agents of political economy, find themselves. In this regard, my positions here are decidedly "liberal," as opposed to either communitarian or critical.

Second, I am concerned with actual human beings in actual historical settings, not with representations of human beings abstracted from their settings. These commitments mean that freedom and justice for historically situated individuals are measured from a first-person, practical perspective. No constraints on individual freedom and no sources of inequality are categorically exempt from review, nor are any considered privileged under this view. Neither economy nor cultural heritage is given independent moral weight. A person whose life and relations are fully regimented by external forces is unfree, no matter whether the source of regimentation can be understood as market-based, authoritarian, or traditional community values. This does not entail a radical anarchism or libertarianism. Organizations, communities, and other external structures are pervasively necessary for human beings to flourish and to act freely and effectively. This does mean, however, that I think of these structures only from the perspective of their effects on human beings. Their value is purely derivative from their importance to the actual human beings that inhabit them and are structured-for better or worse-by them. As a practical matter, this places concern with market structure and economic organization much closer to the core of questions of freedom than liberal theory usually is willing to do. Liberals have tended to leave the basic structure of property and markets either to libertarians-who, like Friedrich Hayek, accepted its present contours as "natural," and a core constituent element of freedom-or to Marxists and neo-Marxists. I treat property and markets as just one domain of human action, with affordances and limitations. Their presence enhances freedom along some dimensions, but their institutional requirements can become sources of constraint when they squelch freedom of action in nonmarket contexts. Calibrating the reach of the market, then, becomes central not only to the shape of justice or welfare in a society, but also to freedom.

Whither the State?

The fourth and last point emerges in various places throughout this book, but deserves explicit note here. What I find new and interesting about the networked information economy is the rise of individual practical capabilities, and the role that these new capabilities play in increasing the relative salience of nonproprietary, often nonmarket individual and social behavior. In my discussion of autonomy and democracy, of justice and a critical culture, I emphasize the rise of individual and cooperative private action and the relative decrease in the dominance of market-based and proprietary action. Where in all this is the state? For the most part, as you will see particularly in chapter 11, the state in both the United States and Europe has played a role in supporting the market-based industrial incumbents of the twentieth-century information production system at the expense of the individuals who make up the emerging networked information economy. Most state interventions have been in the form of either captured legislation catering to incumbents, or, at best, well-intentioned but wrongheaded efforts to optimize the institutional ecology for outdated modes of information and cultural production. In the traditional mapping of political theory, a position such as the one I present here-that freedom and justice can and should best be achieved by a combination of market action and private, voluntary (not to say charitable) nonmarket action, and that the state is a relatively suspect actor-is libertarian. Perhaps, given that I subject to similar criticism rules styled by their proponents as "property"-like "intellectual property" or "spectrum property rights"-it is anarchist, focused on the role of mutual aid and highly skeptical of the state. (It is quite fashionable nowadays to be libertarian, as it has been for a few decades, and more fashionable to be anarchist than it has been in a century.)

The more modest truth is that my position is not rooted in a theoretical skepticism about the state, but in a practical diagnosis of opportunities, barriers, and strategies for achieving improvements in human freedom and development given the actual conditions of technology, economy, and politics. I have no objection in principle to an effective, liberal state pursuing one of a range of liberal projects and commitments. Here and there throughout this book you will encounter instances where I suggest that the state could play constructive roles, if it stopped listening to incumbents for long enough to realize this. These include, for example, municipal funding of neutral broadband networks, state funding of basic research, and possible strategic regulatory interventions to negate monopoly control over essential resources in the digital environment. However, the necessity for the state's affirmative role is muted because of my diagnosis of the particular trajectory of markets, on the one hand, and individual and social action, on the other hand, in the digitally networked information environment. The particular economics of computation and communications; the particular economics of information, knowledge, and cultural production; and the relative role of information in contemporary, advanced economies have coalesced to make nonmarket individual and social action the most important domain of action in the furtherance of the core liberal commitments. Given these particular characteristics, there is more freedom to be found through opening up institutional spaces for voluntary individual and cooperative action than there is in intentional public action through the state. Nevertheless, I offer no particular reasons to resist many of the roles traditionally played by the liberal state. I offer no reason to think that, for example, education should stop being primarily a state-funded, public activity and a core responsibility of the liberal state, or that public health should not be so. I have every reason to think that the rise of nonmarket production enhances, rather than decreases, the justifiability of state funding for basic science and research, as the spillover effects of publicly funded information production can now be much greater and more effectively disseminated and used to enhance the general welfare.

The important new fact about the networked environment, however, is the efficacy and centrality of individual and collective social action. In most domains, freedom of action for individuals, alone

and in loose cooperation with others, can achieve much of the liberal desiderata I consider throughout this book. From a global perspective, enabling individuals to act in this way also extends the benefits of liberalization across borders, increasing the capacities of individuals in nonliberal states to grab greater freedom than those who control their political systems would like. By contrast, as long as states in the most advanced market-based economies continue to try to optimize their institutional frameworks to support the incumbents of the industrial information economy, they tend to threaten rather than support liberal commitments. Once the networked information economy has stabilized and we come to understand the relative importance of voluntary private action outside of markets, the state can begin to adjust its policies to facilitate nonmarket action and to take advantage of its outputs to improve its own support for core liberal commitments.

The Stakes of It All: The Battle Over the Institutional Ecology of the Digital Environment

No benevolent historical force will inexorably lead this technological-economic moment to develop toward an open, diverse, liberal equilibrium. If the transformation I describe as possible occurs, it will lead to substantial redistribution of power and money from the twentieth-century industrial producers of information, culture, and communications-like Hollywood, the recording industry, and perhaps the broadcasters and some of the telecommunications services giants-to a combination of widely diffuse populations around the globe, and the market actors that will build the tools that make this population better able to produce its own information environment rather than buying it ready-made. None of the industrial giants of yore are taking this reallocation lying down. The technology will not overcome their resistance through an insurmountable progressive impulse. The reorganization of production and the advances it can bring in freedom and justice will emerge, therefore, only as a result of social and political action aimed at protecting the new social patterns from the incumbents' assaults. It is precisely to develop an understanding of what is at stake and why it is worth fighting for that I write this book. I offer no reassurances, however, that any of this will in fact come to pass.

The battle over the relative salience of the proprietary, industrial models of information production and exchange and the emerging networked information economy is being carried out in the domain of the institutional ecology of the digital environment. In a wide range of contexts, a similar set of institutional questions is being contested: To what extent will resources necessary for information production and exchange be governed as a commons, free for all to use and biased in their availability in favor of none? To what extent will these resources be entirely proprietary, and available only to those functioning within the market or within traditional forms of well-funded nonmarket action like the state and organized philanthropy? We see this battle played out at all layers of the information environment: the physical devices and network channels necessary to communicate; the existing information and cultural resources out of which new statements must be made; and the logical resources-the software and standards-necessary to translate what human beings want to say to each other into signals that machines can process and transmit. Its central question is whether there will, or will not, be a core common infrastructure that is governed as a commons and therefore available to anyone who wishes to participate in the networked information environment outside of the market-based, proprietary framework.

This is not to say that property is in some sense inherently bad. Property, together with contract, is the core institutional component of markets, and a core institutional element of liberal societies. It is what enables sellers to extract prices from buyers, and buyers to know that when they pay, they will be secure in their ability to use what they bought. It underlies our capacity to plan actions that require use of resources that, without exclusivity, would be unavailable for us to use. But property

also constrains action. The rules of property are circumscribed and intended to elicit a particular datum-willingness and ability to pay for exclusive control over a resource. They constrain what one person or another can do with regard to a resource; that is, use it in some ways but not others, reveal or hide information with regard to it, and so forth. These constraints are necessary so that people must transact with each other through markets, rather than through force or social networks, but they do so at the expense of constraining action outside of the market to the extent that it depends on access to these resources.

Commons are another core institutional component of freedom of action in free societies, but they are structured to enable action that is not based on exclusive control over the resources necessary for action. For example, I can plan an outdoor party with some degree of certainty by renting a private garden or beach, through the property system. Alternatively, I can plan to meet my friends on a public beach or at Sheep's Meadow in Central Park. I can buy an easement from my neighbor to reach a nearby river, or I can walk around her property using the public road that makes up our transportation commons. Each institutional framework-property and commons-allows for a certain freedom of action and a certain degree of predictability of access to resources. Their complementary coexistence and relative salience as institutional frameworks for action determine the relative reach of the market and the domain of nonmarket action, both individual and social, in the resources they govern and the activities that depend on access to those resources. Now that material conditions have enabled the emergence of greater scope for nonmarket action, the scope and existence of a core common infrastructure that includes the basic resources necessary to produce and exchange information will shape the degree to which individuals will be able to act in all the ways that I describe as central to the emergence of a networked information economy and the freedoms it makes possible.

At the physical layer, the transition to broadband has been accompanied by a more concentrated market structure for physical wires and connections, and less regulation of the degree to which owners can control the flow of information on their networks. The emergence of open wireless networks, based on "spectrum commons," counteracts this trend to some extent, as does the current apparent business practice of broadband owners not to use their ownership to control the flow of information over their networks. Efforts to overcome the broadband market concentration through the development of municipal broadband networks are currently highly contested in legislation and courts. The single most threatening development at the physical layer has been an effort driven primarily by Hollywood, over the past few years, to require the manufacturers of computation devices to design their systems so as to enforce the copyright claims and permissions imposed by the owners of digital copyrighted works. Should this effort succeed, the core characteristic of computers-that they are general-purpose devices whose abilities can be configured and changed over time by their owners as uses and preferences change-will be abandoned in favor of machines that can be trusted to perform according to factory specifications, irrespective of what their owners wish. The primary reason that these laws have not yet passed, and are unlikely to pass, is that the computer hardware and software, and electronics and telecommunications industries all understand that such a law would undermine their innovation and creativity. At the logical layer, we are seeing a concerted effort, again headed primarily by Hollywood and the recording industry, to shape the software and standards to make sure that digitally encoded cultural products can continue to be sold as packaged goods. The Digital Millennium Copyright Act and the assault on peer-to-peer technologies are the most obvious in this regard.

More generally information, knowledge, and culture are being subjected to a second enclosure movement, as James Boyle has recently explored in depth. The freedom of action for individuals who wish to produce information, knowledge, and culture is being systematically curtailed in order

to secure the economic returns demanded by the manufacturers of the industrial information economy. A rich literature in law has developed in response to this increasing enclosure over the past twenty years. It started with David Lange's evocative exploration of the public domain and Pamela Samuelson's prescient critique of the application of copyright to computer programs and digital materials, and continued through Jessica Litman's work on the public domain and digital copyright and Boyle's exploration of the basic romantic assumptions underlying our emerging "intellectual property" construct and the need for an environmentalist framework for preserving the public domain. It reached its most eloquent expression in Lawrence Lessig's arguments for the centrality of free exchange of ideas and information to our most creative endeavors, and his diagnoses of the destructive effects of the present enclosure movement. This growing skepticism among legal academics has been matched by a long-standing skepticism among economists (to which I devote much discussion in chapter 2). The lack of either analytic or empirical foundation for the regulatory drive toward ever-stronger proprietary rights has not, however, resulted in a transformed politics of the regulation of intellectual production. Only recently have we begun to see a politics of information policy and "intellectual property" emerge from a combination of popular politics among computer engineers, college students, and activists concerned with the global poor; a reorientation of traditional media advocates; and a very gradual realization by high-technology firms that rules pushed by Hollywood can impede the growth of computer-based businesses. This political countermovement is tied to quite basic characteristics of the technology of computer communications, and to the persistent and growing social practices of sharing-some, like p2p (peer-to-peer) file sharing, in direct opposition to proprietary claims; others, increasingly, are instances of the emerging practices of making information on nonproprietary models and of individuals sharing what they themselves made in social, rather than market patterns. These economic and social forces are pushing at each other in opposite directions, and each is trying to mold the legal environment to better accommodate its requirements. We still stand at a point where information production could be regulated so that, for most users, it will be forced back into the industrial model, squelching the emerging model of individual, radically decentralized, and nonmarket production and its attendant improvements in freedom and justice.

Social and economic organization is not infinitely malleable. Neither is it always equally open to affirmative design. The actual practices of human interaction with information, knowledge, and culture and with production and consumption are the consequence of a feedback effect between social practices, economic organization, technological affordances, and formal constraints on behavior through law and similar institutional forms. These components of the constraints and affordances of human behavior tend to adapt dynamically to each other, so that the tension between the technological affordances, the social and economic practices, and the law are often not too great. During periods of stability, these components of the structure within which human beings live are mostly aligned and mutually reinforce each other, but the stability is subject to shock at any one of these dimensions. Sometimes shock can come in the form of economic crisis, as it did in the United States during the Great Depression. Often it can come from an external physical threat to social institutions, like a war. Sometimes, though probably rarely, it can come from law, as, some would argue, it came from the desegregation decision in *Brown v. Board of Education*. Sometimes it can come from technology; the introduction of print was such a perturbation, as was, surely, the steam engine. The introduction of the high-capacity mechanical presses and telegraph ushered in the era of mass media. The introduction of radio created a similar perturbation, which for a brief moment destabilized the mass-media model, but quickly converged to it. In each case, the period of perturbation offered more opportunities and greater risks than the periods of relative stability. During periods of perturbation, more of the ways in which society organizes itself are up for grabs; more can be renegotiated, as the various other components of human stability adjust to the changes. To borrow Stephen Jay Gould's term from evolutionary theory, human societies exist in a

series of punctuated equilibria. The periods of disequilibrium are not necessarily long. A mere twenty-five years passed between the invention of radio and its adaptation to the mass-media model. A similar period passed between the introduction of telephony and its adoption of the monopoly utility form that enabled only one-to-one limited communications. In each of these periods, various paths could have been taken. Radio showed us even within the past century how, in some societies, different paths were in fact taken and then sustained over decades. After a period of instability, however, the various elements of human behavioral constraint and affordances settled on a new stable alignment. During periods of stability, we can probably hope for little more than tinkering at the edges of the human condition.

This book is offered, then, as a challenge to contemporary liberal democracies. We are in the midst of a technological, economic, and organizational transformation that allows us to renegotiate the terms of freedom, justice, and productivity in the information society. How we shall live in this new environment will in some significant measure depend on policy choices that we make over the next decade or so. To be able to understand these choices, to be able to make them well, we must recognize that they are part of what is fundamentally a social and political choice-a choice about how to be free, equal, productive human beings under a new set of technological and economic conditions. As economic policy, allowing yesterday's winners to dictate the terms of tomorrow's economic competition would be disastrous. As social policy, missing an opportunity to enrich democracy, freedom, and justice in our society while maintaining or even enhancing our productivity would be unforgivable.

Notes

1. Barry Wellman et al., "The Social Affordances of the Internet for Networked Individualism," *JCMC* 8, no. 3 (April 2003).

2. Langdon Winner, ed., "Do Artifacts Have Politics?" in *The Whale and The Reactor: A Search for Limits in an Age of High Technology* (Chicago: University of Chicago Press, 1986), 19-39.

3. Harold Innis, *The Bias of Communication* (Toronto: University of Toronto Press, 1951). Innis too is often lumped with McLuhan and Walter Ong as a technological determinist. His work was, however, one of a political economist, and he emphasized the relationship between technology and economic and social organization, much more than the deterministic operation of technology on human cognition and capability.

4. Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999).

5. Manuel Castells, *The Rise of Networked Society* (Cambridge, MA, and Oxford: Blackwell Publishers, 1996).

Part I
The Networked Information Economy

For more than 150 years, new communications technologies have tended to concentrate and commercialize the production and exchange of information, while extending the geographic and social reach of information distribution networks. High-volume mechanical presses and the telegraph combined with new business practices to change newspapers from small-circulation local efforts into mass media. Newspapers became means of communications intended to reach ever-larger and more dispersed audiences, and their management required substantial capital investment. As the size of the audience and its geographic and social dispersion increased, public discourse developed an increasingly one-way model. Information and opinion that was widely known and formed the shared basis for political conversation and broad social relations flowed from ever more capital-intensive commercial and professional producers to passive, undifferentiated consumers. It was a model easily adopted and amplified by radio, television, and later cable and satellite communications. This trend did not cover all forms of communication and culture. Telephones and personal interactions, most importantly, and small-scale distributions, like mimeographed handbills, were obvious alternatives. Yet the growth of efficient transportation and effective large-scale managerial and administrative structures meant that the sources of effective political and economic power extended over larger geographic areas and required reaching a larger and more geographically dispersed population. The economics of long-distance mass distribution systems necessary to reach this constantly increasing and more dispersed relevant population were typified by high up-front costs and low marginal costs of distribution. These cost characteristics drove cultural production toward delivery to ever-wider audiences of increasingly high production-value goods, whose fixed costs could be spread over ever-larger audiences - like television series, recorded music, and movies. Because of these economic characteristics, the mass-media model of information and cultural production and transmission became the dominant form of public communication in the twentieth century.

The Internet presents the possibility of a radical reversal of this long trend. It is the first modern communications medium that expands its reach by decentralizing the capital structure of production and distribution of information, culture, and knowledge. Much of the physical capital that embeds most of the intelligence in the network is widely diffused and owned by end users. Network routers and servers are not qualitatively different from the computers that end users own, unlike broadcast stations or cable systems, which are radically different in economic and technical terms from the televisions that receive their signals. This basic change in the material conditions of information and cultural production and distribution have substantial effects on how we come to know the world we occupy and the alternative courses of action open to us as individuals and as social actors. Through these effects, the emerging networked environment structures how we perceive and pursue core values in modern liberal societies.

Technology alone does not, however, determine social structure. The introduction of print in China and Korea did not induce the kind of profound religious and political reformation that followed the printed Bible and disputations in Europe. But technology is not irrelevant, either. Luther's were not the first disputations nailed to a church door. Print, however, made it practically feasible for more than 300,000 copies of Luther's publications to be circulated between 1517 and 1520 in a way that earlier disputations could not have been.[1] Vernacular reading of the Bible became a feasible form of religious self-direction only when printing these Bibles and making them available to individual households became economically feasible, and not when all copyists were either monks or

otherwise dependent on the church. Technology creates feasibility spaces for social practice. Some things become easier and cheaper, others harder and more expensive to do or to prevent under different technological conditions. The interaction between these technological-economic feasibility spaces, and the social responses to these changes - both in terms of institutional changes, like law and regulation, and in terms of changing social practices - define the qualities of a period. The way life is actually lived by people within a given set of interlocking technological, economic, institutional, and social practices is what makes a society attractive or unattractive, what renders its practices laudable or lamentable.

A particular confluence of technical and economic changes is now altering the way we produce and exchange information, knowledge, and culture in ways that could redefine basic practices, first in the most advanced economies, and eventually around the globe. The potential break from the past 150 years is masked by the somewhat liberal use of the term "information economy" in various permutations since the 1970s. The term has been used widely to signify the dramatic increase in the importance of usable information as a means of controlling production and the flow of inputs, outputs, and services. While often evoked as parallel to the "postindustrial" stage, in fact, the information economy was tightly linked throughout the twentieth century with controlling the processes of the industrial economy. This is clearest in the case of accounting firms and financial markets, but is true of the industrial modalities of organizing cultural production as well. Hollywood, the broadcast networks, and the recording industry were built around a physical production model. Once the cultural utterances, the songs or movies, were initially produced and fixed in some means of storage and transmission, the economics of production and distribution of these physical goods took over. Making the initial utterances and the physical goods that embodied them required high capital investment up front. Making many copies was not much more expensive than making few copies, and very much cheaper on a per-copy basis. These industries therefore organized themselves to invest large sums in making a small number of high production-value cultural "artifacts," which were then either replicated and stamped onto many low-cost copies of each artifact, or broadcast or distributed through high-cost systems for low marginal cost ephemeral consumption on screens and with receivers. This required an effort to manage demand for those products that were in fact recorded and replicated or distributed, so as to make sure that the producers could sell many units of a small number of cultural utterances at a low per-unit cost, rather than few units each of many cultural utterances at higher per-unit costs. Because of its focus around capital-intensive production and distribution techniques, this first stage might best be thought of as the "industrial information economy."

Radical decentralization of intelligence in our communications network and the centrality of information, knowledge, culture, and ideas to advanced economic activity are leading to a new stage of the information economy - the networked information economy. In this new stage, we can harness many more of the diverse paths and mechanisms for cultural transmission that were muted by the economies of scale that led to the rise of the concentrated, controlled form of mass media, whether commercial or state-run. The most important aspect of the networked information economy is the possibility it opens for reversing the control focus of the industrial information economy. In particular, it holds out the possibility of reversing two trends in cultural production central to the project of control: concentration and commercialization.

Two fundamental facts have changed in the economic ecology in which the industrial information enterprises have arisen. First, the basic output that has become dominant in the most advanced economies is human meaning and communication. Second, the basic physical capital necessary to express and communicate human meaning is the connected personal computer. The core functionalities of processing, storage, and communications are widely owned throughout the

population of users. Together, these changes destabilize the industrial stage of the information economy. Both the capacity to make meaning - to encode and decode humanly meaningful statements - and the capacity to communicate one's meaning around the world, are held by, or readily available to, at least many hundreds of millions of users around the globe. Any person who has information can connect with any other person who wants it, and anyone who wants to make it mean something in some context, can do so. The high capital costs that were a prerequisite to gathering, working, and communicating information, knowledge, and culture, have now been widely distributed in the society. The entry barrier they posed no longer offers a condensation point for the large organizations that once dominated the information environment. Instead, emerging models of information and cultural production, radically decentralized and based on emergent patterns of cooperation and sharing, but also of simple coordinate coexistence, are beginning to take on an ever-larger role in how we produce meaning - information, knowledge, and culture - in the networked information economy.

A Google response to a query, which returns dozens or more sites with answers to an information question you may have, is an example of coordinate coexistence producing information. As Jessica Litman demonstrated in *Sharing and Stealing*, hundreds of independent producers of information, acting for reasons ranging from hobby and fun to work and sales, produce information, independently and at widely varying costs, related to what you were looking for. They all coexist without knowing of each other, most of them without thinking or planning on serving you in particular, or even a class of user like you. Yet the sheer volume and diversity of interests and sources allows their distributed, unrelated efforts to be coordinated - through the Google algorithm in this case, but also through many others - into a picture that has meaning and provides the answer to your question. Other, more deeply engaged and cooperative enterprises are also emerging on the Internet. Wikipedia, a multilingual encyclopedia coauthored by fifty thousand volunteers, is one particularly effective example of many such enterprises.

The technical conditions of communication and information processing are enabling the emergence of new social and economic practices of information and knowledge production. Eisenstein carefully documented how print loosened the power of the church over information and knowledge production in Europe, and enabled, particularly in the Protestant North, the emergence of early modern capitalist enterprises in the form of print shops. These printers were able to use their market revenues to become independent of the church or the princes, as copyists never were, and to form the economic and social basis of a liberal, market-based freedom of thought and communication. Over the past century and a half, these early printers turned into the commercial mass media: A particular type of market-based production-concentrated, largely homogenous, and highly commercialized - that came to dominate our information environment by the end of the twentieth century. On the background of that dominant role, the possibility that a radically different form of information production will emerge - decentralized; socially, no less than commercially, driven; and as diverse as human thought itself - offers the promise of a deep change in how we see the world around us, how we come to know about it and evaluate it, and how we are capable of communicating with others about what we know, believe, and plan.

This part of the book is dedicated to explaining the technological-economic transformation that is making these practices possible. Not because economics drives all; not because technology determines the way society or communication go; but because it is the technological shock, combined with the economic sustainability of the emerging social practices, that creates the new set of social and political opportunities that are the subject of this book. By working out the economics of these practices, we can understand the economic parameters within which practical political imagination and fulfillment can operate in the digitally networked environment. I describe sustained

productive enterprises that take the form of decentralized and nonmarket-based production, and explain why productivity and growth are consistent with a shift toward such modes of production. What I describe is not an exercise in pastoral utopianism. It is not a vision of a return to production in a preindustrial world. It is a practical possibility that directly results from our economic understanding of information and culture as objects of production. It flows from fairly standard economic analysis applied to a very nonstandard economic reality: one in which all the means of producing and exchanging information and culture are placed in the hands of hundreds of millions, and eventually billions, of people around the world, available for them to work with not only when they are functioning in the market to keep body and soul together, but also, and with equal efficacy, when they are functioning in society and alone, trying to give meaning to their lives as individuals and as social beings.

Chapter 2: Some Basic Economics of Information Production and Innovation

The Diversity of Strategies in our Current Information Production System

The Effects of Exclusive Rights

When Information Production Meets the Computer Network

Strong Exclusive Rights in the Digital Environment

Chapter 3: Peer Production and Sharing

Free/Open-Source Software

Peer Production of Information, Knowledge, and Culture Generally

Chapter 4: The Economics of Social Production

Motivation

Social Production: Feasibility Conditions and Organizational Form

Transaction Costs and Efficiency

The Emergence of Social Production in the Digitally Networked Environment

The Interface of Social Production and Market-Based Businesses

Notes

1. Elizabeth Eisenstein, *Printing Press as an Agent of Change* (Cambridge: Cambridge University Press, 1979).

The Wealth of Networks:
How Social Production Transforms Markets and Freedom
by Yochai Benkler, Yale University Press

Chapter 2
Some Basic Economics of Information Production and Innovation

The Diversity of Strategies in our Current Information Production System

The Effects of Exclusive Rights

When Information Production Meets the Computer Network

Strong Exclusive Rights in the Digital Environment

Epigraph

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Chapter 2
Some Basic Economics of Information Production and Innovation

There are no noncommercial automobile manufacturers. There are no volunteer steel foundries. You would never choose to have your primary source of bread depend on voluntary contributions from others. Nevertheless, scientists working at noncommercial research institutes funded by nonprofit educational institutions and government grants produce most of our basic science. Widespread cooperative networks of volunteers write the software and standards that run most of the Internet and enable what we do with it. Many people turn to National Public Radio or the BBC as a reliable source of news. What is it about information that explains this difference? Why do we rely almost exclusively on markets and commercial firms to produce cars, steel, and wheat, but much less so for the most critical information our advanced societies depend on? Is this a historical contingency, or is there something about information as an object of production that makes nonmarket production attractive?

The technical economic answer is that certain characteristics of information and culture lead us to understand them as "public goods," rather than as "pure private goods" or standard "economic goods." When economists speak of information, they usually say that it is "nonrival." We consider a good to be nonrival when its consumption by one person does not make it any less available for consumption by another. Once such a good is produced, no more social resources need be invested in creating more of it to satisfy the next consumer. Apples are rival. If I eat this apple, you cannot eat it. If you nonetheless want to eat an apple, more resources (trees, labor) need to be diverted from, say, building chairs, to growing apples, to satisfy you. The social cost of your consuming the second apple is the cost of not using the resources needed to grow the second apple (the wood from the tree) in their next best use. In other words, it is the cost to society of not having the additional chairs that could have been made from the tree. Information is nonrival. Once a scientist has established a fact, or once Tolstoy has written *War and Peace*, neither the scientist nor Tolstoy need spend a single second on producing additional *War and Peace* manuscripts or studies for the one-hundredth, one-thousandth, or one-millionth user of what they wrote. The physical paper for the book or journal costs something, but the information itself need only be created once. Economists call such goods "public" because a market will not produce them if priced at their marginal cost - zero. In order to provide Tolstoy or the scientist with income, we regulate publishing: We pass laws that enable their publishers to prevent competitors from entering the market. Because no competitors are permitted into the market for copies of *War and Peace*, the publishers can price the contents of the book or journal at above their actual marginal cost of zero. They can then turn some of that excess revenue over to Tolstoy. Even if these laws are therefore necessary to create the incentives for publication, the market that develops based on them will, from the technical economic perspective, systematically be inefficient. As Kenneth Arrow put it in 1962, "precisely to the extent that [property] is effective, there is underutilization of the information."[/1] Because welfare economics defines a market as producing a good efficiently only when it is pricing the good at its marginal cost, a good like information (and culture and knowledge are, for purposes of economics, forms of information), which can never be sold both at a positive (greater than zero) price and at its marginal cost, is fundamentally a candidate for substantial nonmarket production.

This widely held explanation of the economics of information production has led to an understanding that markets based on patents or copyrights involve a trade-off between static and

dynamic efficiency. That is, looking at the state of the world on any given day, it is inefficient that people and firms sell the information they possess. From the perspective of a society's overall welfare, the most efficient thing would be for those who possess information to give it away for free - or rather, for the cost of communicating it and no more. On any given day, enforcing copyright law leads to inefficient underutilization of copyrighted information. However, looking at the problem of information production over time, the standard defense of exclusive rights like copyright expects firms and people not to produce if they know that their products will be available for anyone to take for free. In order to harness the efforts of individuals and firms that want to make money, we are willing to trade off some static inefficiency to achieve dynamic efficiency. That is, we are willing to have some inefficient lack of access to information every day, in exchange for getting more people involved in information production over time. Authors and inventors or, more commonly, companies that contract with musicians and filmmakers, scientists, and engineers, will invest in research and create cultural goods because they expect to sell their information products. Over time, this incentive effect will give us more innovation and creativity, which will outweigh the inefficiency at any given moment caused by selling the information at above its marginal cost. This defense of exclusive rights is limited by the extent to which it correctly describes the motivations of information producers and the business models open to them to appropriate the benefits of their investments. If some information producers do not need to capture the economic benefits of their particular information outputs, or if some businesses can capture the economic value of their information production by means other than exclusive control over their products, then the justification for regulating access by granting copyrights or patents is weakened. As I will discuss in detail, both of these limits on the standard defense are in fact the case.

Nonrivalry, moreover, is not the only quirky characteristic of information production as an economic phenomenon. The other crucial quirkiness is that information is both input and output of its own production process. In order to write today's academic or news article, I need access to yesterday's articles and reports. In order to write today's novel, movie, or song, I need to use and rework existing cultural forms, such as story lines and twists. This characteristic is known to economists as the "on the shoulders of giants" effect, recalling a statement attributed to Isaac Newton: "If I have seen farther it is because I stand on the shoulders of giants."[2] This second quirkiness of information as a production good makes property-like exclusive rights less appealing as the dominant institutional arrangement for information and cultural production than it would have been had the sole quirky characteristic of information been its nonrivalry. The reason is that if any new information good or innovation builds on existing information, then strengthening intellectual property rights increases the prices that those who invest in producing information today must pay to those who did so yesterday, in addition to increasing the rewards an information producer can get tomorrow. Given the nonrivalry, those payments made today for yesterday's information are all inefficiently too high, from today's perspective. They are all above the marginal cost - zero. Today's users of information are not only today's readers and consumers. They are also today's producers and tomorrow's innovators. Their net benefit from a strengthened patent or copyright regime, given not only increased potential revenues but also the increased costs, may be negative. If we pass a law that regulates information production too strictly, allowing its beneficiaries to impose prices that are too high on today's innovators, then we will have not only too little consumption of information today, but also too little production of new information for tomorrow.

Perhaps the most amazing document of the consensus among economists today that, because of the combination of nonrivalry and the "on the shoulders of giants" effect, excessive expansion of "intellectual property" protection is economically detrimental, was the economists' brief filed in the Supreme Court case of *Eldred v. Ashcroft.*[3] The case challenged a law that extended the term of

copyright protection from lasting for the life of the author plus fifty years, to life of the author plus seventy years, or from seventy-five years to ninety-five years for copyrights owned by corporations. If information were like land or iron, the ideal length of property rights would be infinite from the economists' perspective. In this case, however, where the "property right" was copyright, more than two dozen leading economists volunteered to sign a brief opposing the law, counting among their number five Nobel laureates, including that well-known market skeptic, Milton Friedman.

The efficiency of regulating information, knowledge, and cultural production through strong copyright and patent is not only theoretically ambiguous, it also lacks empirical basis. The empirical work trying to assess the impact of intellectual property on innovation has focused to date on patents. The evidence provides little basis to support stronger and increasing exclusive rights of the type we saw in the last two and a half decades of the twentieth century. Practically no studies show a clear-cut benefit to stronger or longer patents.[/4] In perhaps one of the most startling papers on the economics of innovation published in the past few years, Josh Lerner looked at changes in intellectual property law in sixty countries over a period of 150 years. He studied close to three hundred policy changes, and found that, both in developing countries and in economically advanced countries that already have patent law, patenting both at home and abroad by domestic firms of the country that made the policy change, a proxy for their investment in research and development, decreases slightly when patent law is strengthened![/5] The implication is that when a country - either one that already has a significant patent system, or a developing nation - increases its patent protection, it slightly decreases the level of investment in innovation by local firms. Going on intuitions alone, without understanding the background theory, this seems implausible - why would inventors or companies innovate less when they get more protection? Once you understand the interaction of nonrivalry and the "on the shoulders of giants" effect, the findings are entirely consistent with theory. Increasing patent protection, both in developing nations that are net importers of existing technology and science, and in developed nations that already have a degree of patent protection, and therefore some nontrivial protection for inventors, increases the costs that current innovators have to pay on existing knowledge more than it increases their ability to appropriate the value of their own contributions. When one cuts through the rent-seeking politics of intellectual property lobbies like the pharmaceutical companies or Hollywood and the recording industry; when one overcomes the honestly erroneous, but nonetheless conscience-soothing beliefs of lawyers who defend the copyright and patent-dependent industries and the judges they later become, the reality of both theory and empirics in the economics of intellectual property is that both in theory and as far as empirical evidence shows, there is remarkably little support in economics for regulating information, knowledge, and cultural production through the tools of intellectual property law.

Where does innovation and information production come from, then, if it does not come as much from intellectual-property-based market actors, as many generally believe? The answer is that it comes mostly from a mixture of (1) nonmarket sources - both state and nonstate - and (2) market actors whose business models do not depend on the regulatory framework of intellectual property. The former type of producer is the expected answer, within mainstream economics, for a public goods problem like information production. The National Institutes of Health, the National Science Foundation, and the Defense Department are major sources of funding for research in the United States, as are government agencies in Europe, at the national and European level, Japan, and other major industrialized nations. The latter type - that is, the presence and importance of market-based producers whose business models do not require and do not depend on intellectual property protection - is not theoretically predicted by that model, but is entirely obvious once you begin to think about it.

Consider a daily newspaper. Normally, we think of newspapers as dependent on copyrights. In fact, however, that would be a mistake. No daily newspaper would survive if it depended for its business on waiting until a competitor came out with an edition, then copied the stories, and reproduced them in a competing edition. Daily newspapers earn their revenue from a combination of low-priced newsstand sales or subscriptions together with advertising revenues. Neither of those is copyright dependent once we understand that consumers will not wait half a day until the competitor's paper comes out to save a nickel or a quarter on the price of the newspaper. If all copyright on newspapers were abolished, the revenues of newspapers would be little affected./6 Take, for example, the 2003 annual reports of a few of the leading newspaper companies in the United States. The New York Times Company receives a little more than $3 billion a year from advertising and circulation revenues, and a little more than $200 million a year in revenues from all other sources. Even if the entire amount of "other sources" were from syndication of stories and photos - which likely overstates the role of these copyright-dependent sources - it would account for little more than 6 percent of total revenues. The net operating revenues for the Gannett Company were more than $5.6 billion in newspaper advertising and circulation revenue, relative to about $380 million in all other revenues. As with the *New York Times*, at most a little more than 6 percent of revenues could be attributed to copyright-dependent activities. For Knight Ridder, the 2003 numbers were $2.8 billion and $100 million, respectively, or a maximum of about 3.5 percent from copyrights. Given these numbers, it is safe to say that daily newspapers are not a copyright-dependent industry, although they are clearly a market-based information production industry.

As it turns out, repeated survey studies since 1981 have shown that in all industrial sectors except for very few - most notably pharmaceuticals - firm managers do not see patents as the most important way they capture the benefits of their research and developments./7 They rank the advantages that strong research and development gives them in lowering the cost or improving the quality of manufacture, being the first in the market, or developing strong marketing relationships as more important than patents. The term "intellectual property" has high cultural visibility today. Hollywood, the recording industry, and pharmaceuticals occupy center stage on the national and international policy agenda for information policy. However, in the overall mix of our information, knowledge, and cultural production system, the total weight of these exclusivity-based market actors is surprisingly small relative to the combination of nonmarket sectors, government and nonprofit, and market-based actors whose business models do not depend on proprietary exclusion from their information outputs.

The upshot of the mainstream economic analysis of information production today is that the widely held intuition that markets are more or less the best way to produce goods, that property rights and contracts are efficient ways of organizing production decisions, and that subsidies distort production decisions, is only very ambiguously applicable to information. While exclusive rights-based production can partially solve the problem of how information will be produced in our society, a comprehensive regulatory system that tries to mimic property in this area - such as both the United States and the European Union have tried to implement internally and through international agreements - simply cannot work perfectly, even in an ideal market posited by the most abstract economics models. Instead, we find the majority of businesses in most sectors reporting that they do not rely on intellectual property as a primary mechanism for appropriating the benefits of their research and development investments. In addition, we find mainstream economists believing that there is a substantial role for government funding; that nonprofit research can be more efficient than for-profit research; and, otherwise, that nonproprietary production can play an important role in our information production system.

The Diversity of Strategies in our Current Information Production System

The actual universe of information production in the economy then, is not as dependent on property rights and markets in information goods as the last quarter century's increasing obsession with "intellectual property" might suggest. Instead, what we see both from empirical work and theoretical work is that individuals and firms in the economy produce information using a wide range of strategies. Some of these strategies indeed rely on exclusive rights like patents or copyrights, and aim at selling information as a good into an information market. Many, however, do not. In order to provide some texture to what these models look like, we can outline a series of ideal-type "business" strategies for producing information. The point here is not to provide an exhaustive map of the empirical business literature. It is, instead, to offer a simple analytic framework within which to understand the mix of strategies available for firms and individuals to appropriate the benefits of their investments - of time, money, or both, in activities that result in the production of information, knowledge, and culture. The differentiating parameters are simple: cost minimization and benefit maximization. Any of these strategies could use inputs that are already owned - such as existing lyrics for a song or a patented invention to improve on - by buying a license from the owner of the exclusive rights for the existing information. Cost minimization here refers purely to ideal-type strategies for obtaining as many of the information inputs as possible at their marginal cost of zero, instead of buying licenses to inputs at a positive market price. It can be pursued by using materials from the public domain, by using materials the producer itself owns, or by sharing/bartering for information inputs owned by others in exchange for one's own information inputs. Benefits can be obtained either in reliance on asserting one's exclusive rights, or by following a non-exclusive strategy, using some other mechanism that improves the position of the information producer because they invested in producing the information. Nonexclusive strategies for benefit maximization can be pursued both by market actors and by nonmarket actors. Table 2.1 maps nine ideal-type strategies characterized by these components.

The ideal-type strategy that underlies patents and copyrights can be thought of as the "Romantic Maximizer." It conceives of the information producer as a single author or inventor laboring creatively - hence romantic - but in expectation of royalties, rather than immortality, beauty, or truth. An individual or small start-up firm that sells software it developed to a larger firm, or an author selling rights to a book or a film typify this model. The second ideal type that arises within exclusive-rights based industries, "Mickey," is a larger firm that already owns an inventory of exclusive rights, some through in-house development, some by buying from Romantic Maximizers. A defining cost-reduction mechanism for Mickey is that it applies creative people to work on its own inventory, for which it need not pay above marginal cost prices in the market. This strategy is the most advantageous in an environment of very strong exclusive rights protection for a number of reasons. First, the ability to extract higher rents from the existing inventory of information goods is greatest for firms that (a) have an inventory and (b) rely on asserting exclusive rights as their mode of extracting value. Second, the increased costs of production associated with strong exclusive rights are cushioned by the ability of such firms to rework their existing inventory, rather than trying to work with materials from an ever-shrinking public domain or paying for every source of inspiration and element of a new composition. The coarsest version of this strategy might be found if Disney were to produce a "winter sports" thirty-minute television program by tying together scenes from existing cartoons, say, one in which Goofy plays hockey followed by a snippet of Donald Duck ice skating, and so on. More subtle, and representative of the type of reuse relevant to the analysis here, would be the case where Disney buys the rights to Winnie-the-Pooh, and, after producing an animated version of stories from the original books, then continues to work with the same characters and relationships to create a new film, say, *Winnie-the-Pooh-Frankenpooh* (or *Beauty and the Beast-Enchanted Christmas*; or *The Little Mermaid-Stormy the Wild Seahorse*).

Table 2.1: Ideal-Type Information Production Strategies

Cost Minimization/ Benefit Acquisition	Public Domain	Intrafirm	Barter/Sharing
Rights-based exclusion (make money by exercising exclusive rights - licensing or blocking competition)	*Romantic Maximizers* (authors, composers; sell to publishers; sometimes sell to Mickeys)	*Mickey* (Disney reuses inventory for derivative works; buy outputs of Romantic Maximizers)	*RCA* (small number of companies hold blocking patents; they create patent pools to build valuable goods)
Nonexclusion-Market (make money from information production but not by exercising the exclusive rights)	*Scholarly Lawyers* (write articles to get clients; other examples include bands that give music out for free as advertisements for touring and charge money for performance; software developers who develop software and make money from customizing it to a particular client, on-site management, advice and training, not from licensing)	*Know-How* (firms that have cheaper or better production processes because of their research, lower their costs or improve the quality of other goods or services; lawyer offices that build on existing forms)	*Learning Networks* (share information with similar organizations - make money from early access to information. For example, newspapers join together to create a wire service; firms where engineers and scientists from different firms attend professional societies to diffuse knowledge)
Nonexclusion-Nonmarket	*Joe Einstein* (give away information for free in return for status, benefits to reputation, value of the innovation to themselves; wide range of motivations. Includes members of amateur choirs who perform for free, academics who write articles for fame, people who write op-eds, contribute to mailing lists; many free software developers and free software generally for most uses)	*Los Alamos* (share in-house information, rely on in-house inputs to produce valuable public goods used to secure additional government funding and status)	*Limited sharing networks* (release paper to small number of colleagues to get comments so you can improve it before publication. Make use of time delay to gain relative advantage later on using Joe Einstein strategy. Share one's information on formal condition of reciprocity: like "copyleft" conditions on derivative works for distribution)

The third exclusive-rights-based strategy, which I call "RCA," is barter among the owners of inventories. Patent pools, cross-licensing, and market-sharing agreements among the radio patents holders in 1920-1921, which I describe in chapter 6, are a perfect example. RCA, GE, AT&T, and Westinghouse held blocking patents that prevented each other and anyone else from manufacturing the best radios possible given technology at that time. The four companies entered an agreement to combine their patents and divide the radio equipment and services markets, which they used throughout the 1920s to exclude competitors and to capture precisely the postinnovation monopoly rents sought to be created by patents.

Exclusive-rights-based business models, however, represent only a fraction of our information production system. There are both market-based and nonmarket models to sustain and organize information production. Together, these account for a substantial portion of our information output. Indeed, industry surveys concerned with patents have shown that the vast majority of industrial R&D is pursued with strategies that do not rely primarily on patents. This does not mean that most or any of the firms that pursue these strategies possess or seek no exclusive rights in their information products. It simply means that their production strategy does not depend on asserting these rights through exclusion. One such cluster of strategies, which I call "Scholarly Lawyers," relies on demand-side effects of access to the information the producer distributes. It relies on the fact that sometimes using an information good that one has produced makes its users seek out a relationship with the author. The author then charges for the relationship, not for the information. Doctors or lawyers who publish in trade journals, become known, and get business as a result are an instance of this strategy. An enormously creative industry, much of which operates on this model, is software. About two-thirds of industry revenues in software development come from activities that the Economic Census describes as: (1) writing, modifying, testing, and supporting software to meet the needs of a particular customer; (2) planning and designing computer systems that integrate computer hardware, software, and communication technologies; (3) on-site management and operation of clients' computer systems and/or data processing facilities; and (4) other professional and technical computer-related advice and services, systems consultants, and computer training. "Software publishing," by contrast, the business model that relies on sales based on copyright, accounts for a little more than one-third of the industry's revenues./8 Interestingly, this is the model of appropriation that more than a decade ago, Esther Dyson and John Perry Barlow heralded as the future of music and musicians. They argued in the early 1990s for more or less free access to copies of recordings distributed online, which would lead to greater attendance at live gigs. Revenue from performances, rather than recording, would pay artists.

The most common models of industrial R&D outside of pharmaceuticals, however, depend on supply-side effects of information production. One central reason to pursue research is its effects on firm-specific advantages, like production know-how, which permit the firm to produce more efficiently than competitors and sell better or cheaper competing products. Daily newspapers collectively fund news agencies, and individually fund reporters, because their ability to find information and report it is a necessary input into their product - timely news. As I have already suggested, they do not need copyright to protect their revenues. Those are protected by the short half-life of dailies. The investments come in order to be able to play in the market for daily newspapers. Similarly, the learning curve and know-how effects in semiconductors are such that early entry into the market for a new chip will give the first mover significant advantages over competitors. Investment is then made to capture that position, and the investment is captured by the quasi-rents available from the first-mover advantage. In some cases, innovation is necessary in order to be able to produce at the state of the art. Firms participate in "Learning Networks" to gain the benefits of being at the state of the art, and sharing their respective improvements. However, they can only participate if they innovate. If they do not innovate, they lack the in-house capacity to

understand the state of the art and play at it. Their investments are then recouped not from asserting their exclusive rights, but from the fact that they sell into one of a set of markets, access into which is protected by the relatively small number of firms with such absorption capacity, or the ability to function at the edge of the state of the art. Firms of this sort might barter their information for access, or simply be part of a small group of organizations with enough knowledge to exploit the information generated and informally shared by all participants in these learning networks. They obtain rents from the concentrated market structure, not from assertion of property rights./9

An excellent example of a business strategy based on nonexclusivity is IBM's. The firm has obtained the largest number of patents every year from 1993 to 2004, amassing in total more than 29,000 patents. IBM has also, however, been one of the firms most aggressively engaged in adapting its business model to the emergence of free software.
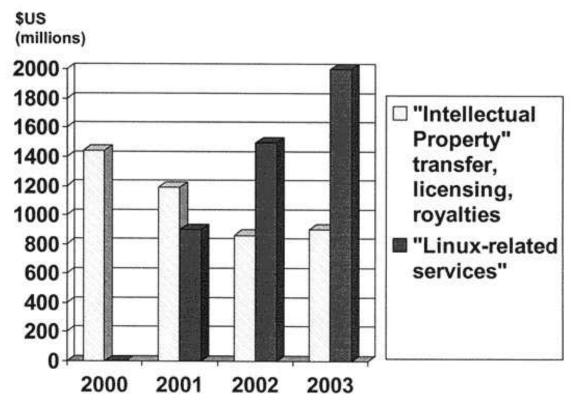
Figure 2.1: Selected IBM Revenues, 2000-2003

Figure 2.1 shows what happened to the relative weight of patent royalties, licenses, and sales in IBM's revenues and revenues that the firm described as coming from "Linux-related services." Within a span of four years, the Linux-related services category moved from accounting for practically no revenues, to providing double the revenues from all patent-related sources, of the firm that has been the most patent-productive in the United States. IBM has described itself as investing more than a billion dollars in free software developers, hired programmers to help develop the Linux kernel and other free software; and donated patents to the Free Software Foundation. What this does for the firm is provide it with a better operating system for its server business - making the servers better, faster, more reliable, and therefore more valuable to consumers. Participating in free software development has also allowed IBM to develop service relationships with its customers, building on free software to offer customer-specific solutions. In other words, IBM has combined both supply-side and demand-side strategies to adopt a nonproprietary business model that has generated more than $2 billion yearly of business for the firm. Its strategy is, if not symbiotic, certainly complementary to free software.

I began this chapter with a puzzle - advanced economies rely on nonmarket organizations for information production much more than they do in other sectors. The puzzle reflects the fact that alongside the diversity of market-oriented business models for information production there is a wide diversity of nonmarket models as well. At a broad level of abstraction, I designate this diversity of motivations and organizational forms as "Joe Einstein" - to underscore the breadth of the range of social practices and practitioners of nonmarket production. These include universities and other research institutes; government research labs that publicize their work, or government information agencies like the Census Bureau. They also include individuals, like academics; authors and artists who play to "immortality" rather than seek to maximize the revenue from their creation. Eric von Hippel has for many years documented user innovation in areas ranging from surfboard design to new mechanisms for pushing electric wiring through insulation tiles./10 The Oratorio Society of New York, whose chorus members are all volunteers, has filled Carnegie Hall every December with a performance of Handel's *Messiah* since the theatre's first season in 1891. Political parties, advocacy groups, and churches are but few of the stable social organizations that fill our information environment with news and views. For symmetry purposes in table 2.1, we also see reliance on internal inventories by some nonmarket organizations, like secret government labs that do not release their information outputs, but use it to continue to obtain public funding. This is what I call "Los Alamos." Sharing in limited networks also occurs in nonmarket relationships, as when academic colleagues circulate a draft to get comments. In the nonmarket, nonproprietary domain, however, these strategies were in the past relatively smaller in scope and significance than the simple act of taking from the public domain and contributing back to it that typifies most Joe Einstein behaviors. Only since the mid-1980s have we begun to see a shift from releasing into the public domain to adoption of commons-binding licensing, like the "copyleft" strategies I describe in chapter 3. What makes these strategies distinct from Joe Einstein is that they formalize the requirement of reciprocity, at least for some set of rights shared.

My point is not to provide an exhaustive list of all the ways we produce information. It is simply to offer some texture to the statement that information, knowledge, and culture are produced in diverse ways in contemporary society. Doing so allows us to understand the comparatively limited role that production based purely on exclusive rights-like patents, copyrights, and similar regulatory constraints on the use and exchange of information - has played in our information production system to this day. It is not new or mysterious to suggest that nonmarket production is important to information production. It is not new or mysterious to suggest that efficiency increases whenever it is possible to produce information in a way that allows the producer - whether market actor or not - to appropriate the benefits of production without actually charging a price for use of the

information itself. Such strategies are legion among both market and nonmarket actors. Recognizing this raises two distinct questions: First, how does the cluster of mechanisms that make up intellectual property law affect this mix? Second, how do we account for the mix of strategies at any given time? Why, for example, did proprietary, market-based production become so salient in music and movies in the twentieth century, and what is it about the digitally networked environment that could change this mix?

The Effects of Exclusive Rights

Once we recognize that there are diverse strategies of appropriation for information production, we come to see a new source of inefficiency caused by strong "intellectual property"-type rights. Recall that in the mainstream analysis, exclusive rights always cause static inefficiency - that is, they allow producers to charge positive prices for products (information) that have a zero marginal cost. Exclusive rights have a more ambiguous effect dynamically. They raise the expected returns from information production, and thereby are thought to induce investment in information production and innovation. However, they also increase the costs of information inputs. If existing innovations are more likely covered by patent, then current producers will more likely have to pay for innovations or uses that in the past would have been available freely from the public domain. Whether, overall, any given regulatory change that increases the scope of exclusive rights improves or undermines new innovation therefore depends on whether, given the level of appropriability that preceded it, it increased input costs more or less than it increased the prospect of being paid for one's outputs.

The diversity of appropriation strategies adds one more kink to this story. Consider the following very simple hypothetical. Imagine an industry that produces "infowidgets." There are ten firms in the business. Two of them are infowidget publishers on the Romantic Maximizer model. They produce infowidgets as finished goods, and sell them based on patent. Six firms produce infowidgets on supply-side (Know-How) or demand-side (Scholarly Lawyer) effects: they make their Realwidgets or Servicewidgets more efficient or desirable to consumers, respectively. Two firms are nonprofit infowidget producers that exist on a fixed, philanthropically endowed income. Each firm produces five infowidgets, for a total market supply of fifty. Now imagine a change in law that increases exclusivity. Assume that this is a change in law that, absent diversity of appropriation, would be considered efficient. Say it increases input costs by 10 percent and appropriability by 20 percent, for a net expected gain of 10 percent. The two infowidget publishers would each see a 10 percent net gain, and let us assume that this would cause each to increase its efforts by 10 percent and produce 10 percent more infowidgets. Looking at these two firms alone, the change in law caused an increase from ten infowidgets to eleven - a gain for the policy change. Looking at the market as a whole, however, eight firms see an increase of 10 percent in costs, and no gain in appropriability. This is because none of these firms actually relies on exclusive rights to appropriate its product's value. If, commensurate with our assumption for the publishers, we assume that this results in a decline in effort and productivity of 10 percent for the eight firms, we would see these firms decline from forty infowidgets to thirty-six, and total market production would decline from fifty infowidgets to forty-seven.

Another kind of effect for the change in law may be to persuade some of the firms to shift strategies or to consolidate. Imagine, for example, that most of the inputs required by the two publishers were owned by the other infowidget publisher. If the two firms merged into one Mickey, each could use the outputs of the other at its marginal cost - zero - instead of at its exclusive-rights market price. The increase in exclusive rights would then not affect the merged firm's costs, only the costs of outside firms that would have to buy the merged firm's outputs from the market.

Given this dynamic, strong exclusive rights drive concentration of inventory owners. We see this very clearly in the increasing sizes of inventory-based firms like Disney. Moreover, the increased appropriability in the exclusive-rights market will likely shift some firms at the margin of the nonproprietary business models to adopt proprietary business models. This, in turn, will increase the amount of information available only from proprietary sources. The feedback effect will further accelerate the rise in information input costs, increasing the gains from shifting to a proprietary strategy and to consolidating larger inventories with new production.

Given diverse strategies, the primary unambiguous effect of increasing the scope and force of exclusive rights is to shape the population of business strategies. Strong exclusive rights increase the attractiveness of exclusive-rights-based strategies at the expense of nonproprietary strategies, whether market-based or nonmarket based. They also increase the value and attraction of consolidation of large inventories of existing information with new production.

When Information Production Meets the Computer Network

Music in the nineteenth century was largely a relational good. It was something people did in the physical presence of each other: in the folk way through hearing, repeating, and improvising; in the middle-class way of buying sheet music and playing for guests or attending public performances; or in the upper-class way of hiring musicians. Capital was widely distributed among musicians in the form of instruments, or geographically dispersed in the hands of performance hall (and drawing room) owners. Market-based production depended on performance through presence. It provided opportunities for artists to live and perform locally, or to reach stardom in cultural centers, but without displacing the local performers. With the introduction of the phonograph, a new, more passive relationship to played music was made possible in reliance on the high-capital requirements of recording, copying, and distributing specific instantiations of recorded music - records. What developed was a concentrated, commercial industry, based on massive financial investments in advertising, or preference formation, aimed at getting ever-larger crowds to want those recordings that the recording executives had chosen. In other words, the music industry took on a more industrial model of production, and many of the local venues - from the living room to the local dance hall - came to be occupied by mechanical recordings rather than amateur and professional local performances. This model crowded out some, but not all, of the live-performance-based markets (for example, jazz clubs, piano bars, or weddings), and created new live-performance markets - the megastar concert tour. The music industry shifted from a reliance on Scholarly Lawyer and Joe Einstein models to reliance on Romantic Maximizer and Mickey models. As computers became more music-capable and digital networks became a ubiquitously available distribution medium, we saw the emergence of the present conflict over the regulation of cultural production - the law of copyright - between the twentieth-century, industrial model recording industry and the emerging amateur distribution systems coupled, at least according to its supporters, to a reemergence of decentralized, relation-based markets for professional performance artists.

This stylized story of the music industry typifies the mass media more generally. Since the introduction of the mechanical press and the telegraph, followed by the phonograph, film, the high-powered radio transmitter, and through to the cable plant or satellite, the capital costs of fixing information and cultural goods in a transmission medium - a high-circulation newspaper, a record or movie, a radio or television program - have been high and increasing. The high physical and financial capital costs involved in making a widely accessible information good and distributing it to the increasingly larger communities (brought together by better transportation systems and more interlinked economic and political systems) muted the relative role of nonmarket production, and emphasized the role of those firms that could muster the financial and physical capital necessary to

communicate on a mass scale. Just as these large, industrial-age machine requirements increased the capital costs involved in information and cultural production, thereby triggering commercialization and concentration of much of this sector, so too ubiquitously available cheap processors have dramatically reduced the capital input costs required to fix information and cultural expressions and communicate them globally. By doing so, they have rendered feasible a radical reorganization of our information and cultural production system, away from heavy reliance on commercial, concentrated business models and toward greater reliance on nonproprietary appropriation strategies, in particular nonmarket strategies whose efficacy was dampened throughout the industrial period by the high capital costs of effective communication.

Information and cultural production have three primary categories of inputs. The first is existing information and culture. We already know that existing information is a nonrival good - that is, its real marginal cost at any given moment is zero. The second major cost is that of the mechanical means of sensing our environment, processing it, and communicating new information goods. This is the high cost that typified the industrial model, and which has drastically declined in computer networks. The third factor is human communicative capacity - the creativity, experience, and cultural awareness necessary to take from the universe of existing information and cultural resources and turn them into new insights, symbols, or representations meaningful to others with whom we converse. Given the zero cost of existing information and the declining cost of communication and processing, human capacity becomes the primary scarce resource in the networked information economy.

Human communicative capacity, however, is an input with radically different characteristics than those of, say, printing presses or satellites. It is held by each individual, and cannot be "transferred" from one person to another or aggregated like so many machines. It is something each of us innately has, though in divergent quanta and qualities. Individual human capacities, rather than the capacity to aggregate financial capital, become the economic core of our information and cultural production. Some of that human capacity is currently, and will continue to be, traded through markets in creative labor. However, its liberation from the constraints of physical capital leaves creative human beings much freer to engage in a wide range of information and cultural production practices than those they could afford to participate in when, in addition to creativity, experience, cultural awareness and time, one needed a few million dollars to engage in information production. From our friendships to our communities we live life and exchange ideas, insights, and expressions in many more diverse relations than those mediated by the market. In the physical economy, these relationships were largely relegated to spaces outside of our economic production system. The promise of the networked information economy is to bring this rich diversity of social life smack into the middle of our economy and our productive lives.

Let's do a little experiment. Imagine that you were performing a Web search with me. Imagine that we were using Google as our search engine, and that what we wanted to do was answer the questions of an inquisitive six-year-old about Viking ships. What would we get, sitting in front of our computers and plugging in a search request for "Viking Ships"? The first site is Canadian, and includes a collection of resources, essays, and worksheets. An enterprising elementary school teacher at the Gander Academy in Newfoundland seems to have put these together. He has essays on different questions, and links to sites hosted by a wide range of individuals and organizations, such as a Swedish museum, individual sites hosted on geocities, and even to a specific picture of a replica Viking ship, hosted on a commercial site dedicated to selling nautical replicas. In other words, it is a Joe Einstein site that points to other sites, which in turn use either Joe Einstein or Scholarly Lawyer strategies. This multiplicity of sources of information that show up on the very first site is then replicated as one continues to explore the remaining links. The second link is to a

Norwegian site called "the Viking Network," a Web ring dedicated to preparing and hosting short essays on Vikings. It includes brief essays, maps, and external links, such as one to an article in *Scientific American*. "To become a member you must produce an Information Sheet on the Vikings in your local area and send it in electronic format to Viking Network. Your info-sheet will then be included in the Viking Network web." The third site is maintained by a Danish commercial photographer, and hosted in Copenhagen, in a portion dedicated to photographs of archeological finds and replicas of Danish Viking ships. A retired professor from the University of Pittsburgh runs the fourth. The fifth is somewhere between a hobby and a showcase for the services of an individual, independent Web publisher offering publishing-related services. The sixth and seventh are museums, in Norway and Virginia, respectively. The eighth is the Web site of a hobbyists' group dedicated to building Viking Ship replicas. The ninth includes classroom materials and teaching guides made freely available on the Internet by PBS, the American Public Broadcasting Service. Certainly, if you perform this search now, as you read this book, the rankings will change from those I saw when I ran it; but I venture that the mix, the range and diversity of producers, and the relative salience of nonmarket producers will not change significantly.

The difference that the digitally networked environment makes is its capacity to increase the efficacy, and therefore the importance, of many more, and more diverse, nonmarket producers falling within the general category of Joe Einstein. It makes nonmarket strategies - from individual hobbyists to formal, well-funded nonprofits - vastly more effective than they could be in the mass-media environment. The economics of this phenomenon are neither mysterious nor complex. Imagine the grade-school teacher who wishes to put together ten to twenty pages of materials on Viking ships for schoolchildren. Pre-Internet, he would need to go to one or more libraries and museums, find books with pictures, maps, and text, or take his own photographs (assuming he was permitted by the museums) and write his own texts, combining this research. He would then need to select portions, clear the copyrights to reprint them, find a printing house that would set his text and pictures in a press, pay to print a number of copies, and then distribute them to all children who wanted them. Clearly, research today is simpler and cheaper. Cutting and pasting pictures and texts that are digital is cheaper. Depending on where the teacher is located, it is possible that these initial steps would have been insurmountable, particularly for a teacher in a poorly endowed community without easy access to books on the subject, where research would have required substantial travel. Even once these barriers were surmounted, in the precomputer, pre-Internet days, turning out materials that looked and felt like a high quality product, with high-resolution pictures and maps, and legible print required access to capital-intensive facilities. The cost of creating even one copy of such a product would likely dissuade the teacher from producing the booklet. At most, he might have produced a mimeographed bibliography, and perhaps some text reproduced on a photocopier. Now, place the teacher with a computer and a high-speed Internet connection, at home or in the school library. The cost of production and distribution of the products of his effort are trivial. A Web site can be maintained for a few dollars a month. The computer itself is widely accessible throughout the developed world. It becomes trivial for a teacher to produce the "booklet" - with more information, available to anyone in the world, anywhere, at any time, as long as he is willing to spend some of his free time putting together the booklet rather than watching television or reading a book.

When you multiply these very simple stylized facts by the roughly billion people who live in societies sufficiently wealthy to allow cheap ubiquitous Internet access, the breadth and depth of the transformation we are undergoing begins to become clear. A billion people in advanced economies may have between two billion and six billion spare hours among them, every day. In order to harness these billions of hours, it would take the whole workforce of almost 340,000 workers employed by the entire motion picture and recording industries in the United States put together,

assuming each worker worked forty-hour weeks without taking a single vacation, for between three and eight and a half years! Beyond the sheer potential quantitative capacity, however one wishes to discount it to account for different levels of talent, knowledge, and motivation, a billion volunteers have qualities that make them more likely to produce what others want to read, see, listen to, or experience. They have diverse interests - as diverse as human culture itself. Some care about Viking ships, others about the integrity of voting machines. Some care about obscure music bands, others share a passion for baking. As Eben Moglen put it, "if you wrap the Internet around every person on the planet and spin the planet, software flows in the network. It's an emergent property of connected human minds that they create things for one another's pleasure and to conquer their uneasy sense of being too alone."/11 It is this combination of a will to create and to communicate with others, and a shared cultural experience that makes it likely that each of us wants to talk about something that we believe others will also want to talk about, that makes the billion potential participants in today's online conversation, and the six billion in tomorrow's conversation, affirmatively better than the commercial industrial model. When the economics of industrial production require high up-front costs and low marginal costs, the producers must focus on creating a few superstars and making sure that everyone tunes in to listen or watch them. This requires that they focus on averaging out what consumers are most likely to buy. This works reasonably well as long as there is no better substitute. As long as it is expensive to produce music or the evening news, there are indeed few competitors for top billing, and the star system can function. Once every person on the planet, or even only every person living in a wealthy economy and 10-20 percent of those living in poorer countries, can easily talk to their friends and compatriots, the competition becomes tougher. It does not mean that there is no continued role for the mass-produced and mass-marketed cultural products - be they Britney Spears or the broadcast news. It does, however, mean that many more "niche markets" - if markets, rather than conversations, are what they should be called - begin to play an ever-increasing role in the total mix of our cultural production system. The economics of production in a digital environment should lead us to expect an increase in the relative salience of nonmarket production models in the overall mix of our information production system, and it is efficient for this to happen - more information will be produced, and much of it will be available for its users at its marginal cost.

The known quirky characteristics of information and knowledge as production goods have always given nonmarket production a much greater role in this production system than was common in capitalist economies for tangible goods. The dramatic decline in the cost of the material means of producing and exchanging information, knowledge, and culture has substantially decreased the costs of information expression and exchange, and thereby increased the relative efficacy of nonmarket production. When these facts are layered over the fact that information, knowledge, and culture have become the central high-value-added economic activities of the most advanced economies, we find ourselves in a new and unfamiliar social and economic condition. Social behavior that traditionally was relegated to the peripheries of the economy has become central to the most advanced economies. Nonmarket behavior is becoming central to producing our information and cultural environment. Sources of knowledge and cultural edification, through which we come to know and comprehend the world, to form our opinions about it, and to express ourselves in communication with others about what we see and believe have shifted from heavy reliance on commercial, concentrated media, to being produced on a much more widely distributed model, by many actors who are not driven by the imperatives of advertising or the sale of entertainment goods.

Strong Exclusive Rights in the Digital Environment

We now have the basic elements of a clash between incumbent institutions and emerging social practice. Technologies of information and cultural production initially led to the increasing salience of commercial, industrial-model production in these areas. Over the course of the twentieth century, in some of the most culturally visible industries like movies and music, copyright law coevolved with the industrial model. By the end of the twentieth century, copyright was longer, broader, and vastly more encompassing than it had been at the beginning of that century. Other exclusive rights in information, culture, and the fruits of innovation expanded following a similar logic. Strong, broad, exclusive rights like these have predictable effects. They preferentially improve the returns to business models that rely on exclusive rights, like copyrights and patents, at the expense of information and cultural production outside the market or in market relationships that do not depend on exclusive appropriation. They make it more lucrative to consolidate inventories of existing materials. The businesses that developed around the material capital required for production fed back into the political system, which responded by serially optimizing the institutional ecology to fit the needs of the industrial information economy firms at the expense of other information producers.

The networked information economy has upset the apple cart on the technical, material cost side of information production and exchange. The institutional ecology, the political framework (the lobbyists, the habits of legislatures), and the legal culture (the beliefs of judges, the practices of lawyers) have not changed. They are as they developed over the course of the twentieth century - centered on optimizing the conditions of those commercial firms that thrive in the presence of strong exclusive rights in information and culture. The outcome of the conflict between the industrial information economy and its emerging networked alternative will determine whether we evolve into a permission culture, as Lessig warns and projects, or into a society marked by social practice of nonmarket production and cooperative sharing of information, knowledge, and culture of the type I describe throughout this book, and which I argue will improve freedom and justice in liberal societies. Chapter 11 chronicles many of the arenas in which this basic conflict is played out. However, for the remainder of this part and part II, the basic economic understanding I offer here is all that is necessary.

There are diverse motivations and strategies for organizing information production. Their relative attractiveness is to some extent dependent on technology, to some extent on institutional arrangements. The rise that we see today in the efficacy and scope of nonmarket production, and of the peer production that I describe and analyze in the following two chapters, are well within the predictable, given our understanding of the economics of information production. The social practices of information production that form the basis of much of the normative analysis I offer in part II are internally sustainable given the material conditions of information production and exchange in the digitally networked environment. These patterns are unfamiliar to us. They grate on our intuitions about how production happens. They grate on the institutional arrangements we developed over the course of the twentieth century to regulate information and cultural production. But that is because they arise from a quite basically different set of material conditions. We must understand these new modes of production. We must learn to evaluate them and compare their advantages and disadvantages to those of the industrial information producers. And then we must adjust our institutional environment to make way for the new social practices made possible by the networked environment.

Notes

1. The full statement was: "[A]ny information obtained, say a new method of production, should, from the welfare point of view, be available free of charge (apart from the costs of transmitting information). This insures optimal utilization of the information but of course provides no incentive for investment in research. In a free enterprise economy, inventive activity is supported by using the invention to create property rights; precisely to the extent that it is successful, there is an underutilization of information." Kenneth Arrow, "Economic Welfare and the Allocation of Resources for Invention," in *Rate and Direction of Inventive Activity: Economic and Social Factors*, ed. Richard R. Nelson (Princeton, NJ: Princeton University Press, 1962), 616-617.

2. Suzanne Scotchmer, "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives* 5 (1991): 29-41.

3. *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

4. Adam Jaffe, "The U.S. Patent System in Transition: Policy Innovation and the Innovation Process," *Research Policy* 29 (2000): 531.

5. Josh Lerner, "Patent Protection and Innovation Over 150 Years" (working paper no. 8977, National Bureau of Economic Research, Cambridge, MA, 2002).

6. At most, a "hot news" exception on the model of *International News Service v. Associated Press*, 248 U.S. 215 (1918), might be required. Even that, however, would only be applicable to online editions that are for pay. In paper, habits of reading, accreditation of the original paper, and first-to-market advantages of even a few hours would be enough. Online, where the first-to-market advantage could shrink to seconds, "hot news" protection may be worthwhile. However, almost all papers are available for free and rely solely on advertising. The benefits of reading a copied version are, at that point, practically insignificant to the reader.

7. Wesley Cohen, R. Nelson, and J. Walsh, "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)" (working paper no. 7552, National Bureau Economic Research, Cambridge, MA, 2000); Richard Levin et al., "Appropriating the Returns from Industrial Research and Development" *Brookings Papers on Economic Activity* 3 (1987): 783; Mansfield et al., "Imitation Costs and Patents: An Empirical Study," *The Economic Journal* 91 (1981): 907.

8. In the 2002 Economic Census, compare NAICS categories 5415 (computer systems and related services) to NAICS 5112 (software publishing). Between the 1997 Economic Census and the 2002 census, this ratio remained stable, at about 36 percent in 1997 and 37 percent in 2002. See 2002 Economic Census, "Industry Series, Information, Software Publishers, and Computer Systems, Design and Related Services" (Washington, DC: U.S. Census Bureau, 2004).

9. Levin et al., "Appropriating the Returns," 794-796 (secrecy, lead time, and learning-curve advantages regarded as more effective than patents by most firms). See also F. M. Scherer, "Learning by Doing and International Trade in Semiconductors" (faculty research working paper series R94-13, John F. Kennedy School of Government, Harvard University, Cambridge, MA, 1994), an empirical study of semiconductor industry suggesting that for industries with steep learning curves, investment in information production is driven by advantages of being first down

the learning curve rather than the expectation of legal rights of exclusion. The absorption effect is described in Wesley M. Cohen and Daniel A. Leventhal, "Innovation and Learning: The Two Faces of R&D," *The Economic Journal* 99 (1989): 569-596. The collaboration effect was initially described in Richard R. Nelson, "The Simple Economics of Basic Scientific Research,"*Journal of Political Economy* 67 (June 1959): 297-306. The most extensive work over the past fifteen years, and the source of the term of learning networks, has been from Woody Powell on knowledge and learning networks. Identifying the role of markets made concentrated by the limited ability to use information, rather than through exclusive rights, was made in F. M. Scherer, "Nordhaus's Theory of Optimal Patent Life: A Geometric Reinterpretation," *American Economic Review* 62 (1972): 422-427.

10. Eric von Hippel, *Democratizing Innovation* (Cambridge, MA: MIT Press, 2005).

11. Eben Moglen, "Anarchism Triumphant: Free Software and the Death of Copyright," *First Monday* (1999), http://www.firstmonday.dk/issues/issue4_8/moglen/

Chapter 3
Peer Production and Sharing

At the heart of the economic engine of the world's most advanced economies, we are beginning to notice a persistent and quite amazing phenomenon. A new model of production has taken root; one that should not be there, at least according to our most widely held beliefs about economic behavior. It should not, the intuitions of the late-twentieth-century American would say, be the case that thousands of volunteers will come together to collaborate on a complex economic project. It certainly should not be that these volunteers will beat the largest and best-financed business enterprises in the world at their own game. And yet, this is precisely what is happening in the software world.

Industrial organization literature provides a prominent place for the transaction costs view of markets and firms, based on insights of Ronald Coase and Oliver Williamson. On this view, people use markets when the gains from doing so, net of transaction costs, exceed the gains from doing the same thing in a managed firm, net of the costs of organizing and managing a firm. Firms emerge when the opposite is true, and transaction costs can best be reduced by bringing an activity into a managed context that requires no individual transactions to allocate this resource or that effort. The emergence of free and open-source software, and the phenomenal success of its flagships, the GNU/Linux operating system, the Apache Web server, Perl, and many others, should cause us to take a second look at this dominant paradigm./1 Free software projects do not rely on markets or on managerial hierarchies to organize production. Programmers do not generally participate in a project because someone who is their boss told them to, though some do. They do not generally participate in a project because someone offers them a price to do so, though some participants do focus on long-term appropriation through money-oriented activities, like consulting or service contracts. However, the critical mass of participation in projects cannot be explained by the direct presence of a price or even a future monetary return. This is particularly true of the all-important, microlevel decisions: who will work, with what software, on what project. In other words, programmers participate in free software projects without following the signals generated by market-based, firm-based, or hybrid models. In chapter 2 I focused on how the networked information economy departs from the industrial information economy by improving the efficacy of nonmarket production generally. Free software offers a glimpse at a more basic and radical challenge. It suggests that the networked environment makes possible a new modality of organizing production: radically decentralized, collaborative, and nonproprietary; based on sharing resources and outputs among widely distributed, loosely connected individuals who cooperate with each other without relying on either market signals or managerial commands. This is what I call "commons-based peer production."

"Commons" refers to a particular institutional form of structuring the rights to access, use, and control resources. It is the opposite of "property" in the following sense: With property, law determines one particular person who has the authority to decide how the resource will be used. That person may sell it, or give it away, more or less as he or she pleases. "More or less" because property doesn't mean anything goes. We cannot, for example, decide that we will give our property away to one branch of our family, as long as that branch has boys, and then if that branch has no boys, decree that the property will revert to some other branch of the family. That type of provision, once common in English property law, is now legally void for public policy reasons. There are many other things we cannot do with our property-like build on wetlands. However, the core characteristic of property as the institutional foundation of markets is that the allocation of

power to decide how a resource will be used is systematically and drastically asymmetric. That asymmetry permits the existence of "an owner" who can decide what to do, and with whom. We know that transactions must be made - rent, purchase, and so forth - if we want the resource to be put to some other use. The salient characteristic of commons, as opposed to property, is that no single person has exclusive control over the use and disposition of any particular resource in the commons. Instead, resources governed by commons may be used or disposed of by anyone among some (more or less well-defined) number of persons, under rules that may range from "anything goes" to quite crisply articulated formal rules that are effectively enforced.

Commons can be divided into four types based on two parameters. The first parameter is whether they are open to anyone or only to a defined group. The oceans, the air, and highway systems are clear examples of open commons. Various traditional pasture arrangements in Swiss villages or irrigation regions in Spain are now classic examples, described by Eleanor Ostrom, of limited-access common resources - where access is limited only to members of the village or association that collectively "owns" some defined pasturelands or irrigation system./2 As Carol Rose noted, these are better thought of as limited common property regimes, rather than commons, because they behave as property vis-à-vis the entire world except members of the group who together hold them in common. The second parameter is whether a commons system is regulated or unregulated. Practically all well-studied, limited common property regimes are regulated by more or less elaborate rules - some formal, some social-conventional - governing the use of the resources. Open commons, on the other hand, vary widely. Some commons, called open access, are governed by no rule. Anyone can use resources within these types of commons at will and without payment. Air is such a resource, with respect to air intake (breathing, feeding a turbine). However, air is a regulated commons with regard to outtake. For individual human beings, breathing out is mildly regulated by social convention - you do not breath too heavily on another human being's face unless forced to. Air is a more extensively regulated commons for industrial exhalation - in the shape of pollution controls. The most successful and obvious regulated commons in contemporary landscapes are the sidewalks, streets, roads, and highways that cover our land and regulate the material foundation of our ability to move from one place to the other. In all these cases, however, the characteristic of commons is that the constraints, if any, are symmetric among all users, and cannot be unilaterally controlled by any single individual. The term "commons-based" is intended to underscore that what is characteristic of the cooperative enterprises I describe in this chapter is that they are not built around the asymmetric exclusion typical of property. Rather, the inputs and outputs of the process are shared, freely or conditionally, in an institutional form that leaves them equally available for all to use as they choose at their individual discretion. This latter characteristic - that commons leave individuals free to make their own choices with regard to resources managed as a commons - is at the foundation of the freedom they make possible. This is a freedom I return to in the discussion of autonomy. Not all commons-based production efforts qualify as peer production. Any production strategy that manages its inputs and outputs as commons locates that production modality outside the proprietary system, in a framework of social relations. It is the freedom to interact with resources and projects without seeking anyone's permission that marks commons-based production generally, and it is also that freedom that underlies the particular efficiencies of peer production, which I explore in chapter 4.

The term "peer production" characterizes a subset of commons-based production practices. It refers to production systems that depend on individual action that is self-selected and decentralized, rather than hierarchically assigned. "Centralization" is a particular response to the problem of how to make the behavior of many individual agents cohere into an effective pattern or achieve an effective result. Its primary attribute is the separation of the locus of opportunities for action from the authority to choose the action that the agent will undertake. Government authorities, firm

managers, teachers in a classroom, all occupy a context in which potentially many individual wills could lead to action, and reduce the number of people whose will is permitted to affect the actual behavior patterns that the agents will adopt. "Decentralization" describes conditions under which the actions of many agents cohere and are effective despite the fact that they do not rely on reducing the number of people whose will counts to direct effective action. A substantial literature in the past twenty years, typified, for example, by Charles Sabel's work, has focused on the ways in which firms have tried to overcome the rigidities of managerial pyramids by decentralizing learning, planning, and execution of the firm's functions in the hands of employees or teams. The most pervasive mode of "decentralization," however, is the ideal market. Each individual agent acts according to his or her will. Coherence and efficacy emerge because individuals signal their wishes, and plan their behavior not in cooperation with others, but by coordinating, understanding the will of others and expressing their own through the price system.

What we are seeing now is the emergence of more effective collective action practices that are decentralized but do not rely on either the price system or a managerial structure for coordination. In this, they complement the increasing salience of uncoordinated nonmarket behavior that we saw in chapter 2. The networked environment not only provides a more effective platform for action to nonprofit organizations that organize action like firms or to hobbyists who merely coexist coordinately. It also provides a platform for new mechanisms for widely dispersed agents to adopt radically decentralized cooperation strategies other than by using proprietary and contractual claims to elicit prices or impose managerial commands. This kind of information production by agents operating on a decentralized, nonproprietary model is not completely new. Science is built by many people contributing incrementally - not operating on market signals, not being handed their research marching orders by a boss - independently deciding what to research, bringing their collaboration together, and creating science. What we see in the networked information economy is a dramatic increase in the importance and the centrality of information produced in this way.

Free/Open-Source Software

The quintessential instance of commons-based peer production has been free software. Free software, or open source, is an approach to software development that is based on shared effort on a nonproprietary model. It depends on many individuals contributing to a common project, with a variety of motivations, and sharing their respective contributions without any single person or entity asserting rights to exclude either from the contributed components or from the resulting whole. In order to avoid having the joint product appropriated by any single party, participants usually retain copyrights in their contribution, but license them to anyone - participant or stranger - on a model that combines a universal license to use the materials with licensing constraints that make it difficult, if not impossible, for any single contributor or third party to appropriate the project. This model of licensing is the most important institutional innovation of the free software movement. Its central instance is the GNU General Public License, or GPL. This requires anyone who modifies software and distributes the modified version to license it under the same free terms as the original software. While there have been many arguments about how widely the provisions that prevent downstream appropriation should be used, the practical adoption patterns have been dominated by forms of licensing that prevent anyone from exclusively appropriating the contributions or the joint product. More than 85 percent of active free software projects include some version of the GPL or similarly structured license./3

Free software has played a critical role in the recognition of peer production, because software is a functional good with measurable qualities. It can be more or less authoritatively tested against its market-based competitors. And, in many instances, free software has prevailed. About 70 percent

of Web server software, in particular for critical e-commerce sites, runs on the Apache Web server-free software.[/4] More than half of all back-office e-mail functions are run by one free software program or another. Google, Amazon, and CNN.com, for example, run their Web servers on the GNU/Linux operating system. They do this, presumably, because they believe this peer-produced operating system is more reliable than the alternatives, not because the system is "free." It would be absurd to risk a higher rate of failure in their core business activities in order to save a few hundred thousand dollars on licensing fees. Companies like IBM and Hewlett Packard, consumer electronics manufacturers, as well as military and other mission-critical government agencies around the world have begun to adopt business and service strategies that rely and extend free software. They do this because it allows them to build better equipment, sell better services, or better fulfill their public role, even though they do not control the software development process and cannot claim proprietary rights of exclusion in the products of their contributions.

The story of free software begins in 1984, when Richard Stallman started working on a project of building a nonproprietary operating system he called GNU (GNU's Not Unix). Stallman, then at the Massachusetts Institute of Technology (MIT), operated from political conviction. He wanted a world in which software enabled people to use information freely, where no one would have to ask permission to change the software they use to fit their needs or to share it with a friend for whom it would be helpful. These freedoms to share and to make your own software were fundamentally incompatible with a model of production that relies on property rights and markets, he thought, because in order for there to be a market in uses of software, owners must be able to make the software unavailable to people who need it. These people would then pay the provider in exchange for access to the software or modification they need. If anyone can make software or share software they possess with friends, it becomes very difficult to write software on a business model that relies on excluding people from software they need unless they pay. As a practical matter, Stallman started writing software himself, and wrote a good bit of it. More fundamentally, he adopted a legal technique that started a snowball rolling. He could not write a whole operating system by himself. Instead, he released pieces of his code under a license that allowed anyone to copy, distribute, and modify the software in whatever way they pleased. He required only that, if the person who modified the software then distributed it to others, he or she do so under the exact same conditions that he had distributed his software. In this way, he invited all other programmers to collaborate with him on this development program, if they wanted to, on the condition that they be as generous with making their contributions available to others as he had been with his. Because he retained the copyright to the software he distributed, he could write this condition into the license that he attached to the software. This meant that anyone using or distributing the software as is, without modifying it, would not violate Stallman's license. They could also modify the software for their own use, and this would not violate the license. However, if they chose to distribute the modified software, they would violate Stallman's copyright unless they included a license identical to his with the software they distributed. This license became the GNU General Public License, or GPL. The legal jujitsu Stallman used - asserting his own copyright claims, but only to force all downstream users who wanted to rely on his contributions to make their own contributions available to everyone else - came to be known as "copyleft," an ironic twist on copyright. This legal artifice allowed anyone to contribute to the GNU project without worrying that one day they would wake up and find that someone had locked them out of the system they had helped to build.

The next major step came when a person with a more practical, rather than prophetic, approach to his work began developing one central component of the operating system - the kernel. Linus Torvalds began to share the early implementations of his kernel, called Linux, with others, under the GPL. These others then modified, added, contributed, and shared among themselves these pieces of the operating system. Building on top of Stallman's foundation, Torvalds crystallized a

model of production that was fundamentally different from those that preceded it. His model was based on voluntary contributions and ubiquitous, recursive sharing; on small incremental improvements to a project by widely dispersed people, some of whom contributed a lot, others a little. Based on our usual assumptions about volunteer projects and decentralized production processes that have no managers, this was a model that could not succeed. But it did.

It took almost a decade for the mainstream technology industry to recognize the value of free or open-source software development and its collaborative production methodology. As the process expanded and came to encompass more participants, and produce more of the basic tools of Internet connectivity - Web server, e-mail server, scripting - more of those who participated sought to "normalize" it, or, more specifically, to render it apolitical. Free software is about freedom ("free as in free speech, not free beer" is Stallman's epitaph for it). "Open-source software" was chosen as a term that would not carry the political connotations. It was simply a mode of organizing software production that may be more effective than market-based production. This move to depoliticize peer production of software led to something of a schism between the free software movement and the communities of open source software developers. It is important to understand, however, that from the perspective of society at large and the historical trajectory of information production generally the abandonment of political motivation and the importation of free software into the mainstream have not made it less politically interesting, but more so. Open source and its wide adoption in the business and bureaucratic mainstream allowed free software to emerge from the fringes of the software world and move to the center of the public debate about practical alternatives to the current way of doing things.

So what is open-source software development? The best source for a phenomenology of open-source development continues to be Eric Raymond's *Cathedral and Bazaar*, written in 1998. Imagine that one person, or a small group of friends, wants a utility. It could be a text editor, photo-retouching software, or an operating system. The person or small group starts by developing a part of this project, up to a point where the whole utility - if it is simple enough - or some important part of it, is functional, though it might have much room for improvement. At this point, the person makes the program freely available to others, with its source code - instructions in a human-readable language that explain how the software does whatever it does when compiled into a machine-readable language. When others begin to use it, they may find bugs, or related utilities that they want to add (e.g., the photo-retouching software only increases size and sharpness, and one of its users wants it to allow changing colors as well). The person who has found the bug or is interested in how to add functions to the software may or may not be the best person in the world to actually write the software fix. Nevertheless, he reports the bug or the new need in an Internet forum of users of the software. That person, or someone else, then thinks that they have a way of tweaking the software to fix the bug or add the new utility. They then do so, just as the first person did, and release a new version of the software with the fix or the added utility. The result is a collaboration between three people - the first author, who wrote the initial software; the second person, who identified a problem or shortcoming; and the third person, who fixed it. This collaboration is not managed by anyone who organizes the three, but is instead the outcome of them all reading the same Internet-based forum and using the same software, which is released under an open, rather than proprietary, license. This enables some of its users to identify problems and others to fix these problems without asking anyone's permission and without engaging in any transactions.

The most surprising thing that the open source movement has shown, in real life, is that this simple model can operate on very different scales, from the small, three-person model I described for simple projects, up to the many thousands of people involved in writing the Linux kernel and the

GNU/Linux operating system - an immensely difficult production task. SourceForge, the most popular hosting-meeting place of such projects, has close to 100,000 registered projects, and nearly a million registered users. The economics of this phenomenon are complex. In the larger-scale models, actual organization form is more diverse than the simple, three-person model. In particular, in some of the larger projects, most prominently the Linux kernel development process, a certain kind of meritocratic hierarchy is clearly present. However, it is a hierarchy that is very different in style, practical implementation, and organizational role than that of the manager in the firm. I explain this in chapter 4, as part of the analysis of the organizational forms of peer production. For now, all we need is a broad outline of how peer-production projects look, as we turn to observe case studies of kindred production models in areas outside of software.

Peer Production of Information, Knowledge, and Culture Generally

Free software is, without a doubt, the most visible instance of peer production at the turn of the twenty-first century. It is by no means, however, the only instance. Ubiquitous computer communications networks are bringing about a dramatic change in the scope, scale, and efficacy of peer production throughout the information and cultural production system. As computers become cheaper and as network connections become faster, cheaper, and ubiquitous, we are seeing the phenomenon of peer production of information scale to much larger sizes, performing more complex tasks than were possible in the past for nonprofessional production. To make this phenomenon more tangible, I describe a number of such enterprises, organized to demonstrate the feasibility of this approach throughout the information production and exchange chain. While it is possible to break an act of communication into finer-grained subcomponents, largely we see three distinct functions involved in the process. First, there is an initial utterance of a humanly meaningful statement. Writing an article or drawing a picture, whether done by a professional or an amateur, whether high quality or low, is such an action. Second, there is a separate function of mapping the initial utterances on a knowledge map. In particular, an utterance must be understood as "relevant" in some sense, and "credible." Relevance is a subjective question of mapping an utterance on the conceptual map of a given user seeking information for a particular purpose defined by that individual. Credibility is a question of quality by some objective measure that the individual adopts as appropriate for purposes of evaluating a given utterance. The distinction between the two is somewhat artificial, however, because very often the utility of a piece of information will depend on a combined valuation of its credibility and relevance. I therefore refer to "relevance/accreditation" as a single function for purposes of this discussion, keeping in mind that the two are complementary and not entirely separable functions that an individual requires as part of being able to use utterances that others have uttered in putting together the user's understanding of the world. Finally, there is the function of distribution, or how one takes an utterance produced by one person and distributes it to other people who find it credible and relevant. In the mass-media world, these functions were often, though by no means always, integrated. NBC news produced the utterances, gave them credibility by clearing them on the evening news, and distributed them simultaneously. What the Internet is permitting is much greater disaggregation of these functions.

Uttering Content

NASA Clickworkers was "an experiment to see if public volunteers, each working for a few minutes here and there can do some routine science analysis that would normally be done by a scientist or graduate student working for months on end." Users could mark craters on maps of Mars, classify craters that have already been marked, or search the Mars landscape for "honeycomb" terrain. The project was "a pilot study with limited funding, run part-time by one software engineer, with occasional input from two scientists." In its first six months of operation, more than 85,000 users

visited the site, with many contributing to the effort, making more than 1.9 million entries (including redundant entries of the same craters, used to average out errors). An analysis of the quality of markings showed "that the automatically-computed consensus of a large number of clickworkers is virtually indistinguishable from the inputs of a geologist with years of experience in identifying Mars craters."[/5] The tasks performed by clickworkers (like marking craters) were discrete, each easily performed in a matter of minutes. As a result, users could choose to work for a few minutes doing a single iteration or for hours by doing many. An early study of the project suggested that some clickworkers indeed worked on the project for weeks, but that 37 percent of the work was done by one-time contributors.[/6]

The clickworkers project was a particularly clear example of how a complex professional task that requires a number of highly trained individuals on full-time salaries can be reorganized so as to be performed by tens of thousands of volunteers in increments so minute that the tasks could be performed on a much lower budget. The low budget would be devoted to coordinating the volunteer effort. However, the raw human capital needed would be contributed for the fun of it. The professionalism of the original scientists was replaced by a combination of high modularization of the task. The organizers broke a large, complex task into small, independent modules. They built in redundancy and automated averaging out of both errors and purposeful erroneous markings - like those of an errant art student who thought it amusing to mark concentric circles on the map. What the NASA scientists running this experiment had tapped into was a vast pool of five-minute increments of human judgment, applied with motivation to participate in a task unrelated to "making a living."

While clickworkers was a distinct, self-conscious experiment, it suggests characteristics of distributed production that are, in fact, quite widely observable. We have already seen in chapter 2, in our little search for Viking ships, how the Internet can produce encyclopedic or almanac-type information. The power of the Web to answer such an encyclopedic question comes not from the fact that one particular site has all the great answers. It is not an *Encyclopedia Britannica*. The power comes from the fact that it allows a user looking for specific information at a given time to collect answers from a sufficiently large number of contributions. The task of sifting and accrediting falls to the user, motivated by the need to find an answer to the question posed. As long as there are tools to lower the cost of that task to a level acceptable to the user, the Web shall have "produced" the information content the user was looking for. These are not trivial considerations, but they are also not intractable. As we shall see, some of the solutions can themselves be peer produced, and some solutions are emerging as a function of the speed of computation and communication, which enables more efficient technological solutions.

Encyclopedic and almanac-type information emerges on the Web out of the coordinate but entirely independent action of millions of users. This type of information also provides the focus on one of the most successful collaborative enterprises that has developed in the first five years of the twenty-first century, *Wikipedia*. *Wikipedia* was founded by an Internet entrepreneur, Jimmy Wales. Wales had earlier tried to organize an encyclopedia named Nupedia, which was built on a traditional production model, but whose outputs were to be released freely: its contributors were to be PhDs, using a formal, peer-reviewed process. That project appears to have failed to generate a sufficient number of high-quality contributions, but its outputs were used in *Wikipedia* as the seeds for a radically new form of encyclopedia writing. Founded in January 2001, *Wikipedia* combines three core characteristics: First, it uses a collaborative authorship tool, Wiki. This platform enables anyone, including anonymous passersby, to edit almost any page in the entire project. It stores all versions, makes changes easily visible, and enables anyone to revert a document to any prior version as well as to add changes, small and large. All contributions and changes are rendered transparent by

the software and database. Second, it is a self-conscious effort at creating an encyclopedia - governed first and foremost by a collective informal undertaking to strive for a neutral point of view, within the limits of substantial self-awareness as to the difficulties of such an enterprise. An effort to represent sympathetically all views on a subject, rather than to achieve objectivity, is the core operative characteristic of this effort. Third, all the content generated by this collaboration is released under the GNU Free Documentation License, an adaptation of the GNU GPL to texts.

The shift in strategy toward an open, peer-produced model proved enormously successful. The site saw tremendous growth both in the number of contributors, including the number of active and very active contributors, and in the number of articles included in the encyclopedia (table 3.1). Most of the early growth was in English, but more recently there has been an increase in the number of articles in many other languages: most notably in German (more than 200,000 articles), Japanese (more than 120,000 articles), and French (about 100,000), but also in another five languages that have between 40,000 and 70,000 articles each, another eleven languages with 10,000 to 40,000 articles each, and thirty-five languages with between 1,000 and 10,000 articles each.

Table 3.1: Contributors to *Wikipedia*, January 2001-June 2005

	Jan. 2001	Jan. 2002	Jan. 2003	Jan. 2004	July. 2004	June 2005
Contributors*	10	472	2,188	9,653	25,011	48,721
Active contributors**	9	212	846	3,228	8,442	16,945
Very active contributors***	0	31	190	692	1,637	3,016
No. of English language articles	25	16,000	101,000	190,000	320,000	630,000
No. of articles, all languages	25	19,000	138,000	409,000	862,000	1,600,000

* Contributed at least ten times; ** at least 5 times in last month; *** more than 100 times in last month.

The first systematic study of the quality of *Wikipedia* articles was published as this book was going to press. The journal *Nature* compared 42 science articles from *Wikipedia* to the gold standard of the *Encyclopedia Britannica*, and concluded that "the difference in accuracy was not particularly great."[/7] On November 15, 2004, Robert McHenry, a former editor in chief of the *Encyclopedia Britannica*, published an article criticizing *Wikipedia* as "The Faith-Based Encyclopedia."[/8] As an example, McHenry mocked the *Wikipedia* article on Alexander Hamilton. He noted that Hamilton biographers have a problem fixing his birth year - whether it is 1755 or 1757. *Wikipedia* glossed over this error, fixing the date at 1755. McHenry then went on to criticize the way the dates were treated throughout the article, using it as an anchor to his general claim: *Wikipedia* is unreliable because it is not professionally produced. What McHenry did not note was that the other major online encyclopedias - like *Columbia* or *Encarta* - similarly failed to deal with the ambiguity surrounding Hamilton's birth date. Only the *Britannica* did. However, McHenry's critique triggered the *Wikipedia* distributed correction mechanism. Within hours of the publication of McHenry's Web article, the

reference was corrected. The following few days saw intensive cleanup efforts to conform all references in the biography to the newly corrected version. Within a week or so, *Wikipedia* had a correct, reasonably clean version. It now stood alone with the *Encyclopedia Britannica* as a source of accurate basic encyclopedic information. In coming to curse it, McHenry found himself blessing *Wikipedia*. He had demonstrated precisely the correction mechanism that makes *Wikipedia*, in the long term, a robust model of reasonably reliable information.

Perhaps the most interesting characteristic about *Wikipedia* is the self-conscious social-norms-based dedication to objective writing. Unlike some of the other projects that I describe in this chapter, *Wikipedia* does not include elaborate software-controlled access and editing capabilities. It is generally open for anyone to edit the materials, delete another's change, debate the desirable contents, survey archives for prior changes, and so forth. It depends on self-conscious use of open discourse, usually aimed at consensus. While there is the possibility that a user will call for a vote of the participants on any given definition, such calls can, and usually are, ignored by the community unless a sufficiently large number of users have decided that debate has been exhausted. While the system operators and server host - Wales - have the practical power to block users who are systematically disruptive, this power seems to be used rarely. The project relies instead on social norms to secure the dedication of project participants to objective writing. So, while not entirely anarchic, the project is nonetheless substantially more social, human, and intensively discourse- and trust-based than the other major projects described here. The following fragments from an early version of the self-described essential characteristics and basic policies of *Wikipedia* are illustrative:

First and foremost, the Wikipedia project is self-consciously an encyclopedia - rather than a dictionary, discussion forum, web portal, etc. Wikipedia's participants commonly follow, and enforce, a few basic policies that seem essential to keeping the project running smoothly and productively. First, because we have a huge variety of participants of all ideologies, and from around the world, Wikipedia is committed to making its articles as unbiased as possible. The aim is not to write articles from a single *objective* point of view - this is a common misunderstanding of the policy - but rather, to fairly and sympathetically present all views on an issue. See "neutral point of view" page for further explanation.[/9]

The point to see from this quotation is that the participants of *Wikipedia* are plainly people who like to write. Some of them participate in other collaborative authorship projects. However, when they enter the common project of *Wikipedia*, they undertake to participate in a particular way - a way that the group has adopted to make its product be an encyclopedia. On their interpretation, that means conveying in brief terms the state of the art on the item, including divergent opinions about it, but not the author's opinion. Whether that is an attainable goal is a subject of interpretive theory, and is a question as applicable to a professional encyclopedia as it is to *Wikipedia*. As the project has grown, it has developed more elaborate spaces for discussing governance and for conflict resolution. It has developed structures for mediation, and if that fails, arbitration, of disputes about particular articles.

The important point is that *Wikipedia* requires not only mechanical cooperation among people, but a commitment to a particular style of writing and describing concepts that is far from intuitive or natural to people. It requires self-discipline. It enforces the behavior it requires primarily through appeal to the common enterprise that the participants are engaged in, coupled with a thoroughly transparent platform that faithfully records and renders all individual interventions in the common project and facilitates discourse among participants about how their contributions do, or do not, contribute to this common enterprise. This combination of an explicit statement of common purpose, transparency, and the ability of participants to identify each other's actions and counteract

them - that is, edit out "bad" or "faithless" definitions - seems to have succeeded in keeping this community from devolving into inefficacy or worse. A case study by IBM showed, for example, that while there were many instances of vandalism on *Wikipedia*, including deletion of entire versions of articles on controversial topics like "abortion," the ability of users to see what was done and to fix it with a single click by reverting to a past version meant that acts of vandalism were corrected within minutes. Indeed, corrections were so rapid that vandalism acts and their corrections did not even appear on a mechanically generated image of the abortion definition as it changed over time./10 What is perhaps surprising is that this success occurs not in a tightly knit community with many social relations to reinforce the sense of common purpose and the social norms embodying it, but in a large and geographically dispersed group of otherwise unrelated participants. It suggests that even in a group of this size, social norms coupled with a facility to allow any participant to edit out purposeful or mistaken deviations in contravention of the social norms, and a robust platform for largely unmediated conversation, keep the group on track.

A very different cultural form of distributed content production is presented by the rise of massive multiplayer online games (MMOGs) as immersive entertainment. These fall in the same cultural "time slot" as television shows and movies of the twentieth century. The interesting thing about these types of games is that they organize the production of "scripts" very differently from movies or television shows. In a game like Ultima Online or EverQuest, the role of the commercial provider is not to tell a finished, highly polished story to be consumed start to finish by passive consumers. Rather, the role of the game provider is to build tools with which users collaborate to tell a story. There have been observations about this approach for years, regarding MUDs (Multi-User Dungeons) and MOOs (Multi-User Object Oriented games). The point to understand about MMOGs is that they produce a discrete element of "content" that was in the past dominated by centralized professional production. The screenwriter of an immersive entertainment product like a movie is like the scientist marking Mars craters - a professional producer of a finished good. In MMOGs, this function is produced by using the appropriate software platform to allow the story to be written by the many users as they experience it. The individual contributions of the users/coauthors of the story line are literally done for fun - they are playing a game. However, they are spending real economic goods - their attention and substantial subscription fees - on a form of entertainment that uses a platform for active coproduction of a story line to displace what was once passive reception of a finished, commercially and professionally manufactured good.

By 2003, a company called Linden Lab took this concept a major step forward by building an online game environment called Second Life. Second Life began almost entirely devoid of content. It was tools all the way down. Within a matter of months, it had thousands of subscribers, inhabiting a "world" that had thousands of characters, hundreds of thousands of objects, multiple areas, villages, and "story lines." The individual users themselves had created more than 99 percent of all objects in the game environment, and all story lines and substantive frameworks for interaction - such as a particular village or group of theme-based participants. The interactions in the game environment involved a good deal of gift giving and a good deal of trade, but also some very surprising structured behaviors. Some users set up a university, where lessons were given in both in-game skills and in programming. Others designed spaceships and engaged in alien abductions (undergoing one seemed to become a status symbol within the game). At one point, aiming (successfully) to prevent the company from changing its pricing policy, users staged a demonstration by making signs and picketing the entry point to the game; and a "tax revolt" by placing large numbers of "tea crates" around an in-game reproduction of the Washington Monument. Within months, Second Life had become an immersive experience, like a movie or book, but one where the commercial provider offered a platform and tools, while the users wrote the story lines, rendered the "set," and performed the entire play.

Relevance/Accreditation

How are we to know that the content produced by widely dispersed individuals is not sheer gobbledygook? Can relevance and accreditation itself be produced on a peer-production model? One type of answer is provided by looking at commercial businesses that successfully break off precisely the "accreditation and relevance" piece of their product, and rely on peer production to perform that function. Amazon and Google are probably the two most prominent examples of this strategy.

Amazon uses a mix of mechanisms to get in front of their buyers of books and other products that the users are likely to purchase. A number of these mechanisms produce relevance and accreditation by harnessing the users themselves. At the simplest level, the recommendation "customers who bought items you recently viewed also bought these items" is a mechanical means of extracting judgments of relevance and accreditation from the actions of many individuals, who produce the datum of relevance as by-product of making their own purchasing decisions. Amazon also allows users to create topical lists and track other users as their "friends and favorites." Amazon, like many consumer sites today, also provides users with the ability to rate books they buy, generating a peer-produced rating by averaging the ratings. More fundamentally, the core innovation of Google, widely recognized as the most efficient general search engine during the first half of the 2000s, was to introduce peer-based judgments of relevance. Like other search engines at the time, Google used a text-based algorithm to retrieve a given universe of Web pages initially. Its major innovation was its PageRank algorithm, which harnesses peer production of ranking in the following way. The engine treats links from other Web sites pointing to a given Web site as votes of confidence. Whenever someone who authors a Web site links to someone else's page, that person has stated quite explicitly that the linked page is worth a visit. Google's search engine counts these links as distributed votes of confidence in the quality of the page pointed to. Pages that are heavily linked-to count as more important votes of confidence. If a highly linked-to site links to a given page, that vote counts for more than the vote of a site that no one else thinks is worth visiting. The point to take home from looking at Google and Amazon is that corporations that have done immensely well at acquiring and retaining users have harnessed peer production to enable users to find things they want quickly and efficiently.

The most prominent example of a distributed project self-consciously devoted to peer production of relevance is the Open Directory Project. The site relies on more than sixty thousand volunteer editors to determine which links should be included in the directory. Acceptance as a volunteer requires application. Quality relies on a peer-review process based substantially on seniority as a volunteer and level of engagement with the site. The site is hosted and administered by Netscape, which pays for server space and a small number of employees to administer the site and set up the initial guidelines. Licensing is free and presumably adds value partly to America Online's (AOL's) and Netscape's commercial search engine/portal and partly through goodwill. Volunteers are not affiliated with Netscape and receive no compensation. They spend time selecting sites for inclusion in the directory (in small increments of perhaps fifteen minutes per site reviewed), producing the most comprehensive, highest-quality human-edited directory of the Web - at this point outshining the directory produced by the company that pioneered human edited directories of the Web: Yahoo!.

Perhaps the most elaborate platform for peer production of relevance and accreditation, at multiple layers, is used by Slashdot. Billed as "News for Nerds," Slashdot has become a leading technology newsletter on the Web, coproduced by hundreds of thousands of users. Slashdot primarily consists of users commenting on initial submissions that cover a variety of technology-related topics. The

submissions are typically a link to an off-site story, coupled with commentary from the person who submits the piece. Users follow up the initial submission with comments that often number in the hundreds. The initial submissions themselves, and more importantly, the approach to sifting through the comments of users for relevance and accreditation, provide a rich example of how this function can be performed on a distributed, peer-production model.

First, it is important to understand that the function of posting a story from another site onto Slashdot, the first "utterance" in a chain of comments on Slashdot, is itself an act of relevance production. The person submitting the story is telling the community of Slashdot users, "here is a story that 'News for Nerds' readers should be interested in." This initial submission of a link is itself very coarsely filtered by editors who are paid employees of Open Source Technology Group (OSTG), which runs a number of similar platforms - like SourceForge, the most important platform for free software developers. OSTG is a subsidiary of VA Software, a software services company. The FAQ (Frequently Asked Question) response to, "how do you verify the accuracy of Slashdot stories?" is revealing: "We don't. You do. If something seems outrageous, we might look for some corroboration, but as a rule, we regard this as the responsibility of the submitter and the audience. This is why it's important to read comments. You might find something that refutes, or supports, the story in the main." In other words, Slashdot very self-consciously is organized as a means of facilitating peer production of accreditation; it is at the comments stage that the story undergoes its most important form of accreditation - peer review ex-post.

Filtering and accreditation of comments on Slashdot offer the most interesting case study of peer production of these functions. Users submit comments that are displayed together with the initial submission of a story. Think of the "content" produced in these comments as a cross between academic peer review of journal submissions and a peer-produced substitute for television's "talking heads." It is in the means of accrediting and evaluating these comments that Slashdot's system provides a comprehensive example of peer production of relevance and accreditation. Slashdot implements an automated system to select moderators from the pool of users. Moderators are chosen according to several criteria; they must be logged in (not anonymous), they must be regular users (who use the site averagely, not one-time page loaders or compulsive users), they must have been using the site for a while (this defeats people who try to sign up just to moderate), they must be willing, and they must have positive "karma." Karma is a number assigned to a user that primarily reflects whether he or she has posted good or bad comments (according to ratings from other moderators). If a user meets these criteria, the program assigns the user moderator status and the user gets five "influence points" to review comments. The moderator rates a comment of his choice using a drop-down list with words such as "flamebait" and "informative." A positive word increases the rating of a comment one point and a negative word decreases the rating a point. Each time a moderator rates a comment, it costs one influence point, so he or she can only rate five comments for each moderating period. The period lasts for three days and if the user does not use the influence points, they expire. The moderation setup is designed to give many users a small amount of power. This decreases the effect of users with an ax to grind or with poor judgment. The site also implements some automated "troll filters," which prevent users from sabotaging the system. Troll filters stop users from posting more than once every sixty seconds, prevent identical posts, and will ban a user for twenty-four hours if he or she has been moderated down several times within a short time frame. Slashdot then provides users with a "threshold" filter that allows each user to block lower-quality comments. The scheme uses the numerical rating of the comment (ranging from -1 to 5). Comments start out at 0 for anonymous posters, 1 for registered users, and 2 for registered users with good "karma." As a result, if a user sets his or her filter at 1, the user will not see any comments from anonymous posters unless the comments' ratings were increased by a

moderator. A user can set his or her filter anywhere from -1 (viewing all of the comments) to 5 (where only the posts that have been upgraded by several moderators will show up).

Relevance, as distinct from accreditation, is also tied into the Slashdot scheme because off-topic posts should receive an "off topic" rating by the moderators and sink below the threshold level (assuming the user has the threshold set above the minimum). However, the moderation system is limited to choices that sometimes are not mutually exclusive. For instance, a moderator may have to choose between "funny" (+1) and "off topic" (-1) when a post is both funny and off topic. As a result, an irrelevant post can increase in ranking and rise above the threshold level because it is funny or informative. It is unclear, however, whether this is a limitation on relevance, or indeed mimics our own normal behavior, say in reading a newspaper or browsing a library, where we might let our eyes linger longer on a funny or informative tidbit, even after we have ascertained that it is not exactly relevant to what we were looking for.

The primary function of moderation is to provide accreditation. If a user sets a high threshold level, they will only see posts that are considered of high quality by the moderators. Users also receive accreditation through their karma. If their posts consistently receive high ratings, their karma will increase. At a certain karma level, their comments will start off with a rating of 2, thereby giving them a louder voice in the sense that users with a threshold of 2 will now see their posts immediately, and fewer upward moderations are needed to push their comments even higher. Conversely, a user with bad karma from consistently poorly rated comments can lose accreditation by having his or her posts initially start off at 0 or -1. In addition to the mechanized means of selecting moderators and minimizing their power to skew the accreditation system, Slashdot implements a system of peer-review accreditation for the moderators themselves. Slashdot accomplishes this "metamoderation" by making any user that has an account from the first 90 percent of accounts created on the system eligible to evaluate the moderators. Each eligible user who opts to perform metamoderation review is provided with ten random moderator ratings of comments. The user/metamoderator then rates the moderator's rating as either unfair, fair, or neither. The metamoderation process affects the karma of the original moderator, which, when lowered sufficiently by cumulative judgments of unfair ratings, will remove the moderator from the moderation system.

Together, these mechanisms allow for distributed production of both relevance and accreditation. Because there are many moderators who can moderate any given comment, and thanks to the mechanisms that explicitly limit the power of any one moderator to overinfluence the aggregate judgment, the system evens out differences in evaluation by aggregating judgments. It then allows individual users to determine what level of accreditation pronounced by this aggregate system fits their particular time and needs by setting their filter to be more or less inclusive. By introducing "karma," the system also allows users to build reputation over time, and to gain greater control over the accreditation of their own work relative to the power of the critics. Users, moderators, and metamoderators are all volunteers.

The primary point to take from the Slashdot example is that the same dynamic that we saw used for peer production of initial utterances, or content, can be implemented to produce relevance and accreditation. Rather than using the full-time effort of professional accreditation experts, the system is designed to permit the aggregation of many small judgments, each of which entails a trivial effort for the contributor, regarding both relevance and accreditation of the materials. The software that mediates the communication among the collaborating peers embeds both the means to facilitate the participation and a variety of mechanisms designed to defend the common effort from poor judgment or defection.

Value-Added Distribution

Finally, when we speak of information or cultural goods that exist (content has been produced) and are made usable through some relevance and accreditation mechanisms, there remains the question of distribution. To some extent, this is a nonissue on the Internet. Distribution is cheap. All one needs is a server and large pipes connecting one's server to the world. Nonetheless, this segment of the publication process has also provided us with important examples of peer production, including one of its earliest examples - Project Gutenberg.

Project Gutenberg entails hundreds of volunteers who scan in and correct books so that they are freely available in digital form. It has amassed more than 13,000 books, and makes the collection available to everyone for free. The vast majority of the "e-texts" offered are public domain materials. The site itself presents the e-texts in ASCII format, the lowest technical common denominator, but does not discourage volunteers from offering the e-texts in markup languages. It contains a search engine that allows a reader to search for typical fields such as subject, author, and title. Project Gutenberg volunteers can select any book that is in the public domain to transform into an e-text. The volunteer submits a copy of the title page of the book to Michael Hart - who founded the project - for copyright research. The volunteer is notified to proceed if the book passes the copyright clearance. The decision on which book to convert to e-text is left up to the volunteer, subject to copyright limitations. Typically, a volunteer converts a book to ASCII format using OCR (optical character recognition) and proofreads it one time in order to screen it for major errors. He or she then passes the ASCII file to a volunteer proofreader. This exchange is orchestrated with very little supervision. The volunteers use a Listserv mailing list and a bulletin board to initiate and supervise the exchange. In addition, books are labeled with a version number indicating how many times they have been proofed. The site encourages volunteers to select a book that has a low number and proof it. The Project Gutenberg proofing process is simple. Proofreaders (aside from the first pass) are not expected to have access to the book, but merely review the e-text for self-evident errors.

Distributed Proofreading, a site originally unaffiliated with Project Gutenberg, is devoted to proofing Project Gutenberg e-texts more efficiently, by distributing the volunteer proofreading function in smaller and more information-rich modules. Charles Franks, a computer programmer from Las Vegas, decided that he had a more efficient way to proofread these e-texts. He built an interface that allowed volunteers to compare scanned images of original texts with the e-texts available on Project Gutenberg. In the Distributed Proofreading process, scanned pages are stored on the site, and volunteers are shown a scanned page and a page of the e-text simultaneously so that they can compare the e-text to the original page. Because of the fine-grained modularity, proofreaders can come on the site and proof one or a few pages and submit them. By contrast, on the Project Gutenberg site, the entire book is typically exchanged, or at minimum, a chapter. In this fashion, Distributed Proofreading clears the proofing of tens of thousands of pages every month. After a couple of years of working independently, Franks joined forces with Hart. By late 2004, the site had proofread more than five thousand volumes using this method.

Sharing of Processing, Storage, and Communications Platforms

All the examples of peer production that we have seen up to this point have been examples where individuals pool their time, experience, wisdom, and creativity to form new information, knowledge, and cultural goods. As we look around the Internet, however, we find that users also cooperate in similar loosely affiliated groups, without market signals or managerial commands, to build supercomputers and massive data storage and retrieval systems. In their radical decentralization and

reliance on social relations and motivations, these sharing practices are similar to peer production of information, knowledge, and culture. They differ in one important aspect: Users are not sharing their innate and acquired human capabilities, and, unlike information, their inputs and outputs are not public goods. The participants are, instead, sharing material goods that they privately own, mostly personal computers and their components. They produce economic, not public, goods - computation, storage, and communications capacity.

As of the middle of 2004, the fastest supercomputer in the world was SETI@home. It ran about 75 percent faster than the supercomputer that was then formally known as "the fastest supercomputer in the world": the IBM Blue Gene/L. And yet, there was and is no single SETI@home computer. Instead, the SETI@home project has developed software and a collaboration platform that have enabled millions of participants to pool their computation resources into a single powerful computer. Every user who participates in the project must download a small screen saver. When a user's personal computer is idle, the screen saver starts up, downloads problems for calculation - in SETI@home, these are radio astronomy signals to be analyzed for regularities - and calculates the problem it has downloaded. Once the program calculates a solution, it automatically sends its results to the main site. The cycle continues for as long as, and repeats every time that, the computer is idle from its user's perspective. As of the middle of 2004, the project had harnessed the computers of 4.5 million users, allowing it to run computations at speeds greater than those achieved by the fastest supercomputers in the world that private firms, using full-time engineers, developed for the largest and best-funded government laboratories in the world. SETI@home is the most prominent, but is only one among dozens of similarly structured Internet-based distributed computing platforms. Another, whose structure has been the subject of the most extensive formal analysis by its creators, is Folding@home. As of mid-2004, Folding@home had amassed contributions of about 840,000 processors contributed by more than 365,000 users.

SETI@home and Folding@home provide a good basis for describing the fairly common characteristics of Internet-based distributed computation projects. First, these are noncommercial projects, engaged in pursuits understood as scientific, for the general good, seeking to harness contributions of individuals who wish to contribute to such larger-than-themselves goals. SETI@home helps in the search for extraterrestrial intelligence. Folding@home helps in protein folding research. Fightaids@home is dedicated to running models that screen compounds for the likelihood that they will provide good drug candidates to fight HIV/AIDS. Genome@home is dedicated to modeling artificial genes that would be created to generate useful proteins. Other sites, like those dedicated to cryptography or mathematics, have a narrower appeal, and combine "altruistic" with hobby as their basic motivational appeal. The absence of money is, in any event, typical of the large majority of active distributed computing projects. Less than one-fifth of these projects mention money at all. Most of those that do mention money refer to the contributors' eligibility for a share of a generally available prize for solving a scientific or mathematical challenge, and mix an appeal to hobby and altruism with the promise of money. Only two of about sixty projects active in 2004 were built on a pay-per-contribution basis, and these were quite small-scale by comparison to many of the others.

Most of the distributed computing projects provide a series of utilities and statistics intended to allow contributors to attach meaning to their contributions in a variety of ways. The projects appear to be eclectic in their implicit social and psychological theories of the motivations for participation in the projects. Sites describe the scientific purpose of the models and the specific scientific output, including posting articles that have used the calculations. In these components, the project organizers seem to assume some degree of taste for generalized altruism and the pursuit of meaning in contributing to a common goal. They also implement a variety of mechanisms to reinforce the

sense of purpose, such as providing aggregate statistics about the total computations performed by the project as a whole. However, the sites also seem to assume a healthy dose of what is known in the anthropology of gift literature as agonistic giving - that is, giving intended to show that the person giving is greater than or more important than others, who gave less. For example, most of the sites allow individuals to track their own contributions, and provide "user of the month"-type rankings. An interesting characteristic of quite a few of these is the ability to create "teams" of users, who in turn compete on who has provided more cycles or work units. SETI@home in particular taps into ready-made nationalisms, by offering country-level statistics. Some of the team names on Folding@home also suggest other, out-of-project bonding measures, such as national or ethnic bonds (for example, Overclockers Australia or Alliance Francophone), technical minority status (for example, Linux or MacAddict4Life), and organizational affiliation (University of Tennessee or University of Alabama), as well as shared cultural reference points (Knights who say Ni!). In addition, the sites offer platforms for simple connectedness and mutual companionship, by offering user fora to discuss the science and the social participation involved. It is possible that these sites are shooting in the dark, as far as motivating sharing is concerned. It also possible, however, that they have tapped into a valuable insight, which is that people behave sociably and generously for all sorts of different reasons, and that at least in this domain, adding reasons to participate - some agonistic, some altruistic, some reciprocity-seeking - does not have a crowding-out effect.

Like distributed computing projects, peer-to-peer file-sharing networks are an excellent example of a highly efficient system for storing and accessing data in a computer network. These networks of sharing are much less "mysterious," in terms of understanding the human motivation behind participation. Nevertheless, they provide important lessons about the extent to which large-scale collaboration among strangers or loosely affiliated users can provide effective communications platforms. For fairly obvious reasons, we usually think of peer-to-peer networks, beginning with Napster, as a "problem." This is because they were initially overwhelmingly used to perform an act that, by the analysis of almost any legal scholar, was copyright infringement. To a significant extent, they are still used in this form. There were, and continue to be, many arguments about whether the acts of the firms that provided peer-to-peer software were responsible for the violations. However, there has been little argument that anyone who allows thousands of other users to make copies of his or her music files is violating copyright - hence the public interpretation of the creation of peer-to-peer networks as primarily a problem. From the narrow perspective of the law of copyright or of the business model of the recording industry and Hollywood, this may be an appropriate focus. From the perspective of diagnosing what is happening to our social and economic structure, the fact that the files traded on these networks were mostly music in the first few years of this technology's implementation is little more than a distraction. Let me explain why.

Imagine for a moment that someone - be it a legislator defining a policy goal or a businessperson defining a desired service - had stood up in mid-1999 and set the following requirements: "We would like to develop a new music and movie distribution system. We would like it to store all the music and movies ever digitized. We would like it to be available from anywhere in the world. We would like it to be able to serve tens of millions of users at any given moment." Any person at the time would have predicted that building such a system would cost tens if not hundreds of millions of dollars; that running it would require large standing engineering staffs; that managing it so that users could find what they wanted and not drown in the sea of content would require some substantial number of "curators" - DJs and movie buffs - and that it would take at least five to ten years to build. Instead, the system was built cheaply by a wide range of actors, starting with Shawn Fanning's idea and implementation of Napster. Once the idea was out, others perfected the idea further, eliminating the need for even the one centralized feature that Napster included - a list of who had what files on which computer that provided the matchmaking function in the Napster

network. Since then, under the pressure of suits from the recording industry and a steady and persistent demand for peer-to-peer music software, rapid successive generations of Gnutella, and then the FastTrack clients KaZaa and Morpheus, Overnet and eDonkey, the improvements of BitTorrent, and many others have enhanced the reliability, coverage, and speed of the peer-to-peer music distribution system - all under constant threat of litigation, fines, police searches and even, in some countries, imprisonment of the developers or users of these networks.

What is truly unique about peer-to-peer networks as a signal of what is to come is the fact that with ridiculously low financial investment, a few teenagers and twenty-something-year-olds were able to write software and protocols that allowed tens of millions of computer users around the world to cooperate in producing the most efficient and robust file storage and retrieval system in the world. No major investment was necessary in creating a server farm to store and make available the vast quantities of data represented by the media files. The users' computers are themselves the "server farm." No massive investment in dedicated distribution channels made of high-quality fiber optics was necessary. The standard Internet connections of users, with some very intelligent file transfer protocols, sufficed. Architecture oriented toward enabling users to cooperate with each other in storage, search, retrieval, and delivery of files was all that was necessary to build a content distribution network that dwarfed anything that existed before.

Again, there is nothing mysterious about why users participate in peer-to-peer networks. They want music; they can get it from these networks for free; so they participate. The broader point to take from looking at peer-to-peer file-sharing networks, however, is the sheer effectiveness of large-scale collaboration among individuals once they possess, under their individual control, the physical capital necessary to make their cooperation effective. These systems are not "subsidized," in the sense that they do not pay the full marginal cost of their service. Remember, music, like all information, is a nonrival public good whose marginal cost, once produced, is zero. Moreover, digital files are not "taken" from one place in order to be played in the other. They are replicated wherever they are wanted, and thereby made more ubiquitous, not scarce. The only actual social cost involved at the time of the transmission is the storage capacity, communications capacity, and processing capacity necessary to store, catalog, search, retrieve, and transfer the information necessary to replicate the files from where copies reside to where more copies are desired. As with any nonrival good, if Jane is willing to spend the actual social costs involved in replicating the music file that already exists and that Jack possesses, then it is efficient that she do so without paying the creator a dime. It may throw a monkey wrench into the particular way in which our society has chosen to pay musicians and recording executives. This, as we saw in chapter 2, trades off efficiency for longer-term incentive effects for the recording industry. However, it is efficient within the normal meaning of the term in economics in a way that it would not have been had Jane and Jack used subsidized computers or network connections.

As with distributed computing, peer-to-peer file-sharing systems build on the fact that individual users own vast quantities of excess capacity embedded in their personal computers. As with distributed computing, peer-to-peer networks developed architectures that allowed users to share this excess capacity with each other. By cooperating in these sharing practices, users construct together systems with capabilities far exceeding those that they could have developed by themselves, as well as the capabilities that even the best-financed corporations could provide using techniques that rely on components they fully owned. The network components owned by any single music delivery service cannot match the collective storage and retrieval capabilities of the universe of users' hard drives and network connections. Similarly, the processors arrayed in the supercomputers find it difficult to compete with the vast computation resource available on the millions of personal computers connected to the Internet, and the proprietary software

development firms find themselves competing, and in some areas losing to, the vast pool of programming talent connected to the Internet in the form of participants in free and open source software development projects.

In addition to computation and storage, the last major element of computer communications networks is connectivity. Here, too, perhaps more dramatically than in either of the two other functionalities, we have seen the development of sharing-based techniques. The most direct transfer of the design characteristics of peer-to-peer networks to communications has been the successful development of Skype - an Internet telephony utility that allows the owners of computers to have voice conversations with each other over the Internet for free, and to dial into the public telephone network for a fee. As of this writing, Skype is already used by more than two million users at any given moment in time. They use a FastTrack-like architecture to share their computing and communications resources to create a global telephone system running on top of the Internet. It was created, and is run by, the developers of KaZaa.

Most dramatically, however, we have seen these techniques emerging in wireless communications. Throughout almost the entire twentieth century, radio communications used a single engineering approach to allow multiple messages to be sent wirelessly in a single geographic area. This approach was to transmit each of the different simultaneous messages by generating separate electromagnetic waves for each, which differed from each other by the frequency of oscillation, or wavelength. The receiver could then separate out the messages by ignoring all electromagnetic energy received at its antenna unless it oscillated at the frequency of the desired message. This engineering technique, adopted by Marconi in 1900, formed the basis of our notion of "spectrum": the range of frequencies at which we know how to generate electromagnetic waves with sufficient control and predictability that we can encode and decode information with them, as well as the notion that there are "channels" of spectrum that are "used" by a communication. For more than half a century, radio communications regulation was thought necessary because spectrum was scarce, and unless regulated, everyone would transmit at all frequencies causing chaos and an inability to send messages. From 1959, when Ronald Coase first published his critique of this regulatory approach, until the early 1990s, when spectrum auctions began, the terms of the debate over "spectrum policy," or wireless communications regulation, revolved around whether the exclusive right to transmit radio signals in a given geographic area should be granted as a regulatory license or a tradable property right. In the 1990s, with the introduction of auctions, we began to see the adoption of a primitive version of a property-based system through "spectrum auctions." By the early 2000s, this system allowed the new "owners" of these exclusive rights to begin to shift what were initially purely mobile telephony systems to mobile data communications as well.

By this time, however, the century-old engineering assumptions that underlay the regulation-versus-property conceptualization of the possibilities open for the institutional framework of wireless communications had been rendered obsolete by new computation and network technologies./11 The dramatic decline in computation cost and improvements in digital signal processing, network architecture, and antenna systems had fundamentally changed the design space of wireless communications systems. Instead of having one primary parameter with which to separate out messages - the frequency of oscillation of the carrier wave - engineers could now use many different mechanisms to allow much smarter receivers to separate out the message they wanted to receive from all other sources of electromagnetic radiation in the geographic area they occupied. Radio transmitters could now transmit at the same frequency, simultaneously, without "interfering" with each other - that is, without confusing the receivers as to which radiation carried the required message and which did not. Just like automobiles that can share a commons-based medium - the road - and unlike railroad cars, which must use dedicated, owned, and managed

railroad tracks - these new radios could share "the spectrum" as a commons. It was no longer necessary, or even efficient, to pass laws - be they in the form of regulations or of exclusive property-like rights - that carved up the usable spectrum into exclusively controlled slices. Instead, large numbers of transceivers, owned and operated by end users, could be deployed and use equipment-embedded protocols to coordinate their communications.

The reasons that owners would share the excess capacity of their new radios are relatively straightforward in this case. Users want to have wireless connectivity all the time, to be reachable and immediately available everywhere. However, they do not actually want to communicate every few microseconds. They will therefore be willing to purchase and keep turned on equipment that provides them with such connectivity. Manufacturers, in turn, will develop and adhere to standards that will improve capacity and connectivity. As a matter of engineering, what has been called "cooperation gain" - the improved quality of the system gained when the nodes cooperate - is the most promising source of capacity scaling for distributed wireless systems./12 Cooperation gain is easy to understand from day-to-day interactions. When we sit in a lecture and miss a word or two, we might turn to a neighbor and ask, "Did you hear what she said?" In radio systems, this kind of cooperation among the antennae (just like the ears) of neighbors is called antenna diversity, and is the basis for the design of a number of systems to improve reception. We might stand in a loud crowd without being able to shout or walk over to the other end of the room, but ask a friend: "If you see so and so, tell him *x*"; that friend then bumps into a friend of so and so and tells that person: "If you see so and so, tell him *x*"; and so forth. When we do this, we are using what in radio engineering is called repeater networks. These kinds of cooperative systems can carry much higher loads without interference, sharing wide swaths of spectrum, in ways that are more efficient than systems that rely on explicit market transactions based on property in the right to emit power in discrete frequencies. The design of such "ad hoc mesh networks" - that is, networks of radios that can configure themselves into cooperative networks as need arises, and help each other forward messages and decipher incoming messages over the din of radio emissions - are the most dynamic area in radio engineering today.

This technological shift gave rise to the fastest-growing sector in the wireless communications arena in the first few years of the twenty-first century - WiFi and similar unlicensed wireless devices. The economic success of the equipment market that utilizes the few primitive "spectrum commons" available in the United States - originally intended for low-power devices like garage openers and the spurious emissions of microwave ovens - led toward at first slow, and more recently quite dramatic, change in U.S. wireless policy. In the past two years alone, what have been called "commons-based" approaches to wireless communications policy have come to be seen as a legitimate, indeed a central, component of the Federal Communication Commission's (FCC's) wireless policy./13 We are beginning to see in this space the most prominent example of a system that was entirely oriented toward regulation aimed at improving the institutional conditions of marketbased production of wireless transport capacity sold as a finished good (connectivity minutes), shifting toward enabling the emergence of a market in shareable goods (smart radios) designed to provision transport on a sharing model.

I hope these detailed examples provide a common set of mental pictures of what peer production looks like. In the next chapter I explain the economics of peer production of information and the sharing of material resources for computation, communications, and storage in particular, and of nonmarket, social production more generally: why it is efficient, how we can explain the motivations that lead people to participate in these great enterprises of nonmarket cooperation, and why we see so much more of it online than we do off-line. The moral and political discussion throughout the remainder of the book does not, however, depend on your accepting the particular

analysis I offer in chapter 4 to "domesticate" these phenomena within more or less standard economics. At this point, it is important that the stories have provided a texture for, and established the plausibility of, the claim that nonmarket production in general and peer production in particular are phenomena of much wider application than free software, and exist in important ways throughout the networked information economy. For purposes of understanding the political implications that occupy most of this book, that is all that is necessary.

Notes

1. For an excellent history of the free software movement and of open-source development, see Glyn Moody, *Rebel Code: Inside Linux and the Open Source Revolution* (New York: Perseus Publishing, 2001).

2. Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge: Cambridge University Press, 1990).

3. Josh Lerner and Jean Tirole, "The Scope of Open Source Licensing" (Harvard NOM working paper no. 02-42, table 1, Cambridge, MA, 2002). The figure is computed out of the data reported in this paper for the number of free software development projects that Lerner and Tirole identify as having "restrictive" or "very restrictive" licenses.

4. Netcraft, April 2004 Web Server Survey, http://news.netcraft.com/archives/web_server_survey.html .

5. Clickworkers Results: Crater Marking Activity, July 3, 2001, http://clickworkers.arc.nasa.gov/documents/crater-marking.pdf .

6. B. Kanefsky, N. G. Barlow, and V. C. Gulick, *Can Distributed Volunteers Accomplish Massive Data Analysis Tasks?* http://www.clickworkers.arc.nasa.gov/documents/abstract.pdf .

7. J. Giles, "Special Report: Internet Encyclopedias Go Head to Head," *Nature,* December 14, 2005, available at http://www.nature.com/news/2005/051212/full/438900a.html .

8. http://www.techcentralstation.com/111504A.html .

9. Yochai Benkler, "Coase's Penguin, or Linux and the Nature of the Firm," *Yale Law Journal* 112 (2001): 369.

10. IBM Collaborative User Experience Research Group, History Flows: Results (2003), http://www.research.ibm.com/history/results.htm .

11. For the full argument, see Yochai Benkler, "Some Economics of Wireless Communications," *Harvard Journal of Law and Technology* 16 (2002): 25; and Yochai Benkler, "Overcoming Agoraphobia: Building the Commons of the Digitally Networked Environment," *Harvard Journal of Law and Technology*11 (1998): 287. For an excellent overview of the intellectual history of this debate and a contribution to the institutional design necessary to make space for this change, see Kevin Werbach, "Supercommons: Towards a Unified Theory of Wireless Communication," *Texas Law Review* 82 (2004): 863. The policy implications of computationally intensive radios using wide bands were first raised by George Gilder in "The New Rule of the Wireless," *Forbes ASAP*, March 29, 1993, and Paul Baran, "Visions of the 21st Century Communications: Is the Shortage of Radio Spectrum for Broadband Networks of the Future a Self Made Problem?" (keynote talk transcript,

8th Annual Conference on Next Generation Networks, Washington, DC, November 9, 1994). Both statements focused on the potential abundance of spectrum, and how it renders "spectrum management" obsolete. Eli Noam was the first to point out that, even if one did not buy the idea that computationally intensive radios eliminated scarcity, they still rendered spectrum property rights obsolete, and enabled instead a fluid, dynamic, real-time market in spectrum clearance rights. See Eli Noam, "Taking the Next Step Beyond Spectrum Auctions: Open Spectrum Access," *Institute of Electrical and Electronics Engineers Communications Magazine* 33, no. 12 (1995): 66-73; later elaborated in Eli Noam, "Spectrum Auction: Yesterday's Heresy, Today's Orthodoxy, Tomorrow's Anachronism. Taking the Next Step to Open Spectrum Access," *Journal of Law and Economics* 41 (1998): 765, 778-780. The argument that equipment markets based on a spectrum commons, or free access to frequencies, could replace the role planned for markets in spectrum property rights with computationally intensive equipment and sophisticated network sharing protocols, and would likely be more efficient even assuming that scarcity persists, was made in Benkler, "Overcoming Agoraphobia." Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York: Basic Books, 1999) and Lawrence Lessig, *The Future of Ideas: The Fate of the Commons in a Connected World* (New York: Random House, 2001) developed a rationale based on the innovation dynamic in support of the economic value of open wireless networks. David Reed, "Comments for FCC Spectrum Task Force on Spectrum Policy," filed with the Federal Communications Commission July 10, 2002, crystallized the technical underpinnings and limitations of the idea that spectrum can be regarded as property.

12. See Benkler, "Some Economics," 44-47. The term "cooperation gain" was developed by Reed to describe a somewhat broader concept than "diversity gain" is in multiuser information theory.

13. *Spectrum Policy Task Force Report to the Commission* (Federal Communications Commission, Washington, DC, 2002); Michael K. Powell, "Broadband Migration III: New Directions in Wireless Policy" (Remarks at the Silicon Flatiron Telecommunications Program, University of Colorado at Boulder, October 30, 2002).

Chapter 4
The Economics of Social Production

The increasing salience of nonmarket production in general, and peer production in particular, raises three puzzles from an economics perspective. First, why do people participate? What is their motivation when they work for or contribute resources to a project for which they are not paid or directly rewarded? Second, why now, why here? What, if anything, is special about the digitally networked environment that would lead us to believe that peer production is here to stay as an important economic phenomenon, as opposed to a fad that will pass as the medium matures and patterns of behavior settle toward those more familiar to us from the economy of steel, coal, and temp agencies. Third, is it efficient to have all these people sharing their computers and donating their time and creative effort? Moving through the answers to these questions, it becomes clear that the diverse and complex patterns of behavior observed on the Internet, from Viking ship hobbyists to the developers of the GNU/Linux operating system, are perfectly consistent with much of our contemporary understanding of human economic behavior. We need to assume no fundamental change in the nature of humanity; we need not declare the end of economics as we know it. We merely need to see that the material conditions of production in the networked information economy have changed in ways that increase the relative salience of social sharing and exchange as a modality of economic production. That is, behaviors and motivation patterns familiar to us from social relations generally continue to cohere in their own patterns. What has changed is that now these patterns of behavior have become effective beyond the domains of building social relations of mutual interest and fulfilling our emotional and psychological needs of companionship and mutual recognition. They have come to play a substantial role as modes of motivating, informing, and organizing productive behavior at the very core of the information economy. And it is this increasing role as a modality of information production that ripples through the rest this book. It is the feasibility of producing information, knowledge, and culture through social, rather than market and proprietary relations - through cooperative peer production and coordinate individual action - that creates the opportunities for greater autonomous action, a more critical culture, a more discursively engaged and better informed republic, and perhaps a more equitable global community.

Motivation

Much of economics achieves analytic tractability by adopting a very simple model of human motivation. The basic assumption is that all human motivations can be more or less reduced to something like positive and negative utilities - things people want, and things people want to avoid. These are capable of being summed, and are usually translatable into a universal medium of exchange, like money. Adding more of something people want, like money, to any given interaction will, all things considered, make that interaction more desirable to rational people. While simplistic, this highly tractable model of human motivation has enabled policy prescriptions that have proven far more productive than prescriptions that depended on other models of human motivation - such as assuming that benign administrators will be motivated to serve their people, or that individuals will undertake self-sacrifice for the good of the nation or the commune.

Of course, this simple model underlying much of contemporary economics is wrong. At least it is wrong as a universal description of human motivation. If you leave a fifty-dollar check on the table at the end of a dinner party at a friend's house, you do not increase the probability that you will be invited again. We live our lives in diverse social frames, and money has a complex relationship with these - sometimes it adds to the motivation to participate, sometimes it detracts from it. While this

is probably a trivial observation outside of the field of economics, it is quite radical within that analytic framework. The present generation's efforts to formalize and engage it began with the Titmuss-Arrow debate of the early 1970s. In a major work, Richard Titmuss compared the U.S. and British blood supply systems. The former was largely commercial at the time, organized by a mix of private for-profit and nonprofit actors; the latter entirely voluntary and organized by the National Health Service. Titmuss found that the British system had higher-quality blood (as measured by the likelihood of recipients contracting hepatitis from transfusions), less blood waste, and fewer blood shortages at hospitals. Titmuss also attacked the U.S. system as inequitable, arguing that the rich exploited the poor and desperate by buying their blood. He concluded that an altruistic blood procurement system is both more ethical and more efficient than a market system, and recommended that the market be kept out of blood donation to protect the "right to give."[1] Titmuss's argument came under immediate attack from economists. Most relevant for our purposes here, Kenneth Arrow agreed that the differences in blood quality indicated that the U.S. blood system was flawed, but rejected Titmuss's central theoretical claim that markets reduce donative activity. Arrow reported the alternative hypothesis held by "economists typically," that if some people respond to exhortation/moral incentives (donors), while others respond to prices and market incentives (sellers), these two groups likely behave independently - neither responds to the other's incentives. Thus, the decision to allow or ban markets should have no effect on donative behavior. Removing a market could, however, remove incentives of the "bad blood" suppliers to sell blood, thereby improving the overall quality of the blood supply. Titmuss had not established his hypothesis analytically, Arrow argued, and its proof or refutation would lie in empirical study.[2] Theoretical differences aside, the U.S. blood supply system did in fact transition to an all-volunteer system of social donation since the 1970s. In surveys since, blood donors have reported that they "enjoy helping" others, experienced a sense of moral obligation or responsibility, or exhibited characteristics of reciprocators after they or their relatives received blood.

A number of scholars, primarily in psychology and economics, have attempted to resolve this question both empirically and theoretically. The most systematic work within economics is that of Swiss economist Bruno Frey and various collaborators, building on the work of psychologist Edward Deci.[3] A simple statement of this model is that individuals have intrinsic and extrinsic motivations. Extrinsic motivations are imposed on individuals from the outside. They take the form of either offers of money for, or prices imposed on, behavior, or threats of punishment or reward from a manager or a judge for complying with, or failing to comply with, specifically prescribed behavior. Intrinsic motivations are reasons for action that come from within the person, such as pleasure or personal satisfaction. Extrinsic motivations are said to "crowd out" intrinsic motivations because they (a) impair self-determination - that is, people feel pressured by an external force, and therefore feel overjustified in maintaining their intrinsic motivation rather than complying with the will of the source of the extrinsic reward; or (b) impair self-esteem - they cause individuals to feel that their internal motivation is rejected, not valued, and as a result, their self-esteem is diminished, causing them to reduce effort. Intuitively, this model relies on there being a culturally contingent notion of what one "ought" to do if one is a well-adjusted human being and member of a decent society. Being offered money to do something you know you "ought" to do, and that self-respecting members of society usually in fact do, implies that the person offering the money believes that you are not a well-adjusted human being or an equally respectable member of society. This causes the person offered the money either to believe the offerer, and thereby lose self-esteem and reduce effort, or to resent him and resist the offer. A similar causal explanation is formalized by Roland Benabou and Jean Tirole, who claim that the person receiving the monetary incentives infers that the person offering the compensation does not trust the offeree to do the right thing, or to do it well of their own accord. The offeree's self-confidence and intrinsic motivation to succeed

are reduced to the extent that the offeree believes that the offerer - a manager or parent, for example - is better situated to judge the offeree's abilities.[/4](#)

More powerful than the theoretical literature is the substantial empirical literature - including field and laboratory experiments, econometrics, and surveys - that has developed since the mid-1990s to test the hypotheses of this model of human motivation. Across many different settings, researchers have found substantial evidence that, under some circumstances, adding money for an activity previously undertaken without price compensation reduces, rather than increases, the level of activity. The work has covered contexts as diverse as the willingness of employees to work more or to share their experience and knowledge with team members, of communities to accept locally undesirable land uses, or of parents to pick up children from day-care centers punctually.[/5](#) The results of this empirical literature strongly suggest that across various domains some displacement or crowding out can be identified between monetary rewards and nonmonetary motivations. This does not mean that offering monetary incentives does not increase extrinsic rewards - it does. Where extrinsic rewards dominate, this will increase the activity rewarded as usually predicted in economics. However, the effect on intrinsic motivation, at least sometimes, operates in the opposite direction. Where intrinsic motivation is an important factor because pricing and contracting are difficult to achieve, or because the payment that can be offered is relatively low, the aggregate effect may be negative. Persuading experienced employees to communicate their tacit knowledge to the teams they work with is a good example of the type of behavior that is very hard to specify for efficient pricing, and therefore occurs more effectively through social motivations for teamwork than through payments. Negative effects of small payments on participation in work that was otherwise volunteer-based are an example of low payments recruiting relatively few people, but making others shift their efforts elsewhere and thereby reducing, rather than increasing, the total level of volunteering for the job.

The psychology-based alternative to the "more money for an activity will mean more of the activity" assumption implicit in most of these new economic models is complemented by a sociology-based alternative. This comes from one branch of the social capital literature - the branch that relates back to Mark Granovetter's 1974 book, *Getting a Job*, and was initiated as a crossover from sociology to economics by James Coleman.[/6](#) This line of literature rests on the claim that, as Nan Lin puts it, "there are two ultimate (or primitive) rewards for human beings in a social structure: economic standing and social standing."[/7](#) These rewards are understood as instrumental and, in this regard, are highly amenable to economics. Both economic and social aspects represent "standing" - that is, a relational measure expressed in terms of one's capacity to mobilize resources. Some resources can be mobilized by money. Social relations can mobilize others. For a wide range of reasons - institutional, cultural, and possibly technological - some resources are more readily capable of being mobilized by social relations than by money. If you want to get your nephew a job at a law firm in the United States today, a friendly relationship with the firm's hiring partner is more likely to help than passing on an envelope full of cash. If this theory of social capital is correct, then sometimes you should be willing to trade off financial rewards for social capital. Critically, the two are not fungible or cumulative. A hiring partner paid in an economy where monetary bribes for job interviews are standard does not acquire a social obligation. That same hiring partner in that same culture, who is also a friend and therefore forgoes payment, however, probably does acquire a social obligation, tenable for a similar social situation in the future. The magnitude of the social debt, however, may now be smaller. It is likely measured by the amount of money saved from not having to pay the price, not by the value of getting the nephew a job, as it would likely be in an economy where jobs cannot be had for bribes. There are things and behaviors, then, that simply cannot be commodified for market exchange, like friendship. Any effort to mix the two, to pay for one's friendship, would render it something completely different - perhaps a psychoanalysis session in our

culture. There are things that, even if commodified, can still be used for social exchange, but the meaning of the social exchange would be diminished. One thinks of borrowing eggs from a neighbor, or lending a hand to friends who are moving their furniture to a new apartment. And there are things that, even when commodified, continue to be available for social exchange with its full force. Consider gamete donations as an example in contemporary American culture. It is important to see, though, that there is nothing intrinsic about any given "thing" or behavior that makes it fall into one or another of these categories. The categories are culturally contingent and cross-culturally diverse. What matters for our purposes here, though, is only the realization that for any given culture, there will be some acts that a person would prefer to perform not for money, but for social standing, recognition, and probably, ultimately, instrumental value obtainable only if that person has performed the action through a social, rather than a market, transaction.

It is not necessary to pin down precisely the correct or most complete theory of motivation, or the full extent and dimensions of crowding out nonmarket rewards by the introduction or use of market rewards. All that is required to outline the framework for analysis is recognition that there is some form of social and psychological motivation that is neither fungible with money nor simply cumulative with it. Transacting within the price system may either increase or decrease the social-psychological rewards (be they intrinsic or extrinsic, functional or symbolic). The intuition is simple. As I have already said, leaving a fifty-dollar check on the table after one has finished a pleasant dinner at a friend's house would not increase the host's social and psychological gains from the evening. Most likely, it would diminish them sufficiently that one would never again be invited. A bottle of wine or a bouquet of flowers would, to the contrary, improve the social gains. And if dinner is not intuitively obvious, think of sex. The point is simple. Money-oriented motivations are different from socially oriented motivations. Sometimes they align. Sometimes they collide. Which of the two will be the case is historically and culturally contingent. The presence of money in sports or entertainment reduced the social psychological gains from performance in late-nineteenth-century Victorian England, at least for members of the middle and upper classes. This is reflected in the long-standing insistence on the "amateur" status of the Olympics, or the status of "actors" in the Victorian society. This has changed dramatically more than a century later, where athletes' and popular entertainers' social standing is practically measured in the millions of dollars their performances can command.

The relative relationships of money and social-psychological rewards are, then, dependent on culture and context. Similar actions may have different meanings in different social or cultural contexts. Consider three lawyers contemplating whether to write a paper presenting their opinion - one is a practicing attorney, the second is a judge, and the third is an academic. For the first, money and honor are often, though not always, positively correlated. Being able to command a very high hourly fee for writing the requested paper is a mode of expressing one's standing in the profession, as well as a means of putting caviar on the table. Yet, there are modes of acquiring esteem - like writing the paper as a report for a bar committee - that are not improved by the presence of money, and are in fact undermined by it. This latter effect is sharpest for the judge. If a judge is approached with an offer of money for writing an opinion, not only is this not a mark of honor, it is a subversion of the social role and would render corrupt the writing of the opinion. For the judge, the intrinsic "rewards" for writing the opinion when matched by a payment for the product would be guilt and shame, and the offer therefore an expression of disrespect. Finally, if the same paper is requested of the academic, the presence of money is located somewhere in between the judge and the practitioner. To a high degree, like the judge, the academic who writes for money is rendered suspect in her community of scholarship. A paper clearly funded by a party, whose results support the party's regulatory or litigation position, is practically worthless as an academic work. In a mirror image of the practitioner, however, there are some forms of money that add to and

reinforce an academic's social psychological rewards - peer-reviewed grants and prizes most prominent among them.

Moreover, individuals are not monolithic agents. While it is possible to posit idealized avaricious money-grubbers, altruistic saints, or social climbers, the reality of most people is a composite of these all, and one that is not like any of them. Clearly, some people are more focused on making money, and others are more generous; some more driven by social standing and esteem, others by a psychological sense of well-being. The for-profit and nonprofit systems probably draw people with different tastes for these desiderata. Academic science and commercial science also probably draw scientists with similar training but different tastes for types of rewards. However, well-adjusted, healthy individuals are rarely monolithic in their requirements. We would normally think of someone who chose to ignore and betray friends and family to obtain either more money or greater social recognition as a fetishist of some form or another. We spend some of our time making money, some of our time enjoying it hedonically; some of our time being with and helping family, friends, and neighbors; some of our time creatively expressing ourselves, exploring who we are and what we would like to become. Some of us, because of economic conditions we occupy, or because of our tastes, spend very large amounts of time trying to make money - whether to become rich or, more commonly, just to make ends meet. Others spend more time volunteering, chatting, or writing.

For all of us, there comes a time on any given day, week, and month, every year and in different degrees over our lifetimes, when we choose to act in some way that is oriented toward fulfilling our social and psychological needs, not our market-exchangeable needs. It is that part of our lives and our motivational structure that social production taps, and on which it thrives. There is nothing mysterious about this. It is evident to any of us who rush home to our family or to a restaurant or bar with friends at the end of a workday, rather than staying on for another hour of overtime or to increase our billable hours; or at least regret it when we cannot. It is evident to any of us who has ever brought a cup of tea to a sick friend or relative, or received one; to anyone who has lent a hand moving a friend's belongings; played a game; told a joke, or enjoyed one told by a friend. What needs to be understood now, however, is under what conditions these many and diverse social actions can turn into an important modality of economic production. When can all these acts, distinct from our desire for money and motivated by social and psychological needs, be mobilized, directed, and made effective in ways that we recognize as economically valuable?

Social Production: Feasibility Conditions and Organizational Form

The core technologically contingent fact that enables social relations to become a salient modality of production in the networked information economy is that all the inputs necessary to effective productive activity are under the control of individual users. Human creativity, wisdom, and life experience are all possessed uniquely by individuals. The computer processors, data storage devices, and communications capacity necessary to make new meaningful conversational moves from the existing universe of information and stimuli, and to render and communicate them to others near and far are also under the control of these same individual users - at least in the advanced economies and in some portions of the population of developing economies. This does not mean that all the physical capital necessary to process, store, and communicate information is under individual user control. That is not necessary. It is, rather, that the majority of individuals in these societies have the threshold level of material capacity required to explore the information environment they occupy, to take from it, and to make their own contributions to it.

There is nothing about computation or communication that naturally or necessarily enables this fact. It is a felicitous happenstance of the fabrication technology of computing machines in the last quarter of the twentieth century, and, it seems, in the reasonably foreseeable future. It is cheaper to build freestanding computers that enable their owners to use a wide and dynamically changing range of information applications, and that are cheap enough that each machine is owned by an individual user or household, than it is to build massive supercomputers with incredibly high-speed communications to yet cheaper simple terminals, and to sell information services to individuals on an on-demand or standardized package model. Natural or contingent, it is nevertheless a fact of the industrial base of the networked information economy that individual users - susceptible as they are to acting on diverse motivations, in diverse relationships, some market-based, some social - possess and control the physical capital necessary to make effective the human capacities they uniquely and individually possess.

Now, having the core inputs of information production ubiquitously distributed in society is a core enabling fact, but it alone cannot assure that social production will become economically significant. Children and teenagers, retirees, and very rich individuals can spend most of their lives socializing or volunteering; most other people cannot. While creative capacity and judgment are universally distributed in a population, available time and attention are not, and human creative capacity cannot be fully dedicated to nonmarket, nonproprietary production all the time. Someone needs to work for money, at least some of the time, to pay the rent and put food on the table. Personal computers too are only used for earnings-generating activities some of the time. In both these resources, there remain large quantities of excess capacity - time and interest in human beings; processing, storage, and communications capacity in computers - available to be used for activities whose rewards are not monetary or monetizable, directly or indirectly.

For this excess capacity to be harnessed and become effective, the information production process must effectively integrate widely dispersed contributions, from many individual human beings and machines. These contributions are diverse in their quality, quantity, and focus, in their timing and geographic location. The great success of the Internet generally, and peer-production processes in particular, has been the adoption of technical and organizational architectures that have allowed them to pool such diverse efforts effectively. The core characteristics underlying the success of these enterprises are their modularity and their capacity to integrate many fine-grained contributions.

"Modularity" is a property of a project that describes the extent to which it can be broken down into smaller components, or modules, that can be independently produced before they are assembled into a whole. If modules are independent, individual contributors can choose what and when to contribute independently of each other. This maximizes their autonomy and flexibility to define the nature, extent, and timing of their participation in the project. Breaking up the maps of Mars involved in the clickworkers project (described in chapter 3) and rendering them in small segments with a simple marking tool is a way of modularizing the task of mapping craters. In the SETI@home project (see chapter 3), the task of scanning radio astronomy signals is broken down into millions of little computations as a way of modularizing the calculations involved.

"Granularity" refers to the size of the modules, in terms of the time and effort that an individual must invest in producing them. The five minutes required for moderating a comment on Slashdot, or for metamoderating a moderator, is more fine-grained than the hours necessary to participate in writing a bug fix in an open-source project. More people can participate in the former than in the latter, independent of the differences in the knowledge required for participation. The number of people who can, in principle, participate in a project is therefore inversely related to the size of the

smallest-scale contribution necessary to produce a usable module. The granularity of the modules therefore sets the smallest possible individual investment necessary to participate in a project. If this investment is sufficiently low, then "incentives" for producing that component of a modular project can be of trivial magnitude. Most importantly for our purposes of understanding the rising role of nonmarket production, the time can be drawn from the excess time we normally dedicate to having fun and participating in social interactions. If the finest-grained contributions are relatively large and would require a large investment of time and effort, the universe of potential contributors decreases. A successful large-scale peer-production project must therefore have a predominate portion of its modules be relatively fine-grained.

Perhaps the clearest example of how large-grained modules can make projects falter is the condition, as of the middle of 2005, of efforts to peer produce open textbooks. The largest such effort is Wikibooks, a site associated with *Wikipedia*, which has not taken off as did its famous parent project. Very few texts there have reached maturity to the extent that they could be usable as a partial textbook, and those few that have were largely written by one individual with minor contributions by others. Similarly, an ambitious initiative launched in California in 2004 still had not gone far beyond an impassioned plea for help by mid-2005. The project that seems most successful as of 2005 was a South African project, Free High School Science Texts (FHSST), founded by a physics graduate student, Mark Horner. As of this writing, that three-year-old project had more or less completed a physics text, and was about halfway through chemistry and mathematics textbooks. The whole FHSST project involves a substantially more managed approach than is common in peer-production efforts, with a core group of dedicated graduate student administrators recruiting contributors, assigning tasks, and integrating the contributions. Horner suggests that the basic limiting factor is that in order to write a high school textbook, the output must comply with state-imposed guidelines for content and form. To achieve these requirements, the various modules must cohere to a degree much larger than necessary in a project like *Wikipedia*, which can endure high diversity in style and development without losing its utility. As a result, the individual contributions have been kept at a high level of abstraction - an idea or principle explained at a time. The minimal time commitment required of each contributor is therefore large, and has led many of those who volunteered initially to not complete their contributions. In this case, the guideline requirements constrained the project's granularity, and thereby impeded its ability to grow and capture the necessary thousands of small-grained contributions. With orders of magnitude fewer contributors, each must be much more highly motivated and available than is necessary in *Wikipedia*, Slashdot, and similar successful projects.

It is not necessary, however, that each and every chunk or module be fine grained. Free software projects in particular have shown us that successful peer-production projects may also be structured, technically and culturally, in ways that make it possible for different individuals to contribute vastly different levels of effort commensurate with their ability, motivation, and availability. The large free software projects might integrate thousands of people who are acting primarily for social psychological reasons - because it is fun or cool; a few hundred young programmers aiming to make a name for themselves so as to become employable; and dozens of programmers who are paid to write free software by firms that follow one of the nonproprietary strategies described in chapter 2. IBM and Red Hat are the quintessential examples of firms that contribute paid employee time to peer-production projects in this form. This form of link between a commercial firm and a peer production community is by no means necessary for a peer-production process to succeed; it does, however, provide one constructive interface between market- and nonmarket-motivated behavior, through which actions on the two types of motivation can reinforce, rather than undermine, each other.

The characteristics of planned modularization of a problem are highly visible and explicit in some peer-production projects - the distributed computing projects like SETI@home are particularly good examples of this. However, if we were to step back and look at the entire phenomenon of Web-based publication from a bird's-eye view, we would see that the architecture of the World Wide Web, in particular the persistence of personal Web pages and blogs and their self-contained, technical independence of each other, give the Web as a whole the characteristics of modularity and variable but fine-grained granularity. Imagine that you were trying to evaluate how, if at all, the Web is performing the task of media watchdog. Consider one example, which I return to in chapter 7: The Memory Hole, a Web site created and maintained by Russ Kick, a freelance author and editor. Kick spent some number of hours preparing and filing a Freedom of Information Act request with the Defense Department, seeking photographs of coffins of U.S. military personnel killed in Iraq. He was able to do so over some period, not having to rely on "getting the scoop" to earn his dinner. At the same time, tens of thousands of other individual Web publishers and bloggers were similarly spending their time hunting down stories that moved them, or that they happened to stumble across in their own daily lives. When Kick eventually got the photographs, he could upload them onto his Web site, where they were immediately available for anyone to see. Because each contribution like Kick's can be independently created and stored, because no single permission point or failure point is present in the architecture of the Web - it is merely a way of conveniently labeling documents stored independently by many people who are connected to the Internet and use HTML (hypertext markup language) and HTTP (hypertext transfer protocol) - as an "information service," it is highly modular and diversely granular. Each independent contribution comprises as large or small an investment as its owner-operator chooses to make. Together, they form a vast almanac, trivia trove, and news and commentary facility, to name but a few, produced by millions of people at their leisure - whenever they can or want to, about whatever they want.

The independence of Web sites is what marks their major difference from more organized peer-production processes, where contributions are marked not by their independence but by their interdependence. The Web as a whole requires no formal structure of cooperation. As an "information good" or medium, it emerges as a pattern out of coordinate coexistence of millions of entirely independent acts. All it requires is a pattern recognition utility superimposed over the outputs of these acts - a search engine or directory. Peer-production processes, to the contrary, do generally require some substantive cooperation among users. A single rating of an individual comment on Slashdot does not by itself moderate the comment up or down, neither does an individual marking of a crater. Spotting a bug in free software, proposing a fix, reviewing the proposed fix, and integrating it into the software are interdependent acts that require a level of cooperation. This necessity for cooperation requires peer-production processes to adopt more engaged strategies for assuring that everyone who participates is doing so in good faith, competently, and in ways that do not undermine the whole, and weeding out those would-be participants who are not.

Cooperation in peer-production processes is usually maintained by some combination of technical architecture, social norms, legal rules, and a technically backed hierarchy that is validated by social norms. *Wikipedia* is the strongest example of a discourse-centric model of cooperation based on social norms. However, even *Wikipedia* includes, ultimately, a small number of people with system administrator privileges who can eliminate accounts or block users in the event that someone is being genuinely obstructionist. This technical fallback, however, appears only after substantial play has been given to self-policing by participants, and to informal and quasi-formal community-based dispute resolution mechanisms. Slashdot, by contrast, provides a strong model of a sophisticated technical system intended to assure that no one can "defect" from the cooperative enterprise of commenting and moderating comments. It limits behavior enabled by the system to avoid

destructive behavior before it happens, rather than policing it after the fact. The Slash code does this by technically limiting the power any given person has to moderate anyone else up or down, and by making every moderator the subject of a peer review system whose judgments are enforced technically - that is, when any given user is described by a sufficiently large number of other users as unfair, that user automatically loses the technical ability to moderate the comments of others. The system itself is a free software project, licensed under the GPL (General Public License) - which is itself the quintessential example of how law is used to prevent some types of defection from the common enterprise of peer production of software. The particular type of defection that the GPL protects against is appropriation of the joint product by any single individual or firm, the risk of which would make it less attractive for anyone to contribute to the project to begin with. The GPL assures that, as a legal matter, no one who contributes to a free software project need worry that some other contributor will take the project and make it exclusively their own. The ultimate quality judgments regarding what is incorporated into the "formal" releases of free software projects provide the clearest example of the extent to which a meritocratic hierarchy can be used to integrate diverse contributions into a finished single product. In the case of the Linux kernel development project (see chapter 3), it was always within the power of Linus Torvalds, who initiated the project, to decide which contributions should be included in a new release, and which should not. But it is a funny sort of hierarchy, whose quirkiness Steve Weber well explicates./8 Torvalds's authority is persuasive, not legal or technical, and certainly not determinative. He can do nothing except persuade others to prevent them from developing anything they want and add it to their kernel, or to distribute that alternative version of the kernel. There is nothing he can do to prevent the entire community of users, or some subsection of it, from rejecting his judgment about what ought to be included in the kernel. Anyone is legally free to do as they please. So these projects are based on a hierarchy of meritocratic respect, on social norms, and, to a great extent, on the mutual recognition by most players in this game that it is to everybody's advantage to have someone overlay a peer review system with some leadership.

In combination then, three characteristics make possible the emergence of information production that is not based on exclusive proprietary claims, not aimed toward sales in a market for either motivation or information, and not organized around property and contract claims to form firms or market exchanges. First, the physical machinery necessary to participate in information and cultural production is almost universally distributed in the population of the advanced economies. Certainly, personal computers as capital goods are under the control of numbers of individuals that are orders of magnitude larger than the number of parties controlling the use of mass-production-capable printing presses, broadcast transmitters, satellites, or cable systems, record manufacturing and distribution chains, and film studios and distribution systems. This means that the physical machinery can be put in service and deployed in response to any one of the diverse motivations individual human beings experience. They need not be deployed in order to maximize returns on the financial capital, because financial capital need not be mobilized to acquire and put in service any of the large capital goods typical of the industrial information economy. Second, the primary raw materials in the information economy, unlike the industrial economy, are public goods - existing information, knowledge, and culture. Their actual marginal social cost is zero. Unless regulatory policy makes them purposefully expensive in order to sustain the proprietary business models, acquiring raw materials also requires no financial capital outlay. Again, this means that these raw materials can be deployed for any human motivation. They need not maximize financial returns. Third, the technical architectures, organizational models, and social dynamics of information production and exchange on the Internet have developed so that they allow us to structure the solution to problems - in particular to information production problems - in ways that are highly modular. This allows many diversely motivated people to act for a wide range of reasons that, in combination, cohere into new useful information, knowledge, and cultural goods. These

architectures and organizational models allow both independent creation that coexists and coheres into usable patterns, and interdependent cooperative enterprises in the form of peer-production processes.

Together, these three characteristics suggest that the patterns of social production of information that we are observing in the digitally networked environment are not a fad. They are, rather, a sustainable pattern of human production given the characteristics of the networked information economy. The diversity of human motivation is nothing new. We now have a substantial literature documenting its importance in free and open-source software development projects, from Josh Lerner and Jean Tirole, Rishab Ghosh, Eric Von Hippel and Karim Lakhani, and others. Neither is the public goods nature of information new. What is new are the technological conditions that allow these facts to provide the ingredients of a much larger role in the networked information economy for nonmarket, nonproprietary production to emerge. As long as capitalization and ownership of the physical capital base of this economy remain widely distributed and as long as regulatory policy does not make information inputs artificially expensive, individuals will be able to deploy their own creativity, wisdom, conversational capacities, and connected computers, both independently and in loose interdependent cooperation with others, to create a substantial portion of the information environment we occupy. Moreover, we will be able to do so for whatever reason we choose - through markets or firms to feed and clothe ourselves, or through social relations and open communication with others, to give our lives meaning and context.

Transaction Costs and Efficiency

For purposes of analyzing the political values that are the concern of most of this book, all that is necessary is that we accept that peer production in particular, and nonmarket information production and exchange in general, are sustainable in the networked information economy. Most of the remainder of the book seeks to evaluate why, and to what extent, the presence of a substantial nonmarket, commons-based sector in the information production system is desirable from the perspective of various aspects of freedom and justice. Whether this sector is "efficient" within the meaning of the word in welfare economics is beside the point to most of these considerations. Even a strong commitment to a pragmatic political theory, one that accepts and incorporates into its consideration the limits imposed by material and economic reality, need not aim for "efficient" policy in the welfare sense. It is sufficient that the policy is economically and socially sustainable on its own bottom - in other words, that it does not require constant subsidization at the expense of some other area excluded from the analysis. It is nonetheless worthwhile spending a few pages explaining why, and under what conditions, commons-based peer production, and social production more generally, are not only sustainable but actually efficient ways of organizing information production.

The efficient allocation of two scarce resources and one public good are at stake in the choice between social production - whether it is peer production or independent nonmarket production- and market-based production. Because most of the outputs of these processes are nonrival goods - information, knowledge, and culture - the fact that the social production system releases them freely, without extracting a price for using them, means that it would, all other things being equal, be more efficient for information to be produced on a nonproprietary social model, rather than on a proprietary market model. Indeed, all other things need not even be equal for this to hold. It is enough that the net value of the information produced by commons-based social production processes and released freely for anyone to use as they please is no less than the total value of information produced through property-based systems minus the deadweight loss caused by the above-marginal-cost pricing practices that are the intended result of the intellectual property system.

The two scarce resources are: first, human creativity, time, and attention; and second, the computation and communications resources used in information production and exchange. In both cases, the primary reason to choose among proprietary and nonproprietary strategies, between market-based systems - be they direct market exchange or firm-based hierarchical production - and social systems, are the comparative transaction costs of each, and the extent to which these transaction costs either outweigh the benefits of working through each system, or cause the system to distort the information it generates so as to systematically misallocate resources.

The first thing to recognize is that markets, firms, and social relations are three distinct transactional frameworks. Imagine that I am sitting in a room and need paper for my printer. I could (a) order paper from a store; (b) call the storeroom, if I am in a firm or organization that has one, and ask the clerk to deliver the paper I need; or (c) walk over to a neighbor and borrow some paper. Choice (a) describes the market transactional framework. The store knows I need paper immediately because I am willing to pay for it now. Alternative (b) is an example of the firm as a transactional framework. The paper is in the storeroom because someone in the organization planned that someone else would need paper today, with some probability, and ordered enough to fill that expected need. The clerk in the storeroom gives it to me because that is his job; again, defined by someone who planned to have someone available to deliver paper when someone else in the proper channels of authority says that she needs it. Comparing and improving the efficiency of (a) and (b), respectively, has been a central project in transaction-costs organization theory. We might compare, for example, the costs of taking my call, verifying the credit card information, and sending a delivery truck for my one batch of paper, to the costs of someone planning for the average needs of a group of people like me, who occasionally run out of paper, and stocking a storeroom with enough paper and a clerk to fill our needs in a timely manner. However, notice that (c) is also an alternative transactional framework. I could, rather than incurring the costs of transacting through the market with the local store or of building a firm with sufficient lines of authority to stock and manage the storeroom, pop over to my neighbor and ask for some paper. This would make sense even within an existing firm when, for example, I need two or three pages immediately and do not want to wait for the storeroom clerk to do his rounds, or more generally, if I am working at home and the costs of creating "a firm," stocking a storeroom, and paying a clerk are too high for my neighbors and me. Instead, we develop a set of neighborly social relations, rather than a firm-based organization, to deal with shortfalls during periods when it would be too costly to assure a steady flow of paper from the market - for example, late in the evening, on a weekend, or in a sparsely populated area.

The point is not, of course, to reduce all social relations and human decency to a transaction-costs theory. Too many such straight planks have already been cut from the crooked timber of humanity to make that exercise useful or enlightening. The point is that most of economics internally has been ignoring the social transactional framework as an alternative whose relative efficiency can be accounted for and considered in much the same way as the relative cost advantages of simple markets when compared to the hierarchical organizations that typify much of our economic activity-firms.

A market transaction, in order to be efficient, must be clearly demarcated as to what it includes, so that it can be priced efficiently. That price must then be paid in equally crisply delineated currency. Even if a transaction initially may be declared to involve sale of "an amount reasonably required to produce the required output," for a "customary" price, at some point what was provided and what is owed must be crystallized and fixed for a formal exchange. The crispness is a functional requirement of the price system. It derives from the precision and formality of the medium of exchange - currency - and the ambition to provide refined representations of the comparative value of marginal decisions through denomination in an exchange medium that represents these

incremental value differences. Similarly, managerial hierarchies require a crisp definition of who should be doing what, when, and how, in order to permit the planning and coordination process to be effective.

Social exchange, on the other hand, does not require the same degree of crispness at the margin. As Maurice Godelier put it in *The Enigma of the Gift*, "the mark of the gift between close friends and relatives . . . is not the absence of obligations, it is the absence of 'calculation.' "[/9](9) There are, obviously, elaborate and formally ritualistic systems of social exchange, in both ancient societies and modern. There are common-property regimes that monitor and record calls on the common pool very crisply. However, in many of the common-property regimes, one finds mechanisms of bounding or fairly allocating access to the common pool that more coarsely delineate the entitlements, behaviors, and consequences than is necessary for a proprietary system. In modern market society, where we have money as a formal medium of precise exchange, and where social relations are more fluid than in traditional societies, social exchange certainly occurs as a fuzzier medium. Across many cultures, generosity is understood as imposing a debt of obligation; but none of the precise amount of value given, the precise nature of the debt to be repaid, or the date of repayment need necessarily be specified. Actions enter into a cloud of goodwill or membership, out of which each agent can understand him- or herself as being entitled to a certain flow of dependencies or benefits in exchange for continued cooperative behavior. This may be an ongoing relationship between two people, a small group like a family or group of friends, and up to a general level of generosity among strangers that makes for a decent society. The point is that social exchange does not require defining, for example, "I will lend you my car and help you move these five boxes on Monday, and in exchange you will feed my fish next July," in the same way that the following would: "I will move five boxes on Tuesday for $100, six boxes for $120." This does not mean that social systems are cost free - far from it. They require tremendous investment, acculturation, and maintenance. This is true in this case every bit as much as it is true for markets or states. Once functional, however, social exchanges require less information crispness at the margin.

Both social and market exchange systems require large fixed costs - the setting up of legal institutions and enforcement systems for markets, and creating social networks, norms, and institutions for the social exchange. Once these initial costs have been invested, however, market transactions systematically require a greater degree of precise information about the content of actions, goods, and obligations, and more precision of monitoring and enforcement on a per-transaction basis than do social exchange systems.

This difference between markets and hierarchical organizations, on the one hand, and peer-production processes based on social relations, on the other, is particularly acute in the context of human creative labor - one of the central scarce resources that these systems must allocate in the networked information economy. The levels and focus of individual effort are notoriously hard to specify for pricing or managerial commands, considering all aspects of individual effort and ability - talent, motivation, workload, and focus - as they change in small increments over the span of an individual's full day, let alone months. What we see instead is codification of effort types - a garbage collector, a law professor - that are priced more or less finely. However, we only need to look at the relative homogeneity of law firm starting salaries as compared to the high variability of individual ability and motivation levels of graduating law students to realize that pricing of individual effort can be quite crude. Similarly, these attributes are also difficult to monitor and verify over time, though perhaps not quite as difficult as predicting them *ex ante*. Pricing therefore continues to be a function of relatively crude information about the actual variability among people. More importantly, as aspects of performance that are harder to fully specify in advance or monitor-like creativity over time given the occurrence of new

opportunities to be creative, or implicit know-how - become a more significant aspect of what is valuable about an individual's contribution, market mechanisms become more and more costly to maintain efficiently, and, as a practical matter, simply lose a lot of information.

People have different innate capabilities; personal, social, and educational histories; emotional frameworks; and ongoing lived experiences, which make for immensely diverse associations with, idiosyncratic insights into, and divergent utilization of existing information and cultural inputs at different times and in different contexts. Human creativity is therefore very difficult to standardize and specify in the contracts necessary for either market-cleared or hierarchically organized production. As the weight of human intellectual effort increases in the overall mix of inputs into a given production process, an organization model that does not require contractual specification of the individual effort required to participate in a collective enterprise, and which allows individuals to self-identify for tasks, will be better at gathering and utilizing information about who should be doing what than a system that does require such specification. Some firms try to solve this problem by utilizing market- and social-relations-oriented hybrids, like incentive compensation schemes and employee-of-the-month-type social motivational frameworks. These may be able to improve on firm-only or market-only approaches. It is unclear, though, how well they can overcome the core difficulty: that is, that both markets and firm hierarchies require significant specification of the object of organization and pricing - in this case, human intellectual input. The point here is qualitative. It is not only, or even primarily, that more people can participate in production in a commons-based effort. It is that the widely distributed model of information production will better identify the best person to produce a specific component of a project, considering all abilities and availability to work on the specific module within a specific time frame. With enough uncertainty as to the value of various productive activities, and enough variability in the quality of both information inputs and human creative talent vis-à-vis any set of production opportunities, freedom of action for individuals coupled with continuous communications among the pool of potential producers and consumers can generate better information about the most valuable productive actions, and the best human inputs available to engage in these actions at a given time. Markets and firm incentive schemes are aimed at producing precisely this form of self-identification. However, the rigidities associated with collecting and comprehending bids from individuals through these systems (that is, transaction costs) limit the efficacy of self-identification by comparison to a system in which, once an individual self-identifies for a task, he or she can then undertake it without permission, contract, or instruction from another. The emergence of networked organizations (described and analyzed in the work of Charles Sabel and others) suggests that firms are in fact trying to overcome these limitations by developing parallels to the freedom to learn, innovate, and act on these innovations that is intrinsic to peer-production processes by loosening the managerial bonds, locating more of the conception and execution of problem solving away from the managerial core of the firm, and implementing these through social, as well as monetary, motivations. However, the need to assure that the value created is captured within the organization limits the extent to which these strategies can be implemented within a single enterprise, as opposed to their implementation in an open process of social production. This effect, in turn, is in some sectors attenuated through the use of what Walter Powell and others have described as learning networks. Engineers and scientists often create frameworks that allow them to step out of their organizational affiliations, through conferences or workshops. By reproducing the social production characteristics of academic exchange, they overcome some of the information loss caused by the boundary of the firm. While these organizational strategies attenuate the problem, they also underscore the degree to which it is widespread and understood by organizations as such. The fact that the direction of the solutions business organizations choose tends to shift elements of the production process away from market- or firm-based models and toward networked social production models is revealing. Now, the self-identification that is central to the relative

information efficiency of peer production is not always perfect. Some mechanisms used by firms and markets to codify effort levels and abilities - like formal credentials - are the result of experience with substantial errors or misstatements by individuals of their capacities. To succeed, therefore, peer-production systems must also incorporate mechanisms for smoothing out incorrect self-assessments - as peer review does in traditional academic research or in the major sites like *Wikipedia* or Slashdot, or as redundancy and statistical averaging do in the case of NASA clickworkers. The prevalence of misperceptions that individual contributors have about their own ability and the cost of eliminating such errors will be part of the transaction costs associated with this form of organization. They parallel quality control problems faced by firms and markets.

The lack of crisp specification of who is giving what to whom, and in exchange for what, also bears on the comparative transaction costs associated with the allocation of the second major type of scarce resource in the networked information economy: the physical resources that make up the networked information environment - communications, computation, and storage capacity. It is important to note, however, that these are very different from creativity and information as inputs: they are private goods, not a public good like information, and they are standardized goods with well-specified capacities, not heterogeneous and highly uncertain attributes like human creativity at a given moment and context. Their outputs, unlike information, are not public goods. The reasons that they are nonetheless subject to efficient sharing in the networked environment therefore require a different economic explanation. However, the sharing of these material resources, like the sharing of human creativity, insight, and attention, nonetheless relies on both the comparative transaction costs of markets and social relations and the diversity of human motivation.

Personal computers, wireless transceivers, and Internet connections are "shareable goods." The basic intuition behind the concept of shareable goods is simple. There are goods that are "lumpy": given a state of technology, they can only be produced in certain discrete bundles that offer discontinuous amounts of functionality or capacity. In order to have any ability to run a computation, for example, a consumer must buy a computer processor. These, in turn, only come in discrete units with a certain speed or capacity. One could easily imagine a world where computers are very large and their owners sell computation capacity to consumers "on demand," whenever they needed to run an application. That is basically the way the mainframe world of the 1960s and 1970s worked. However, the economics of microchip fabrication and of network connections over the past thirty years, followed by storage technology, have changed that. For most functions that users need, the price-performance trade-off favors stand-alone, general-purpose personal computers, owned by individuals and capable of running locally most applications users want, over remote facilities capable of selling on-demand computation and storage. So computation and storage today come in discrete, lumpy units. You can decide to buy a faster or slower chip, or a larger or smaller hard drive, but once you buy them, you have the capacity of these machines at your disposal, whether you need it or not.

Lumpy goods can, in turn, be fine-, medium-, or large-grained. A large-grained good is one that is so expensive it can only be used by aggregating demand for it. Industrial capital equipment, like a steam engine, is of this type. Fine-grained goods are of a granularity that allows consumers to buy precisely as much of the goods needed for the amount of capacity they require. Medium-grained goods are small enough for an individual to justify buying for her own use, given their price and her willingness and ability to pay for the functionality she plans to use. A personal computer is a medium-grained lumpy good in the advanced economies and among the more well-to-do in poorer countries, but is a large-grained capital good for most people in poor countries. If, given the price of such a good and the wealth of a society, a large number of individuals buy and use such medium-grained lumpy goods, that society will have a large amount of excess capacity "out there,"

in the hands of individuals. Because these machines are put into service to serve the needs of individuals, their excess capacity is available for these individuals to use as they wish - for their own uses, to sell to others, or to share with others. It is the combination of the fact that these machines are available at prices (relative to wealth) that allow users to put them in service based purely on their value for personal use, and the fact that they have enough capacity to facilitate additionally the action and fulfill the needs of others, that makes them "shareable." If they were so expensive that they could only be bought by pooling the value of a number of users, they would be placed in service either using some market mechanism to aggregate that demand, or through formal arrangements of common ownership by all those whose demand was combined to invest in purchasing the resource. If they were so finely grained in their capacity that there would be nothing left to share, again, sharing would be harder to sustain. The fact that they are both relatively inexpensive and have excess capacity makes them the basis for a stable model of individual ownership of resources combined with social sharing of that excess capacity.

Because social sharing requires less precise specification of the transactional details with each transaction, it has a distinct advantage over market-based mechanisms for reallocating the excess capacity of shareable goods, particularly when they have small quanta of excess capacity relative to the amount necessary to achieve the desired outcome. For example, imagine that there are one thousand people in a population of computer owners. Imagine that each computer is capable of performing one hundred computations per second, and that each computer owner needs to perform about eighty operations per second. Every owner, in other words, has twenty operations of excess capacity every second. Now imagine that the marginal transaction costs of arranging a sale of these twenty operations - exchanging PayPal (a widely used low-cost Internet-based payment system) account information, insurance against nonpayment, specific statement of how much time the computer can be used, and so forth - cost ten cents more than the marginal transaction costs of sharing the excess capacity socially. John wants to render a photograph in one second, which takes two hundred operations per second. Robert wants to model the folding of proteins, which takes ten thousand operations per second. For John, a sharing system would save fifty cents - assuming he can use his own computer for half of the two hundred operations he needs. He needs to transact with five other users to "rent" their excess capacity of twenty operations each. Robert, on the other hand, needs to transact with five hundred individual owners in order to use their excess capacity, and for him, using a sharing system is fifty dollars cheaper. The point of the illustration is simple. The cost advantage of sharing as a transactional framework relative to the price system increases linearly with the number of transactions necessary to acquire the level of resources necessary for an operation. If excess capacity in a society is very widely distributed in small dollops, and for any given use of the excess capacity it is necessary to pool the excess capacity of thousands or even millions of individual users, the transaction-cost advantages of the sharing system become significant.

The transaction-cost effect is reinforced by the motivation crowding out theory. When many discrete chunks of excess capacity need to be pooled, each distinct contributor cannot be paid a very large amount. Motivation crowding out theory would predict that when the monetary rewards to an activity are low, the negative effect of crowding out the social-psychological motivation will weigh more heavily than any increased incentive that is created by the promise of a small payment to transfer one's excess capacity. The upshot is that when the technological state results in excess capacity of physical capital being widely distributed in small dollops, social sharing can outperform secondary markets as a mechanism for harnessing that excess capacity. This is so because of both transaction costs and motivation. Fewer owners will be willing to sell their excess capacity cheaply than to give it away for free in the right social context and the transaction costs of selling will be higher than those of sharing.

From an efficiency perspective, then, there are clear reasons to think that social production systems - both peer production of information, knowledge, and culture and sharing of material resources - can be more efficient than market-based systems to motivate and allocate both human creative effort and the excess computation, storage, and communications capacity that typify the networked information economy. That does not mean that all of us will move out of market-based productive relationships all of the time. It does mean that alongside our market-based behaviors we generate substantial amounts of human creativity and mechanical capacity. The transaction costs of clearing those resources through the price system or through firms are substantial, and considerably larger for the marginal transaction than clearing them through social-sharing mechanisms as a transactional framework. With the right institutional framework and peer-review or quality-control mechanisms, and with well-modularized organization of work, social sharing is likely to identify the best person available for a job and make it feasible for that person to work on that job using freely available information inputs. Similarly, social transactional frameworks are likely to be substantially less expensive than market transactions for pooling large numbers of discrete, small increments of the excess capacity of the personal computer processors, hard drives, and network connections that make up the physical capital base of the networked information economy. In both cases, given that much of what is shared is excess capacity from the perspective of the contributors, available to them after they have fulfilled some threshold level of their market-based consumption requirements, social-sharing systems are likely to tap in to social psychological motivations that money cannot tap, and, indeed, that the presence of money in a transactional framework could nullify. Because of these effects, social sharing and collaboration can provide not only a sustainable alternative to market-based and firm-based models of provisioning information, knowledge, culture, and communications, but also an alternative that more efficiently utilizes the human and physical capital base of the networked information economy. A society whose institutional ecology permitted social production to thrive would be more productive under these conditions than a society that optimized its institutional environment solely for market- and firm-based production, ignoring its detrimental effects to social production.

The Emergence of Social Production in the Digitally Networked Environment

There is a curious congruence between the anthropologists of the gift and mainstream economists today. Both treat the gift literature as being about the periphery, about societies starkly different from modern capitalist societies. As Godelier puts it, "What a contrast between these types of society, these social and mental universes, and today's capitalist society where the majority of social relations are impersonal (involving the individual as citizen and the state, for instance), and where the exchange of things and services is conducted for the most part in an anonymous marketplace, leaving little room for an economy and moral code based on gift-giving."[/10] And yet, sharing is everywhere around us in the advanced economies. Since the 1980s, we have seen an increasing focus, in a number of literatures, on production practices that rely heavily on social rather than price-based or governmental policies. These include, initially, the literature on social norms and social capital, or trust.[/11] Both these lines of literature, however, are statements of the institutional role of social mechanisms for enabling market exchange and production. More direct observations of social production and exchange systems are provided by the literature on social provisioning of public goods - like social norm enforcement as a dimension of policing criminality, and the literature on common property regimes.[/12] The former are limited by their focus on public goods provisioning. The latter are usually limited by their focus on discretely identifiable types of resources-common pool resources - that must be managed as among a group of claimants while retaining a proprietary outer boundary toward nonmembers. The focus of those who study these phenomena is usually on relatively small and tightly knit communities, with clear boundaries between members and nonmembers.[/13]

These lines of literature point to an emerging understanding of social production and exchange as an alternative to markets and firms. Social production is not limited to public goods, to exotic, out-of-the-way places like surviving medieval Spanish irrigation regions or the shores of Maine's lobster fishing grounds, or even to the ubiquitous phenomenon of the household. As SETI@home and Slashdot suggest, it is not necessarily limited to stable communities of individuals who interact often and know each other, or who expect to continue to interact personally. Social production of goods and services, both public and private, is ubiquitous, though unnoticed. It sometimes substitutes for, and sometimes complements, market and state production everywhere. It is, to be fanciful, the dark matter of our economic production universe.

Consider the way in which the following sentences are intuitively familiar, yet as a practical matter, describe the provisioning of goods or services that have well-defined NAICS categories (the categories used by the Economic Census to categorize economic sectors) whose provisioning through the markets is accounted for in the Economic Census, but that are commonly provisioned in a form consistent with the definition of sharing - on a radically distributed model, without price or command.

> NAICS 624410624410 [Babysitting services, child day care]
>
>> "John, could you pick up Bobby today when you take Lauren to soccer? I have a conference call I have to make."
>>
>> "Are you doing homework with Zoe today, or shall I?"
>
> NAICS 484210 [Trucking used household, office, or institutional furniture and equipment]
>
>> "Jane, could you lend a hand moving this table to the dining room?"
>>
>> "Here, let me hold the elevator door for you, this looks heavy."
>
> NAICS 484122 [Trucking, general freight, long-distance, less-than-truckload]
>
>> "Jack, do you mind if I load my box of books in your trunk so you can drop it off at my brother's on your way to Boston?"
>
> NAICS 514110 [Traffic reporting services]
>
>> "Oh, don't take I-95, it's got horrible construction traffic to exit 39."
>
> NAICS 711510 [Newspaper columnists, independent (freelance)]
>
>> "I don't know about Kerry, he doesn't move me, I think he should be more aggressive in criticizing Bush on Iraq."
>
> NAICS 621610 [Home health-care services]
>
>> "Can you please get me my medicine? I'm too wiped to get up."
>>
>> "Would you like a cup of tea?"
>
> NAICS 561591 [Tourist information bureaus]
>
>> "Excuse me, how do I get to Carnegie Hall?"
>
> NAICS 561321 [Temporary help services]
>
>> "I've got a real crunch on the farm, can you come over on Saturday and lend a hand?"

> "This is crazy, I've got to get this document out tonight, could you lend me a hand with proofing and pulling it all together tonight?"

NAICS 71 [Arts, entertainment, and recreation]

> "Did you hear the one about the Buddhist monk, the Rabbi, and the Catholic priest . . . ?"

> "Roger, bring out your guitar. . . . "

> "Anybody up for a game of . . . ?"

The litany of examples generalizes through a combination of four dimensions that require an expansion from the current focus of the literatures related to social production. First, they relate to production of goods and services, not only of norms or rules. Social relations provide the very motivations for, and information relating to, production and exchange, not only the institutional framework for organizing action, which itself is motivated, informed, and organized by markets or managerial commands. Second, they relate to all kinds of goods, not only public goods. In particular, the paradigm cases of free software development and distributed computing involve labor and shareable goods - each plainly utilizing private goods as inputs, and, in the case of distributed computing, producing private goods as outputs. Third, at least some of them relate not only to relations of production within well-defined communities of individuals who have repeated interactions, but extend to cover baseline standards of human decency. These enable strangers to ask one another for the time or for directions, enable drivers to cede the road to each other, and enable strangers to collaborate on software projects, on coauthoring an online encyclopedia, or on running simulations of how proteins fold. Fourth, they may either complement or substitute for market and state production systems, depending on the social construction of mixed provisioning. It is hard to measure the weight that social and sharing-based production has in the economy. Our intuitions about capillary systems would suggest that the total volume of boxes or books moved or lifted, instructions given, news relayed, and meals prepared by family, friends, neighbors, and minimally decent strangers would be very high relative to the amount of substitutable activity carried on through market exchanges or state provisioning.

Why do we, despite the ubiquity of social production, generally ignore it as an economic phenomenon, and why might we now reconsider its importance? A threshold requirement for social sharing to be a modality of economic production, as opposed to one purely of social reproduction, is that sharing-based action be effective. Efficacy of individual action depends on the physical capital requirements for action to become materially effective, which, in turn, depend on technology. Effective action may have very low physical capital requirements, so that every individual has, by natural capacity, "the physical capital" necessary for action. Social production or sharing can then be ubiquitous (though in practice, it may not). Vocal cords to participate in a sing-along or muscles to lift a box are obvious examples. When the capital requirements are nontrivial, but the capital good is widely distributed and available, sharing can similarly be ubiquitous and effective. This is true both when the shared resource or good is the capacity of the capital good itself - as in the case of shareable goods - and when some widely distributed human capacity is made effective through the use of the widely distributed capital goods - as in the case of human creativity, judgment, experience, and labor shared in online peer-production processes - in which participants contribute using the widespread availability of connected computers. When use of larger-scale physical capital goods is a threshold requirement of effective action, we should not expect to see widespread reliance on decentralized sharing as a standard modality of production. Industrial mass-manufacture of automobiles, steel, or plastic toys, for example, is not the sort of thing that is likely to be produced on a social-sharing basis, because of the capital constraints. This is not to say that even for large-scale capital projects, like irrigation systems and dams, social production systems

cannot step into the breach. We have those core examples in the common-property regime literature, and we have worker-owned firms as examples of mixed systems. However, those systems tend to replicate the characteristics of firm, state, or market production - using various combinations of quotas, scrip systems, formal policing by "professional" officers, or management within worker-owned firms. By comparison, the "common property" arrangements described among lobster gangs of Maine or fishing groups in Japan, where capital requirements are much lower, tend to be more social-relations-based systems, with less formalized or crisp measurement of contributions to, and calls on, the production system.

To say that sharing is technology dependent is not to deny that it is a ubiquitous human phenomenon. Sharing is so deeply engrained in so many of our cultures that it would be difficult to argue that with the "right" (or perhaps "wrong") technological contingencies, it would simply disappear. My claim, however, is narrower. It is that the relative economic role of sharing changes with technology. There are technological conditions that require more or less capital, in larger or smaller packets, for effective provisioning of goods, services, and resources the people value. As these conditions change, the relative scope for social-sharing practices to play a role in production changes. When goods, services, and resources are widely dispersed, their owners can choose to engage with each other through social sharing instead of through markets or a formal, state-based relationship, because individuals have available to them the resources necessary to engage in such behavior without recourse to capital markets or the taxation power of the state. If technological changes make the resources necessary for effective action rare or expensive, individuals may wish to interact in social relations, but they can now only do so ineffectively, or in different fields of endeavor that do not similarly require high capitalization. Large-packet, expensive physical capital draws the behavior into one or the other of the modalities of production that can collect the necessary financial capital - through markets or taxation. Nothing, however, prevents change from happening in the opposite direction. Goods, services, and resources that, in the industrial stage of the information economy required large-scale, concentrated capital investment to provision, are now subject to a changing technological environment that can make sharing a better way of achieving the same results than can states, markets, or their hybrid, regulated industries.

Because of changes in the technology of the industrial base of the most advanced economies, social sharing and exchange is becoming a common modality of production at their very core - in the information, culture, education, computation, and communications sectors. Free software, distributed computing, ad hoc mesh wireless networks, and other forms of peer production offer clear examples of large-scale, measurably effective sharing practices. The highly distributed capital structure of contemporary communications and computation systems is largely responsible for this increased salience of social sharing as a modality of economic production in that environment. By lowering the capital costs required for effective individual action, these technologies have allowed various provisioning problems to be structured in forms amenable to decentralized production based on social relations, rather than through markets or hierarchies.

My claim is not, of course, that we live in a unique moment of humanistic sharing. It is, rather, that our own moment in history suggests a more general observation. The technological state of a society, in particular the extent to which individual agents can engage in efficacious production activities with material resources under their individual control, affects the opportunities for, and hence the comparative prevalence and salience of, social, market - both price-based and managerial - and state production modalities. The capital cost of effective economic action in the industrial economy shunted sharing to its economic peripheries - to households in the advanced economies, and to the global economic peripheries that have been the subject of the anthropology of gift or the common-property regime literatures. The emerging restructuring of capital investment in digital

networks - in particular, the phenomenon of user-capitalized computation and communications capabilities - are at least partly reversing that effect. Technology does not determine the level of sharing. It does, however, set threshold constraints on the effective domain of sharing as a modality of economic production. Within the domain of the practically feasible, the actual level of sharing practices will be culturally driven and cross-culturally diverse.

Most practices of production - social or market-based - are already embedded in a given technological context. They present no visible "problem" to solve or policy choice to make. We do not need to be focused consciously on improving the conditions under which friends lend a hand to each other to move boxes, make dinner, or take kids to school. We feel no need to reconsider the appropriateness of market-based firms as the primary modality for the production of automobiles. However, in moments where a field of action is undergoing a technological transition that changes the opportunities for sharing as a modality of production, understanding that sharing *is* a modality of production becomes more important, as does understanding how it functions as such. This is so, as we are seeing today, when prior technologies have already set up market- or state-based production systems that have the law and policy-making systems already designed to fit their requirements. While the prior arrangement may have been the most efficient, or even may have been absolutely necessary for the incumbent production system, its extension under new technological conditions may undermine, rather than improve, the capacity of a society to produce and provision the goods, resources, or capacities that are the object of policy analysis. This is, as I discuss in part III, true of wireless communications regulation, or "spectrum management," as it is usually called; of the regulation of information, knowledge, and cultural production, or "intellectual property," as it is usually now called; and it may be true of policies for computation and wired communications networks, as distributed computing and the emerging peer-to-peer architectures suggest.

The Interface of Social Production and Market-Based Businesses

The rise of social production does not entail a decline in market-based production. Social production first and foremost harnesses impulses, time, and resources that, in the industrial information economy, would have been wasted or used purely for consumption. Its immediate effect is therefore likely to increase overall productivity in the sectors where it is effective. But that does not mean that its effect on market-based enterprises is neutral. A newly effective form of social behavior, coupled with a cultural shift in tastes as well as the development of new technological and social solution spaces to problems that were once solved through market-based firms, exercises a significant force on the shape and conditions of market action. Understanding the threats that these developments pose to some incumbents explains much of the political economy of law in this area, which will occupy chapter 11. At the simplest level, social production in general and peer production in particular present new sources of competition to incumbents that produce information goods for which there are now socially produced substitutes. Open source software development, for example, first received mainstream media attention in 1998 due to publication of a leaked internal memorandum from Microsoft, which came to be known as The Halloween Memo. In it, a Microsoft strategist identified the open source methodology as the one major potential threat to the company's dominance over the desktop. As we have seen since, definitively in the Web server market and gradually in segments of the operating system market, this prediction proved prescient. Similarly, *Wikipedia* now presents a source of competition to online encyclopedias like *Columbia*, *Grolier*, or *Encarta*, and may well come to be seen as an adequate substitute for *Britannica* as well. Most publicly visible, peer-to-peer file sharing networks have come to compete with the recording industry as an alternative music distribution system, to the point where the long-term existence of that industry is in question. Some scholars like William Fisher, and artists like

Jenny Toomey and participants in the Future of Music Coalition, are already looking for alternative ways of securing for artists a living from the music they make.

The competitive threat from social production, however, is merely a surface phenomenon. Businesses often face competition or its potential, and this is a new source, with new economics, which may or may not put some of the incumbents out of business. But there is nothing new about entrants with new business models putting slow incumbents out of business. More basic is the change in opportunity spaces, the relationships of firms to users, and, indeed, the very nature of the boundary of the firm that those businesses that are already adapting to the presence and predicted persistence of social production are exhibiting. Understanding the opportunities social production presents for businesses begins to outline how a stable social production system can coexist and develop a mutually reinforcing relationship with market-based organizations that adapt to and adopt, instead of fight, them.

Consider the example I presented in chapter 2 of IBM's relationship to the free and open source software development community. IBM, as I explained there, has shown more than $2 billion a year in "Linux-related revenues." Prior to IBM's commitment to adapting to what the firm sees as the inevitability of free and open source software, the company either developed in house or bought from external vendors the software it needed as part of its hardware business, on the one hand, and its software services - customization, enterprise solutions, and so forth - on the other hand. In each case, the software development follows a well-recognized supply chain model. Through either an employment contract or a supply contract the company secures a legal right to require either an employee or a vendor to deliver a given output at a given time. In reliance on that notion of a supply chain that is fixed or determined by a contract, the company turns around and promises to its clients that it will deliver the integrated product or service that includes the contracted-for component. With free or open source software, that relationship changes. IBM is effectively relying for its inputs on a loosely defined cloud of people who are engaged in productive social relations. It is making the judgment that the probability that a sufficiently good product will emerge out of this cloud is high enough that it can undertake a contractual obligation to its clients, even though no one in the cloud is specifically contractually committed to it to produce the specific inputs the firm needs in the timeframe it needs it. This apparent shift from a contractually deterministic supply chain to a probabilistic supply chain is less dramatic, however, than it seems. Even when contracts are signed with employees or suppliers, they merely provide a probability that the employee or the supplier will in fact supply in time and at appropriate quality, given the difficulties of coordination and implementation. A broad literature in organization theory has developed around the effort to map the various strategies of collaboration and control intended to improve the likelihood that the different components of the production process will deliver what they are supposed to: from early efforts at vertical integration, to relational contracting, pragmatic collaboration, or Toyota's fabled flexible specialization. The presence of a formalized enforceable contract, for outputs in which the supplier can claim and transfer a property right, may change the probability of the desired outcome, but not the fact that in entering its own contract with its clients, the company is making a prediction about the required availability of necessary inputs in time. When the company turns instead to the cloud of social production for its inputs, it is making a similar prediction. And, as with more engaged forms of relational contracting, pragmatic collaborations, or other models of iterated relations with coproducers, the company may engage with the social process in order to improve the probability that the required inputs will in fact be produced in time. In the case of companies like IBM or Red Hat, this means, at least partly, paying employees to participate in the open source development projects. But managing this relationship is tricky. The firms must do so without seeking to, or even seeming to seek to, take over the project; for to take over the project in order to steer it more "predictably" toward the firm's needs is to kill the goose that lays the golden

eggs. For IBM and more recently Nokia, supporting the social processes on which they rely has also meant contributing hundreds of patents to the Free Software Foundation, or openly licensing them to the software development community, so as to extend the protective umbrella created by these patents against suits by competitors. As the companies that adopt this strategic reorientation become more integrated into the peer-production process itself, the boundary of the firm becomes more porous. Participation in the discussions and governance of open source development projects creates new ambiguity as to where, in relation to what is "inside" and "outside" of the firm boundary, the social process is. In some cases, a firm may begin to provide utilities or platforms for the users whose outputs it then uses in its own products. The Open Source Development Group (OSDG), for example, provides platforms for Slashdot and SourceForge. In these cases, the notion that there are discrete "suppliers" and "consumers," and that each of these is clearly demarcated from the other and outside of the set of stable relations that form the inside of the firm becomes somewhat attenuated.

As firms have begun to experience these newly ambiguous relationships with individuals and social groups, they have come to wrestle with questions of leadership and coexistence. Businesses like IBM, or eBay, which uses peer production as a critical component of its business ecology - the peer reviewed system of creating trustworthiness, without which person-to-person transactions among individual strangers at a distance would be impossible - have to structure their relationship to the peer-production processes that they co-exist with in a helpful and non-threatening way. Sometimes, as we saw in the case of IBM's contributions to the social process, this may mean support without attempting to assume "leadership" of the project. Sometimes, as when peer production is integrated more directly into what is otherwise a commercially created and owned platform - as in the case of eBay - the relationship is more like that of a peer-production leader than of a commercial actor. Here, the critical and difficult point for business managers to accept is that bringing the peer-production community into the newly semi-porous boundary of the firm - taking those who used to be customers and turning them into participants in a process of coproduction - changes the relationship of the firm's managers and its users. Linden Labs, which runs Second Life, learned this in the context of the tax revolt described in chapter 3. Users cannot be ordered around like employees. Nor can they be simply advertised - to and manipulated, or even passively surveyed, like customers. To do that would be to lose the creative and generative social character that makes integration of peer production into a commercial business model so valuable for those businesses that adopt it. Instead, managers must be able to identify patterns that emerge in the community and inspire trust that they are correctly judging the patterns that are valuable from the perspective of the users, not only the enterprise, so that the users in fact coalesce around and extend these patterns.

The other quite basic change wrought by the emergence of social production, from the perspective of businesses, is a change in taste. Active users require and value new and different things than passive consumers did. The industrial information economy specialized in producing finished goods, like movies or music, to be consumed passively, and well-behaved appliances, like televisions, whose use was fully specified at the factory door. The emerging businesses of the networked information economy are focusing on serving the demand of active users for platforms and tools that are much more loosely designed, late-binding - that is, optimized only at the moment of use and not in advance - variable in their uses, and oriented toward providing users with new, flexible platforms for relationships. Personal computers, camera phones, audio and video editing software, and similar utilities are examples of tools whose value increases for users as they are enabled to explore new ways to be creative and productively engaged with others. In the network, we are beginning to see business models emerge to allow people to come together, like MeetUp, and to share annotations of Web pages they read, like del.icio.us, or photographs they took, like Flickr.

Services like Blogger and Technorati similarly provide platforms for the new social and cultural practices of personal journals, or the new modes of expression described in chapters 7 and 8.

The overarching point is that social production is reshaping the market conditions under which businesses operate. To some of the incumbents of the industrial information economy, the pressure from social production is experienced as pure threat. It is the clash between these incumbents and the new practices that was most widely reported in the media in the first five years of the twenty-first century, and that has driven much of policy making, legislation, and litigation in this area. But the much more fundamental effect on the business environment is that social production is changing the relationship of firms to individuals outside of them, and through this changing the strategies that firms internally are exploring. It is creating new sources of inputs, and new tastes and opportunities for outputs. Consumers are changing into users - more active and productive than the consumers of the industrial information economy. The change is reshaping the relationships necessary for business success, requiring closer integration of users into the process of production, both in inputs and outputs. It requires different leadership talents and foci. By the time of this writing, in 2005, these new opportunities and adaptations have begun to be seized upon as strategic advantages by some of the most successful companies working around the Internet and information technology, and increasingly now around information and cultural production more generally. Eric von Hippel's work has shown how the model of user innovation has been integrated into the business model of innovative firms even in sectors far removed from either the network or from information production - like designing kite-surfing equipment or mountain bikes. As businesses begin to do this, the platforms and tools for collaboration improve, the opportunities and salience of social production increases, and the political economy begins to shift. And as these firms and social processes coevolve, the dynamic accommodation they are developing provides us with an image of what the future stable interface between market-based businesses and the newly salient social production is likely to look like.

Notes

1. Richard M. Titmuss, *The Gift Relationship: From Human Blood to Social Policy* (New York: Vintage Books, 1971), 94.

2. Kenneth J. Arrow, "Gifts and Exchanges," *Philosophy & Public Affairs* 1 (1972): 343.

3. Bruno S. Frey, *Not Just for Money: An Economic Theory of Personal Motivation* (Brookfield, VT: Edward Elgar, 1997); Bruno S. Frey, *Inspiring Economics: Human Motivation in Political Economy* (Northampton, MA: Edward Elgar, 2001), 52-72. An excellent survey of this literature is Bruno S. Frey and Reto Jegen, "Motivation Crowding Theory," *Journal of Economic Surveys* 15, no. 5 (2001): 589. For a crystallization of the underlying psychological theory, see Edward L. Deci and Richard M. Ryan, *Intrinsic Motivation and Self-Determination in Human Behavior* (New York: Plenum, 1985).

4. Roland Bénabou and Jean Tirole, "Self-Confidence and Social Interactions" (working paper no. 7585, National Bureau of Economic Research, Cambridge, MA, March 2000).

5. Truman F. Bewley, "A Depressed Labor Market as Explained by Participants," *American Economic Review (Papers and Proceedings)* 85 (1995): 250, provides survey data about managers' beliefs about the effects of incentive contracts; Margit Osterloh and Bruno S. Frey, "Motivation, Knowledge Transfer, and Organizational Form," *Organization Science* 11 (2000): 538, provides evidence that employees with tacit knowledge communicate it to coworkers more efficiently without extrinsic

motivations, with the appropriate social motivations, than when money is offered for "teaching" their knowledge; Bruno S. Frey and Felix Oberholzer-Gee, "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review* 87 (1997): 746; and Howard Kunreuther and Douslar Easterling, "Are Risk-Benefit Tradeoffs Possible in Siting Hazardous Facilities?" *American Economic Review (Papers and Proceedings)* 80 (1990): 252-286, describe empirical studies where communities became less willing to accept undesirable public facilities (Not in My Back Yard or NIMBY) when offered compensation, relative to when the arguments made were policy based on the common weal; Uri Gneezy and Aldo Rustichini, "A Fine Is a Price," *Journal of Legal Studies* 29 (2000): 1, found that introducing a fine for tardy pickup of kindergarten kids increased, rather than decreased, the tardiness of parents, and once the sense of social obligation was lost to the sense that it was "merely" a transaction, the parents continued to be late at pickup, even after the fine was removed.

6. James S. Coleman, "Social Capital in the Creation of Human Capital," *American Journal of Sociology* 94, supplement (1988): S95, S108. For important early contributions to this literature, see Mark Granovetter, "The Strength of Weak Ties," *American Journal of Sociology* 78 (1973): 1360; Mark Granovetter, *Getting a Job: A Study of Contacts and Careers* (Cambridge, MA: Harvard University Press, 1974); Yoram Ben-Porath, "The F-Connection: Families, Friends and Firms and the Organization of Exchange," *Population and Development Review* 6 (1980): 1.

7. Nan Lin, *Social Capital: A Theory of Social Structure and Action* (New York: Cambridge University Press, 2001), 150-151.

8. Steve Weber, *The Success of Open Source* (Cambridge, MA: Harvard University Press, 2004).

9. Maurice Godelier, *The Enigma of the Gift*, trans. Nora Scott (Chicago: University of Chicago Press, 1999), 5.

10. Godelier, *The Enigma*, 106.

11. In the legal literature, Robert Ellickson, *Order Without Law: How Neighbors Settle Disputes* (Cambridge, MA: Harvard University Press, 1991), is the locus classicus for showing how social norms can substitute for law. For a bibliography of the social norms literature outside of law, see Richard H. McAdams, "The Origin, Development, and Regulation of Norms," *Michigan Law Review* 96 (1997): 338n1, 339n2. Early contributions were: Edna Ullman-Margalit, *The Emergence of Norms* (Oxford: Clarendon Press, 1977); James Coleman, "Norms as Social Capital," in *Economic Imperialism: The Economic Approach Applied Outside the Field of Economics*, ed. Peter Bernholz and Gerard Radnitsky (New York: Paragon House Publishers, 1987), 133-155; Sally E. Merry, "Rethinking Gossip and Scandal," in *Toward a Theory of Social Control, Fundamentals*, ed. Donald Black (New York: Academic Press, 1984).

12. On policing, see Robert C. Ellickson, "Controlling Chronic Misconduct in City Spaces: Of Panhandlers, Skid Rows, and Public-Space Zoning," *Yale Law Journal* 105 (1996): 1165, 1194-1202; and Dan M. Kahan, "Between Economics and Sociology: The New Path of Deterrence," *Michigan Law Review* 95 (1997): 2477.

13. An early and broad claim in the name of commons in resources for communication and transportation, as well as human community building-like roads, canals, or social-gathering places-is Carol Rose, "The Comedy of the Commons: Custom, Commerce, and Inherently Public Property," *University Chicago Law Review* 53 (1986): 711. Condensing around the work of Elinor Ostrom, a more

narrowly defined literature developed over the course of the 1990s: Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (New York: Cambridge University Press, 1990). Another seminal study was James M. Acheson, *The Lobster Gangs of Maine* (New Hampshire: University Press of New England, 1988). A brief intellectual history of the study of common resource pools and common property regimes can be found in Charlotte Hess and Elinor Ostrom, "Ideas, Artifacts, Facilities, and Content: Information as a Common-Pool Resource," *Law & Contemporary Problems* 66 (2003): 111.

The Wealth of Networks:
How Social Production Transforms Markets and Freedom
by Yochai Benkler, Yale University Press

Part II
The Networked Information Economy

Epigraph

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Part II
The Political Economy of Property and Commons

How a society produces its information environment goes to the very core of freedom. Who gets to say what, to whom? What is the state of the world? What counts as credible information? How will different forms of action affect the way the world can become? These questions go to the foundations of effective human action. They determine what individuals understand to be the range of options open to them, and the range of consequences to their actions. They determine what is understood to be open for debate in a society, and what is considered impossible as a collective goal or a collective path for action. They determine whose views count toward collective action, and whose views are lost and never introduced into the debate of what we should do as political entities or social communities. Freedom depends on the information environment that those individuals and societies occupy. Information underlies the very possibility of individual self-direction. Information and communication constitute the practices that enable a community to form a common range of understandings of what is at stake and what paths are open for the taking. They are constitutive components of both formal and informal mechanisms for deciding on collective action. Societies that embed the emerging networked information economy in an institutional ecology that accommodates nonmarket production, both individual and cooperative, will improve the freedom of their constituents along all these dimensions.

The networked information economy makes individuals better able to do things for and by themselves, and makes them less susceptible to manipulation by others than they were in the mass-media culture. In this sense, the emergence of this new set of technical, economic, social, and institutional relations can increase the relative role that each individual is able to play in authoring his or her own life. The networked information economy also promises to provide a much more robust platform for public debate. It enables citizens to participate in public conversation continuously and pervasively, not as passive recipients of "received wisdom" from professional talking heads, but as active participants in conversations carried out at many levels of political and social structure. Individuals can find out more about what goes on in the world, and share it more effectively with others. They can check the claims of others and produce their own, and they can be heard by others, both those who are like-minded and opponents. At a more foundational level of collective understanding, the shift from an industrial to a networked information economy increases the extent to which individuals can become active participants in producing their own cultural environment. It opens the possibility of a more critical and reflective culture.

Unlike the relationship of information production to freedom, the relationship between the organization of information production and distributive justice is not intrinsic. However, the importance of knowledge in contemporary economic production makes a change in the modality of information production important to justice as well. The networked information economy can provide opportunities for global development and for improvements in the justice of distribution of opportunities and capacities everywhere. Economic opportunity and welfare today - of an individual, a social group, or a nation - depend on the state of knowledge and access to opportunities to learn and apply practical knowledge. Transportation networks, global financial markets, and institutional trade arrangements have made material resources and outputs capable of flowing more efficiently from any one corner of the globe to another than they were at any previous period. Economic welfare and growth now depend more on knowledge and social organization than on natural sources. Knowledge transfer and social reform, probably more than any other set of changes, can affect the economic opportunities and material development of

different parts of the global economic system, within economies both advanced and less developed. The emergence of a substantial nonmarket sector in the networked information economy offers opportunities for providing better access to knowledge and information as input from, and better access for information outputs of, developing and less-developed economies and poorer geographic and social sectors in the advanced economies. Better access to knowledge and the emergence of less capital-dependent forms of productive social organization offer the possibility that the emergence of the networked information economy will open up opportunities for improvement in economic justice, on scales both global and local.

The basic intuition and popular belief that the Internet will bring greater freedom and global equity has been around since the early 1990s. It has been the technophile's basic belief, just as the horrors of cyberporn, cybercrime, or cyberterrorism have been the standard gut-wrenching fears of the technophobe. The technophilic response is reminiscent of claims made in the past for electricity, for radio, or for telegraph, expressing what James Carey described as "the mythos of the electrical sublime." The question this part of the book explores is whether this claim, given the experience of the past decade, can be sustained on careful analysis, or whether it is yet another instance of a long line of technological utopianism. The fact that earlier utopias were overly optimistic does not mean that these previous technologies did not in fact alter the conditions of life - material, social, and intellectual. They did, but they did so differently in different societies, and in ways that diverged from the social utopias attached to them. Different nations absorbed and used these technologies differently, diverging in social and cultural habits, but also in institutional strategies for adoption - some more state-centric, others more market based; some more controlled, others less so. Utopian or at least best-case conceptions of the emerging condition are valuable if they help diagnose the socially and politically significant attributes of the emerging networked information economy correctly and allow us to form a normative conception of their significance. At a minimum, with these in hand, we can begin to design our institutional response to the present technological perturbation in order to improve the conditions of freedom and justice over the next few decades.

The chapters in this part focus on major liberal commitments or concerns. Chapter 5 addresses the question of individual autonomy. Chapters 6, 7, and 8 address democratic participation: first in the political public sphere and then, more broadly, in the construction of culture. Chapter 9 deals with justice and human development. Chapter 10 considers the effects of the networked information economy on community.

Chapter 5 Individual Freedom: Autonomy, Information, and Law

Freedom to Do More For Oneself, by Oneself, and With Others

Autonomy, Property, and Commons

Autonomy and the Information Environment

Autonomy, Mass Media, and Nonmarket Information Producers

Chapter 9 Justice and Development

Liberal Theories of Justice and the Networked Information Economy

Commons-Based Strategies for Human Welfare and Development

Information-Embedded Goods and Tools, Information, and Knowledge

Industrial Organization of HDI-Related Information Industries

Toward Adopting Commons-Based Strategies for Development

Commons-Based Research for Food and Medicines

Commons-Based Strategies for Development: Conclusion

Chapter 10 Social Ties: Networking Together

From "Virtual Communities" to Fear of Disintegration

A More Positive Picture Emerges Over Time

The Internet as a Platform for Human Connection

The Emergence of Social Software

The Internet and Human Community

Chapter 5
Individual Freedom: Autonomy, Information, and Law

The emergence of the networked information economy has the potential to increase individual autonomy. First, it increases the range and diversity of things that individuals can do for and by themselves. It does this by lifting, for one important domain of life, some of the central material constraints on what individuals can do that typified the industrial information economy. The majority of materials, tools, and platforms necessary for effective action in the information environment are in the hands of most individuals in advanced economies. Second, the networked information economy provides nonproprietary alternative sources of communications capacity and information, alongside the proprietary platforms of mediated communications. This decreases the extent to which individuals are subject to being acted upon by the owners of the facilities on which they depend for communications. The construction of consumers as passive objects of manipulation that typified television culture has not disappeared overnight, but it is losing its dominance in the information environment. Third, the networked information environment qualitatively increases the range and diversity of information available to individuals. It does so by enabling sources commercial and noncommercial, mainstream and fringe, domestic or foreign, to produce information and communicate with anyone. This diversity radically changes the universe of options that individuals can consider as open for them to pursue. It provides them a richer basis to form critical judgments about how they could live their lives, and, through this opportunity for critical reflection, why they should value the life they choose.

Freedom to Do More For Oneself, by Oneself, and With Others

Rory Cejas was a twenty-six-year-old firefighter/paramedic with the Miami Fire Department in 2003, when he enlisted the help of his brother, wife, and a friend to make a *Star Wars* - like fan film. Using a simple camcorder and tripod, and widely available film and image generation and editing software on his computer, he made a twenty-minute film he called *The Jedi Saga.* The film is not a parody. It is not social criticism. It is a straightforward effort to make a movie in the genre of *Star Wars*, using the same type of characters and story lines. In the predigital world, it would have been impossible, as a practical matter, for Cejas to do this. It would have been an implausible part of his life plan to cast his wife as a dark femme fatale, or his brother as a Jedi Knight, so they could battle shoulder-to-shoulder, light sabers drawn, against a platoon of Imperial clone soldiers. And it would have been impossible for him to distribute the film he had made to friends and strangers. The material conditions of cultural production have changed, so that it has now become part of his feasible set of options. He needs no help from government to do so. He needs no media access rules that give him access to fancy film studios. He needs no cable access rules to allow him to distribute his fantasy to anyone who wants to watch it. The new set of feasible options open to him includes not only the option passively to sit in the theatre or in front of the television and watch the images created by George Lucas, but also the option of trying his hand at making this type of film by himself.

Jedi Saga will not be a blockbuster. It is not likely to be watched by many people. Those who do watch it are not likely to enjoy it in the same way that they enjoyed any of Lucas's films, but that is not its point. When someone like Cejas makes such a film, he is not displacing what Lucas does. He is changing what he himself does - from sitting in front of a screen that is painted by another to painting his own screen. Those who watch it will enjoy it in the same way that friends and family enjoy speaking to each other or singing together, rather than watching talking heads or listening to

Talking Heads. Television culture, the epitome of the industrial information economy, structured the role of consumers as highly passive. While media scholars like John Fiske noted the continuing role of viewers in construing and interpreting the messages they receive, the role of the consumer in this model is well defined. The media product is a finished good that they consume, not one that they make. Nowhere is this clearer than in the movie theatre, where the absence of light, the enveloping sound, and the size of the screen are all designed to remove the viewer as agent, leaving only a set of receptors - eyes, ears - through which to receive the finished good that is the movie. There is nothing wrong with the movies as one mode of entertainment. The problem emerges, however, when the movie theatre becomes an apt metaphor for the relationship the majority of people have with most of the information environment they occupy. That increasing passivity of television culture came to be a hallmark of life for most people in the late stages of the industrial information economy. The couch potato, the eyeball bought and sold by Madison Avenue, has no part in making the information environment he or she occupies.

Perhaps no single entertainment product better symbolizes the shift that the networked information economy makes possible from television culture than the massive multiplayer online game. These games are typified by two central characteristics. First, they offer a persistent game environment. That is, any action taken or "object" created anywhere in the game world persists over time, unless and until it is destroyed by some agent in the game; and it exists to the same extent for all players. Second, the games are effectively massive collaboration platforms for thousands, tens of thousands - or in the case of Lineage, the most popular game in South Korea, more than four million-users. These platforms therefore provide individual players with various contexts in which to match their wits and skills with other human players. The computer gaming environment provides a persistent relational database of the actions and social interactions of players. The first games that became mass phenomena, like Ultima Online or Everquest, started with an already richly instantiated context. Designers of these games continue to play a large role in defining the range of actions and relations feasible for players. The basic medieval themes, the role of magic and weapons, and the types and ranges of actions that are possible create much of the context, and therefore the types of relationships pursued. Still, these games leave qualitatively greater room for individual effort and personal taste in producing the experience, the relationships, and hence the story line, relative to a television or movie experience. Second Life, a newer game by Linden Labs, offers us a glimpse into the next step in this genre of immersive entertainment. Like other massively multiplayer online games, Second Life is a persistent collaboration platform for its users. Unlike other games, however, Second Life offers only tools, with no story line, stock objects, or any cultural or meaning-oriented context whatsoever. Its users have created 99 percent of the objects in the game environment. The medieval village was nothing but blank space when they started. So was the flying vehicle design shop, the futuristic outpost, or the university, where some of the users are offering courses in basic programming skills and in-game design. Linden Labs charges a flat monthly subscription fee. Its employees focus on building tools that enable users to do everything from basic story concept down to the finest details of their own appearance and of objects they use in the game world. The in-game human relationships are those made by the users as they interact with each other in this immersive entertainment experience. The game's relationship to its users is fundamentally different from that of the movie or television studio. Movies and television seek to control the entire experience - rendering the viewer inert, but satisfied. Second Life sees the users as active makers of the entertainment environment that they occupy, and seeks to provide them with the tools they need to be so. The two models assume fundamentally different conceptions of play. Whereas in front of the television, the consumer is a passive receptacle, limited to selecting which finished good he or she will consume from a relatively narrow range of options, in the world of Second Life, the individual is treated as a fundamentally active, creative human being, capable of building his or her own fantasies, alone and in affiliation with others.

Second Life and *Jedi Saga* are merely examples, perhaps trivial ones, within the entertainment domain. They represent a shift in possibilities open both to human beings in the networked information economy and to the firms that sell them the tools for becoming active creators and users of their information environment. They are stark examples because of the centrality of the couch potato as the image of human action in television culture. Their characteristics are representative of the shift in the individual's role that is typical of the networked information economy in general and of peer production in particular. Linus Torvalds, the original creator of the Linux kernel development community, was, to use Eric Raymond's characterization, a designer with an itch to scratch. Peer-production projects often are composed of people who want to do something in the world and turn to the network to find a community of peers willing to work together to make that wish a reality. Michael Hart had been working in various contexts for more than thirty years when he - at first gradually, and more recently with increasing speed - harnessed the contributions of hundreds of volunteers to Project Gutenberg in pursuit of his goal to create a globally accessible library of public domain e-texts. Charles Franks was a computer programmer from Las Vegas when he decided he had a more efficient way to proofread those e-texts, and built an interface that allowed volunteers to compare scanned images of original texts with the e-texts available on Project Gutenberg. After working independently for a couple of years, he joined forces with Hart. Franks's facility now clears the volunteer work of more than one thousand proofreaders, who proof between two hundred and three hundred books a month. Each of the thousands of volunteers who participate in free software development projects, in *Wikipedia*, in the Open Directory Project, or in any of the many other peer-production projects, is living some version, as a major or minor part of their lives, of the possibilities captured by the stories of a Linus Torvalds, a Michael Hart, or *The Jedi Saga*. Each has decided to take advantage of some combination of technical, organizational, and social conditions within which we have come to live, and to become an active creator in his or her world, rather than merely to accept what was already there. The belief that it is possible to make something valuable happen in the world, and the practice of actually acting on that belief, represent a qualitative improvement in the condition of individual freedom. They mark the emergence of new practices of self-directed agency as a lived experience, going beyond mere formal permissibility and theoretical possibility.

Our conception of autonomy has not only been forged in the context of the rise of the democratic, civil rights-respecting state over its major competitors as a political system. In parallel, we have occupied the context of the increasing dominance of market-based industrial economy over its competitors. The culture we have developed over the past century is suffused with images that speak of the loss of agency imposed by that industrial economy. No cultural image better captures the way that mass industrial production reduced workers to cogs and consumers to receptacles than the one-dimensional curves typical of welfare economics - those that render human beings as mere production and demand functions. Their cultural, if not intellectual, roots are in Fredrick Taylor's *Theory of Scientific Management*: the idea of abstracting and defining all motions and actions of employees in the production process so that all the knowledge was in the system, while the employees were barely more than its replaceable parts. Taylorism, ironically, was a vast improvement over the depredations of the first industrial age, with its sweatshops and child labor. It nonetheless resolved into the kind of mechanical existence depicted in Charlie Chaplin's tragic-comic portrait, *Modern Times*. While the grind of industrial Taylorism seems far from the core of the advanced economies, shunted as it is now to poorer economies, the basic sense of alienation and lack of effective agency persists. Scott Adams's *Dilbert* comic strip, devoted to the life of a white-collar employee in a nameless U.S. corporation, thoroughly alienated from the enterprise, crimped by corporate hierarchy, resisting in all sorts of ways - but trapped in a cubicle - powerfully captures this sense for the industrial information economy in much the same way that Chaplin's *Modern Times* did for the industrial economy itself.

In the industrial economy and its information adjunct, most people live most of their lives within hierarchical relations of production, and within relatively tightly scripted possibilities after work, as consumers. It did not necessarily have to be this way. Michael Piore and Charles Sabel's *Second Industrial Divide* and Roberto Mangabeira Unger's *False Necessity* were central to the emergence of a "third way" literature that developed in the 1980s and 1990s to explore the possible alternative paths to production processes that did not depend so completely on the displacement of individual agency by hierarchical production systems. The emergence of radically decentralized, nonmarket production provides a new outlet for the attenuation of the constrained and constraining roles of employees and consumers. It is not limited to Northern Italian artisan industries or imagined for emerging economies, but is at the very heart of the most advanced market economies. Peer production and otherwise decentralized nonmarket production can alter the producer/consumer relationship with regard to culture, entertainment, and information. We are seeing the emergence of the user as a new category of relationship to information production and exchange. Users are individuals who are sometimes consumers and sometimes producers. They are substantially more engaged participants, both in defining the terms of their productive activity and in defining what they consume and how they consume it. In these two great domains of life - production and consumption, work and play - the networked information economy promises to enrich individual autonomy substantively by creating an environment built less around control and more around facilitating action.

The emergence of radically decentralized nonmarket production in general and of peer production in particular as feasible forms of action opens new classes of behaviors to individuals. Individuals can now justifiably believe that they can in fact do things that they want to do, and build things that they want to build in the digitally networked environment, and that this pursuit of their will need not, perhaps even cannot, be frustrated by insurmountable cost or an alien bureaucracy. Whether their actions are in the domain of political organization (like the organizers of MoveOn.org), or of education and professional attainment (as with the case of Jim Cornish, who decided to create a worldwide center of information on the Vikings from his fifth-grade schoolroom in Gander, Newfoundland), the networked information environment opens new domains for productive life that simply were not there before. In doing so, it has provided us with new ways to imagine our lives as productive human beings. Writing a free operating system or publishing a free encyclopedia may have seemed quixotic a mere few years ago, but these are now far from delusional. Human beings who live in a material and social context that lets them aspire to such things as possible for them to do, in their own lives, by themselves and in loose affiliation with others, are human beings who have a greater realm for their agency. We can live a life more authored by our own will and imagination than by the material and social conditions in which we find ourselves. At least we can do so more effectively than we could until the last decade of the twentieth century.

This new practical individual freedom, made feasible by the digital environment, is at the root of the improvements I describe here for political participation, for justice and human development, for the creation of a more critical culture, and for the emergence of the networked individual as a more fluid member of community. In each of these domains, the improvements in the degree to which these liberal commitments are honored and practiced emerge from new behaviors made possible and effective by the networked information economy. These behaviors emerge now precisely because individuals have a greater degree of freedom to act effectively, unconstrained by a need to ask permission from anyone. It is this freedom that increases the salience of nonmonetizable motivations as drivers of production. It is this freedom to seek out whatever information we wish, to write about it, and to join and leave various projects and associations with others that underlies the new efficiencies we see in the networked information economy. These behaviors underlie the cooperative news and commentary production that form the basis of the networked public sphere,

and in turn enable us to look at the world as potential participants in discourse, rather than as potential viewers only. They are at the root of making a more transparent and reflective culture. They make possible the strategies I suggest as feasible avenues to assure equitable access to opportunities for economic participation and to improve human development globally.

Treating these new practical opportunities for action as improvements in autonomy is not a theoretically unproblematic proposition. For all its intuitive appeal and centrality, autonomy is a notoriously nebulous concept. In particular, there are deep divisions within the literature as to whether it is appropriate to conceive of autonomy in substantive terms - as Gerald Dworkin, Joseph Raz, and Joel Feinberg most prominently have, and as I have here - or in formal terms. Formal conceptions of autonomy are committed to assuming that all people have the capacity for autonomous choice, and do not go further in attempting to measure the degree of freedom people actually exercise in the world in which they are in fact constrained by circumstances, both natural and human. This commitment is not rooted in some stubborn unwillingness to recognize the slings and arrows of outrageous fortune that actually constrain our choices. Rather, it comes from the sense that only by treating people as having these capacities and abilities can we accord them adequate respect as free, rational beings, and avoid sliding into overbearing paternalism. As Robert Post put it, while autonomy may well be something that needs to be "achieved" as a descriptive matter, the "structures of social authority" will be designed differently depending on whether or not individuals are treated as autonomous. "From the point of view of the designer of the structure, therefore, the presence or absence of autonomy functions as an axiomatic and foundational principle."[/1] Autonomy theory that too closely aims to understand the degree of autonomy people actually exercise under different institutional arrangements threatens to form the basis of an overbearing benevolence that would undermine the very possibility of autonomous action.

While the fear of an overbearing bureaucracy benevolently guiding us through life toward becoming more autonomous is justifiable, the formal conception of autonomy pays a high price in its bluntness as a tool to diagnose the autonomy implications of policy. Given how we are: situated, context-bound, messy individuals, it would be a high price to pay to lose the ability to understand how law and policy actually affect whatever capacity we do have to be the authors of our own life choices in some meaningful sense. We are individuals who have the capacity to form beliefs and to change them, to form opinions and plans and defend them - but also to listen to arguments and revise our beliefs. We experience some decisions as being more free than others; we mock or lament ourselves when we find ourselves trapped by the machine or the cubicle, and we do so in terms of a sense of helplessness, a negation of freedom, not only, or even primarily, in terms of lack of welfare; and we cherish whatever conditions those are that we experience as "free" precisely for that freedom, not for other reasons. Certainly, the concerns with an overbearing state, whether professing benevolence or not, are real and immediate. No one who lives with the near past of the totalitarianism of the twentieth century or with contemporary authoritarianism and fundamentalism can belittle these. But the great evils that the state can impose through formal law should not cause us to adopt methodological commitments that would limit our ability to see the many ways in which ordinary life in democratic societies can nonetheless be more or less free, more or less conducive to individual self-authorship.

If we take our question to be one concerned with diagnosing the condition of freedom of individuals, we must observe the conditions of life from a first-person, practical perspective - that is, from the perspective of the person whose autonomy we are considering. If we accept that all individuals are always constrained by personal circumstances both physical and social, then the way to think about autonomy of human agents is to inquire into the relative capacity of individuals to be the authors of their lives within the constraints of context. From this perspective, whether the

sources of constraint are private actors or public law is irrelevant. What matters is the extent to which a particular configuration of material, social, and institutional conditions allows an individual to be the author of his or her life, and to what extent these conditions allow others to act upon the individual as an object of manipulation. As a means of diagnosing the conditions of individual freedom in a given society and context, we must seek to observe the extent to which people are, in fact, able to plan and pursue a life that can reasonably be described as a product of their own choices. It allows us to compare different conditions, and determine that a certain condition allows individuals to do more for themselves, without asking permission from anyone. In this sense, we can say that the conditions that enabled Cejas to make *Jedi Saga* are conditions that made him more autonomous than he would have been without the tools that made that movie possible. It is in this sense that the increased range of actions we can imagine for ourselves in loose affiliation with others - like creating a Project Gutenberg - increases our ability to imagine and pursue life plans that would have been impossible in the recent past.

From the perspective of the implications of autonomy for how people act in the digital environment, and therefore how they are changing the conditions of freedom and justice along the various dimensions explored in these chapters, this kind of freedom to act is central. It is a practical freedom sufficient to sustain the behaviors that underlie the improvements in these other domains. From an internal perspective of the theory of autonomy, however, this basic observation that people can do more by themselves, alone or in loose affiliation with others, is only part of the contribution of the networked information economy to autonomy, and a part that will only be considered an improvement by those who conceive of autonomy as a substantive concept. The implications of the networked information economy for autonomy are, however, broader, in ways that make them attractive across many conceptions of autonomy. To make that point, however, we must focus more specifically on law as the source of constraint, a concern common to both substantive and formal conceptions of autonomy. As a means of analyzing the implications of law to autonomy, the perspective offered here requires that we broaden our analysis beyond laws that directly limit autonomy. We must also look to laws that structure the conditions of action for individuals living within the ambit of their effect. In particular, where we have an opportunity to structure a set of core resources necessary for individuals to perceive the state of the world and the range of possible actions, and to communicate their intentions to others, we must consider whether the way we regulate these resources will create systematic limitations on the capacity of individuals to control their own lives, and in their susceptibility to manipulation and control by others. Once we recognize that there cannot be a person who is ideally "free," in the sense of being unconstrained or uncaused by the decisions of others, we are left to measure the effects of all sorts of constraints that predictably flow from a particular legal arrangement, in terms of the effect they have on the relative role that individuals play in authoring their own lives.

Autonomy, Property, and Commons

The first legal framework whose role is altered by the emergence of the networked information economy is the property-like regulatory structure of patents, copyrights, and similar exclusion mechanisms applicable to information, knowledge, and culture. Property is usually thought in liberal theory to enhance, rather than constrain, individual freedom, in two quite distinct ways. First, it provides security of material context - that is, it allows one to know with some certainty that some set of resources, those that belong to her, will be available for her to use to execute her plans over time. This is the core of Kant's theory of property, which relies on a notion of positive liberty, the freedom to do things successfully based on life plans we can lay for ourselves. Second, property and markets provide greater freedom of action for the individual owner as compared both, as Marx diagnosed, to the feudal arrangements that preceded them, and, as he decidedly did not but Hayek

did, to the models of state ownership and regulation that competed with them throughout most of the twentieth century.

Markets are indeed institutional spaces that enable a substantial degree of free choice. "Free," however, does not mean "anything goes." If John possesses a car and Jane possesses a gun, a market will develop only if John is prohibited from running Jane over and taking her gun, and also if Jane is prohibited from shooting at John or threatening to shoot him if he does not give her his car. A market that is more or less efficient will develop only if many other things are prohibited to, or required of, one or both sides - like monopolization or disclosure. Markets are, in other words, structured relationships intended to elicit a particular datum - the comparative willingness and ability of agents to pay for goods or resources. The most basic set of constraints that structure behavior in order to enable markets are those we usually call property. Property is a cluster of background rules that determine what resources each of us has when we come into relations with others, and, no less important, what "having" or "lacking" a resource entails in our relations with these others. These rules impose constraints on who can do what in the domain of actions that require access to resources that are the subjects of property law. They are aimed to crystallize asymmetries of power over resources, which then form the basis for exchanges - I will allow you to do X, which I am asymmetrically empowered to do (for example, watch television using this cable system), and you, in turn, will allow me to do Y, which you are asymmetrically empowered to do (for example, receive payment from your bank account). While a necessary precondition for markets, property also means that choice in markets is itself not free of constraints, but is instead constrained in a particular pattern. It makes some people more powerful with regard to some things, and must constrain the freedom of action of others in order to achieve this asymmetry./2

Commons are an alternative form of institutional space, where human agents can act free of the particular constraints required for markets, and where they have some degree of confidence that the resources they need for their plans will be available to them. Both freedom of action and security of resource availability are achieved in very different patterns than they are in property-based markets. As with markets, commons do not mean that anything goes. Managing resources as commons does, however, mean that individuals and groups can use those resources under different types of constraints than those imposed by property law. These constraints may be social, physical, or regulatory. They may make individuals more free or less so, in the sense of permitting a greater or lesser freedom of action to choose among a range of actions that require access to resources governed by them than would property rules in the same resources. Whether having a particular type of resource subject to a commons, rather than a property-based market, enhances freedom of action and security, or harms them, is a context-specific question. It depends on how the commons is structured, and how property rights in the resource would have been structured in the absence of a commons. The public spaces in New York City, like Central Park, Union Square, or any sidewalk, afford more people greater freedom than does a private backyard - certainly to all but its owner. Given the diversity of options that these public spaces make possible as compared to the social norms that neighbors enforce against each other, they probably offer more freedom of action than a backyard offers even to its owner in many loosely urban and suburban communities. Swiss pastures or irrigation districts of the type that Elinor Ostrom described as classic cases of long-standing sustainable commons offer their participants security of holdings at least as stable as any property system, but place substantial traditional constraints on who can use the resources, how they can use them, and how, if at all, they can transfer their rights and do something completely different. These types of commons likely afford their participants less, rather than more, freedom of action than would have been afforded had they owned the same resource in a market-alienable property arrangement, although they retain security in much the same way. Commons, like the air, the sidewalk, the road and highway, the ocean, or the public beach, achieve security on a very

different model. I can rely on the resources so managed in a probabilistic, rather than deterministic sense. I can plan to meet my friends for a picnic in the park, not because I own the park and can direct that it be used for my picnic, but because I know there will be a park, that it is free for me to use, and that there will be enough space for us to find a corner to sit in. This is also the sort of security that allows me to plan to leave my house at some hour, and plan to be at work at some other hour, relying not on owning the transportation path, but on the availability to me of the roads and highways on symmetric terms to its availability to everyone else. If we look more closely, we will see that property and markets also offer only a probabilistic security of context, whose parameters are different - for example, the degree of certainty we have as to whether the resource we rely on as our property will be stolen or damaged, whether it will be sufficient for what we need, or if we need more, whether it will be available for sale and whether we will be able to afford it.

Like property and markets, then, commons provide both freedom of action and security of context. They do so, however, through the imposition of different constraints than do property and market rules. In particular, what typifies all these commons in contradistinction to property is that no actor is empowered by law to act upon another as an object of his or her will. I can impose conditions on your behavior when you are walking on my garden path, but I have no authority to impose on you when you walk down the sidewalk. Whether one or the other of the two systems, used exclusively, will provide "greater freedom" in some aggregate sense is not a priori determinable. It will depend on the technical characteristics of the resource, the precise contours of the rules of, respectively, the proprietary market and the commons, and the distribution of wealth in society. Given the diversity of resources and contexts, and the impossibility of a purely "anything goes" absence of rules for either system, some mix of the two different institutional frameworks is likely to provide the greatest diversity of freedom to act in a material context. This diversity, in turn, enables the greatest freedom to plan action within material contexts, allowing individuals to trade off the availabilities of, and constraints on, different resources to forge a context sufficiently provisioned to enable them to execute their plans, while being sufficiently unregulated to permit them to do so. Freedom inheres in diversity of constraint, not in the optimality of the balance of freedom and constraint represented by any single institutional arrangement. It is the diversity of constraint that allows individuals to plan to live out different portions and aspects of their lives in different institutional contexts, taking advantage of the different degrees of freedom and security they make possible.

In the context of information, knowledge, and culture, because of the nonrivalry of information and its characteristic as input as well as output of the production process, the commons provides substantially greater security of context than it does when material resources, like parks or roadways, are at stake. Moreover, peer production and the networked information economy provide an increasingly robust source of new information inputs. This reduces the risk of lacking resources necessary to create new expressions or find out new things, and renders more robust the freedom to act without being susceptible to constraint from someone who holds asymmetrically greater power over the information resources one needs. As to information, then, we can say with a high degree of confidence that a more expansive commons improves individual autonomy, while enclosure of the public domain undermines it. This is less determinate with communications systems. Because computers and network connections are rival goods, there is less certainty that a commons will deliver the required resources. Under present conditions, a mixture of commons-based and proprietary communications systems is likely to improve autonomy. If, however, technological and social conditions change so that, for example, sharing on the model of peer-to-peer networks, distributed computation, or wireless mesh networks will be able to offer as dependable a set of communications and computation resources as the Web offers information and

knowledge resources, the relative attractiveness of commons-oriented communications policies will increase from the perspective of autonomy.

Autonomy and the Information Environment

The structure of our information environment is constitutive of our autonomy, not only functionally significant to it. While the capacity to act free of constraints is most immediately and clearly changed by the networked information economy, information plays an even more foundational role in our very capacity to make and pursue life plans that can properly be called our own. A fundamental requirement of self-direction is the capacity to perceive the state of the world, to conceive of available options for action, to connect actions to consequences, to evaluate alternative outcomes, and to decide upon and pursue an action accordingly. Without these, no action, even if mechanically self-directed in the sense that my brain consciously directs my body to act, can be understood as autonomous in any normatively interesting sense. All of the components of decision making prior to action, and those actions that are themselves communicative moves or require communication as a precondition to efficacy, are constituted by the information and communications environment we, as agents, occupy. Conditions that cause failures at any of these junctures, which place bottlenecks, failures of communication, or provide opportunities for manipulation by a gatekeeper in the information environment, create threats to the autonomy of individuals in that environment. The shape of the information environment, and the distribution of power within it to control information flows to and from individuals, are, as we have seen, the contingent product of a combination of technology, economic behavior, social patterns, and institutional structure or law.

In 1999, Cisco Systems issued a technical white paper, which described a new router that the company planned to sell to cable broadband providers. In describing advantages that these new "policy routers" offer cable providers, the paper explained that if the provider's users want to subscribe to a service that "pushes" information to their computer: "You could restrict the incoming push broadcasts as well as subscribers' outgoing access to the push site to discourage its use. At the same time, you could promote your own or a partner's services with full speed features to encourage adoption of your services."/3

In plain English, the broadband provider could inspect the packets flowing to and from a customer, and decide which packets would go through faster and more reliably, and which would slow down or be lost. Its engineering purpose was to improve quality of service. However, it could readily be used to make it harder for individual users to receive information that they want to subscribe to, and easier for them to receive information from sites preferred by the provider - for example, the provider's own site, or sites of those who pay the cable operator for using this function to help "encourage" users to adopt their services. There are no reports of broadband providers using these capabilities systematically. But occasional events, such as when Canada's second largest telecommunications company blocked access for all its subscribers and those of smaller Internet service providers that relied on its network to the website of the Telecommunications Workers Union in 2005, suggest that the concern is far from imaginary.

It is fairly clear that the new router increases the capacity of cable operators to treat their subscribers as objects, and to manipulate their actions in order to make them act as the provider wills, rather than as they would have had they had perfect information. It is less obvious whether this is a violation of, or a decrease in, the autonomy of the users. At one extreme, imagine the home as a black box with no communications capabilities save one - the cable broadband connection. Whatever comes through that cable is, for all practical purposes, "the state of the world," as far as

the inhabitants of that home know. In this extreme situation, the difference between a completely neutral pipe that carries large amounts of information indiscriminately, and a pipe finely controlled by the cable operator is a large one, in terms of the autonomy of the home's inhabitants. If the pipe is indiscriminate, then the choices of the users determine what they know; decisions based on that knowledge can be said to be autonomous, at least to the extent that whether they are or are not autonomous is a function of the state of the agent's knowledge when forming a decision. If the pipe is finely controlled and purposefully manipulated by the cable operator, by contrast, then decisions that individuals make based on the knowledge they acquire through that pipe are substantially a function of the choices of the controller of the pipe, not of the users. At the other extreme, if each agent has dozens of alternative channels of communication to the home, and knows how the information flow of each one is managed, then the introduction of policy routers into one or some of those channels has no real implications for the agent's autonomy. While it may render one or more channels manipulable by their provider, the presence of alternative, indiscriminate channels, on the one hand, and of competition and choice among various manipulated channels, on the other hand, attenuates the extent to which the choices of the provider structure the universe of information within which the individual agent operates. The provider no longer can be said to shape the individual's choices, even if it tries to shape the information environment observable through its channel with the specific intent of manipulating the actions of users who view the world through its pipe. With sufficient choice among pipes, and sufficient knowledge about the differences between pipes, the very choice to use the manipulated pipe can be seen as an autonomous act. The resulting state of knowledge is self-selected by the user. Even if that state of knowledge then is partial and future actions constrained by it, the limited range of options is itself an expression of the user's autonomy, not a hindrance on it. For example, consider the following: Odysseus and his men mix different forms of freedom and constraint in the face of the Sirens. Odysseus maintains his capacity to acquire new information by leaving his ears unplugged, but binds himself to stay on the ship by having his men tie him to the mast. His men choose the same course at the same time, but bind themselves to the ship by having Odysseus stop their ears with wax, so that they do not get the new information - the siren songs - that might change their minds and cause them not to stay the course. Both are autonomous when they pass by the Sirens, though both are free only because of their current incapacity. Odysseus's incapacity to jump into the water and swim to the Sirens and his men's incapacity to hear the siren songs are a result of their autonomously chosen past actions.

The world we live in is neither black box nor cornucopia of well-specified communications channels. However, characterizing the range of possible configurations of the communications environment we occupy as lying on a spectrum from one to the other provides us with a framework for describing the degree to which actual conditions of a communications environment are conducive to individual autonomy. More important perhaps, it allows us to characterize policy and law that affects the communications environment as improving or undermining individual autonomy. Law can affect the range of channels of communications available to individuals, as well as the rules under which they are used. How many communications channels and sources of information can an individual receive? How many are available for him or her to communicate with others? Who controls these communications channels? What does control over the communications channels to an agent entail? What can the controller do, and what can it not? All of these questions are the subject of various forms of policy and law. Their implications affect the degree of autonomy possessed by individuals operating with the institutional-technical-economic framework thus created.

There are two primary types of effects that information law can have on personal autonomy. The first type is concerned with the relative capacity of some people systematically to constrain the

perceptions or shape the preferences of others. A law that systematically gives some people the power to control the options perceived by, or the preferences of, others, is a law that harms autonomy. Government regulation of the press and its propaganda that attempts to shape its subjects' lives is a special case of this more general concern. This concern is in some measure quantitative, in the sense that a greater degree of control to which one is subject is a greater offense to autonomy. More fundamentally, a law that systematically makes one adult susceptible to the control of another offends the autonomy of the former. Law has created the conditions for one person to act upon another as an object. This is the nonpragmatic offense to autonomy committed by abortion regulations upheld in *Planned Parenthood v. Casey* - such as requirements that women who seek abortions listen to lectures designed to dissuade them. These were justified by the plurality there, not by the claim that they did not impinge on a woman's autonomy, but that the state's interest in the potential life of a child trumps the autonomy of the pregnant woman.

The second type of effect that law can have on autonomy is to reduce significantly the range and variety of options open to people in society generally, or to certain classes of people. This is different from the concern with government intervention generally. It is not focused on whether the state prohibits these options, but only on whether the effect of the law is to remove options. It is less important whether this effect is through prohibition or through a set of predictable or observable behavioral adaptations among individuals and organizations that, as a practical matter, remove these options. I do not mean to argue for the imposition of restraints, in the name of autonomy, on any lawmaking that results in a removal of any single option, irrespective of the quantity and variety of options still open. Much of law does that. Rather, the autonomy concern is implicated by laws that systematically and significantly reduce the number, and more important, impoverish the variety, of options open to people in the society for which the law is passed.

"Number and variety" is intended to suggest two dimensions of effect on the options open to an individual. The first is quantitative. For an individual to author her own life, she must have a significant set of options from which to choose; otherwise, it is the choice set - or whoever, if anyone, made it so - and not the individual, that is governing her life. This quantitative dimension, however, does not mean that more choices are always better, from the individual's perspective. It is sufficient that the individual have some adequate threshold level of options in order for him or her to exercise substantive self-authorship, rather than being authored by circumstances. Beyond that threshold level, additional options may affect one's welfare and success as an autonomous agent, but they do not so constrain an individual's choices as to make one not autonomous. Beyond quantitative adequacy, the options available to an individual must represent meaningfully different paths, not merely slight variations on a theme. Qualitatively, autonomy requires the availability of options in whose adoption or rejection the individual can practice critical reflection and life choices. In order to sustain the autonomy of a person born and raised in a culture with a set of socially embedded conventions about what a good life is, one would want a choice set that included at least some unconventional, non-mainstream, if you will, critical options. If all the options one has - even if, in a purely quantitative sense, they are "adequate" - are conventional or mainstream, then one loses an important dimension of self-creation. The point is not that to be truly autonomous one necessarily must be unconventional. Rather, if self-governance for an individual consists in critical reflection and re-creation by making choices over the course of his life, then some of the options open must be different from what he would choose simply by drifting through life, adopting a life plan for no reason other than that it is accepted by most others. A person who chooses a conventional life in the presence of the option to live otherwise makes that conventional life his or her own in a way that a person who lives a conventional life without knowing about alternatives does not.

As long as our autonomy analysis of information law is sensitive to these two effects on information flow to, from, and among individuals and organizations in the regulated society, it need not conflict with the concerns of those who adopt the formal conception of autonomy. It calls for no therapeutic agenda to educate adults in a wide range of options. It calls for no one to sit in front of educational programs. It merely focuses on two core effects that law can have through the way it structures the relationships among people with regard to the information environment they occupy. If a law - passed for any reason that may or may not be related to autonomy concerns - creates systematic shifts of power among groups in society, so that some have a greater ability to shape the perceptions of others with regard to available options, consequences of action, or the value of preferences, then that law is suspect from an autonomy perspective. It makes the choices of some people less their own and more subject to manipulation by those to whom the law gives the power to control perceptions. Furthermore, a law that systematically and severely limits the range of options known to individuals is one that imposes a normative price, in terms of autonomy, for whatever value it is intended to deliver. As long as the focus of autonomy as an institutional design desideratum is on securing the best possible information flow to the individual, the designer of the legal structure need not assume that individuals are not autonomous, or have failures of autonomy, in order to serve autonomy. All the designer need assume is that individuals will not act in order to optimize the autonomy of their neighbors. Law then responds by avoiding institutional designs that facilitate the capacity of some groups of individuals to act on others in ways that are systematically at the expense of the ability of those others to control their own lives, and by implementing policies that predictably diversify the set of options that all individuals are able to see as open to them.

Throughout most of the 1990s and currently, communications and information policy around the globe was guided by a wish to "let the private sector lead," interpreted in large measure to mean that various property and property-like regulatory frameworks should be strengthened, while various regulatory constraints on property-like rights should be eased. The drive toward proprietary, market-based provisioning of communications and information came from disillusionment with regulatory systems and state-owned communications networks. It saw the privatization of national postal, telephone, and telegraph authorities (PTTs) around the world. Even a country with a long tradition of state-centric communications policy, like France, privatized much of its telecommunications systems. In the United States, this model translated into efforts to shift telecommunications from the regulated monopoly model it followed throughout most of the twentieth century to a competitive market, and to shift Internet development from being primarily a government-funded exercise, as it had been from the late 1960s to the mid 1990s, to being purely private property, market based. This model was declared in the Clinton administration's 1993 *National Information Infrastructure: Agenda for Action*, which pushed for privatization of Internet deployment and development. It was the basis of that administration's 1995 *White Paper on Intellectual Property*, which mapped the most aggressive agenda ever put forward by any American administration in favor of perfect enclosure of the public domain; and it was in those years when the Federal Communications Commission (FCC) first implemented spectrum auctions aimed at more thorough privatization of wireless communications in the United States. The general push for stronger intellectual property rights and more marketcentric telecommunications systems also became a central tenet of international trade regimes, pushing similar policies in smaller and developing economies.

The result of the push toward private provisioning and deregulation has led to the emergence of a near-monopolistic market structure for wired physical broadband services. By the end of 2003, more than 96 percent of homes and small offices in the United States that had any kind of "high-speed" Internet services received their service from either their incumbent cable operator or their incumbent local telephone company. If one focuses on the subset of these homes and offices

that get service that provides more substantial room for autonomous communicative action - that is, those that have upstream service at high-speed, enabling them to publish and participate in online production efforts and not simply to receive information at high speeds - the picture is even more dismal. Less than 2 percent of homes and small offices receive their broadband connectivity from someone other than their cable carrier or incumbent telephone carrier. More than 83 percent of these users get their access from their cable operator. Moreover, the growth rate in adoption of cable broadband and local telephone digital subscriber line (DSL) has been high and positive, whereas the growth rate of the few competing platforms, like satellite broadband, has been stagnant or shrinking. The proprietary wired environment is gravitating toward a high-speed connectivity platform that will be either a lopsided duopoly, or eventually resolve into a monopoly platform./4 These owners are capable, both technically and legally, of installing the kind of policy routers with which I opened the discussion of autonomy and information law - routers that would allow them to speed up some packets and slow down or reject others in ways intended to shape the universe of information available to users of their networks.

The alternative of building some portions of our telecommunications and information production and exchange systems as commons was not understood in the mid-1990s, when the policy that resulted in this market structure for communications was developed. As we saw in chapter 3, however, wireless communications technology has progressed to the point where it is now possible for users to own equipment that cooperates in mesh networks to form a "last-mile" infrastructure that no one other than the users own. Radio networks can now be designed so that their capital structure more closely approximates the Internet and personal computer markets, bringing with it a greater scope for commons-based peer production of telecommunications infrastructure. Throughout most of the twentieth century, wireless communications combined high-cost capital goods (radio transmitters and antennae towers) with cheaper consumer goods (radio receivers), using regulated proprietary infrastructure, to deliver a finished good of wireless communications on an industrial model. Now WiFi is marking the possibility of an inversion of the capital structure of wireless communication. We see end-user equipment manufacturers like Intel, Cisco, and others producing and selling radio "transceivers" that are shareable goods. By using ad hoc mesh networking techniques, some early versions of which are already being deployed, these transceivers allow their individual owners to cooperate and coprovision their own wireless communications network, without depending on any cable carrier or other wired provider as a carrier of last resort. Almost the entire debate around spectrum policy and the relative merits of markets and commons in wireless policy is conducted today in terms of efficiency and innovation. A common question these days is which of the two approaches will lead to greater growth of wireless communications capacity and will more efficiently allocate the capacity we already have. I have contributed my fair share of this form of analysis, but the question that concerns us here is different. We must ask what, if any, are the implications of the emergence of a feasible, sustainable model of a commons-based physical infrastructure for the first and last mile of the communications environment, in terms of individual autonomy?

The choice between proprietary and commons-based wireless data networks takes on new significance in light of the market structure of the wired network, and the power it gives owners of broadband networks to control the information flow into the vast majority of homes. Commons-based wireless systems become the primary legal form of communications capacity that does not systematically subject its users to manipulation by an infrastructure owner.

Imagine a world with four agents - A, B, C, and D - connected to each other by a communications network. Each component, or route, of the network could be owned or unowned. If all components are unowned, that is, are organized as a commons, each agent has an equal privilege to

use any component of the network to communicate with any other agent. If all components are owned, the owner of any network component can deny to any other agent use of that network component to communicate with anyone else. This translates in the real world into whether or not there is a "spectrum owner" who "owns" the link between any two users, or whether the link is simply a consequence of the fact that two users are communicating with each other in a way that no one has a right to prevent them from doing.

In this simple model, if the network is unowned, then for any communication all that is required is a willing sender and a willing recipient. No third agent gets a say as to whether any other pair will communicate with each other. Each agent determines independently of the others whether to participate in a communicative exchange, and communication occurs whenever all its participants, and only they, agree to communicate with each other. For example, A can exchange information with B, as long as B consents. The only person who has a right to prevent A from receiving information from, or sending information to, B, is B, in the exercise of B's own autonomous choice whether to change her information environment. Under these conditions, neither A nor B is subject to control of her information environment by others, except where such control results from denying her the capacity to control the information environment of another. If all network components are owned, on the other hand, then for any communication there must be a willing sender, a willing recipient, and a willing infrastructure owner. In a pure property regime, infrastructure owners have a say over whether, and the conditions under which, others in their society will communicate with each other. It is precisely the power to prevent others from communicating that makes infrastructure ownership a valuable enterprise: One can charge for granting one's permission to communicate. For example, imagine that D owns all lines connecting A to B directly or through D, and C owns all lines connecting A or B to C. As in the previous scenario, A wishes to exchange information with B. Now, in addition to B, A must obtain either C's or D's consent. A now functions under two distinct types of constraint. The first, as before, is a constraint imposed by B's autonomy: A cannot change B's information environment (by exchanging information with her) without B's consent. The second constraint is that A must persuade an owner of whatever carriage medium connects A to B to permit A and B to communicate. The communication is not sent to or from C or D. It does not change C's or D's information environment, and that is not A's intention. C and D's ability to consent or withhold consent is not based on the autonomy principle. It is based, instead, on an instrumental calculus: namely, that creating such property rights in infrastructure will lead to the right incentives for the deployment of infrastructure necessary for A and B to communicate in the first place.

Now imagine that D owns the entire infrastructure. If A wants to get information from B or to communicate to C in order to persuade C to act in a way that is beneficial to A, A needs D's permission. D may grant or withhold permission, and may do so either for a fee or upon the imposition of conditions on the communication. Most significantly, D can choose to prevent anyone from communicating with anyone else, or to expose each participant to the communications of only some, but not all, members of society. This characteristic of her ownership gives D the power to shape A's information environment by selectively exposing A to information in the form of communications from others. Most commonly, we might see this where D decides that B will pay more if all infrastructure is devoted to permitting B to communicate her information to A and C, rather than any of it used to convey A's statements to C. D might then refuse to carry A's message to C and permit only B to communicate to A and C. The point is that from A's perspective, A is dependent upon D's decisions as to what information can be carried on the infrastructure, among whom, and in what directions. To the extent of that dependence, A's autonomy is compromised. We might call the requirement that D can place on A as a precondition to using the infrastructure an "influence exaction."

The magnitude of the negative effect on autonomy, or of the influence exaction, depends primarily on (a) the degree to which it is hard or easy to get around D's facility, and (b) the degree of transparency of the exaction. Compare, for example, Cisco's policy router for cable broadband, which allows the cable operator to speed up and slow down packets based on its preferences, to Amazon's brief experiment in 1998-1999 with accepting undisclosed payments from publishers in exchange for recommending their books. If a cable operator programs its routers to slow down packets of competitors, or of information providers that do not pay, this practice places a significant exaction on users. First, the exaction is entirely nontransparent. There are many reasons that different sites load at different speeds, or even fail to load altogether. Users, the vast majority of whom are unaware that the provider could, if it chose, regulate the flow of information to them, will assume that it is the target site that is failing, not that their own service provider is manipulating what they can see. Second, there is no genuine work-around. Cable broadband covers roughly two-thirds of the home market, in many places without alternative; and where there is an alternative, there is only one - the incumbent telephone company. Without one of these noncompetitive infrastructure owners, the home user has no broadband access to the Internet. In Amazon's case, the consumer outrage when the practice was revealed focused on the lack of transparency. Users had little objection to clearly demarcated advertisement. The resistance was to the nontransparent manipulation of the recommendation system aimed at causing the consumers to act in ways consistent with Amazon's goals, rather than their own. In that case, however, there were alternatives. There are many different places from which to find book reviews and recommendations, and at the time, barnesandnoble.com was already available as an online bookseller - and had not significantly adopted similar practices. The exaction was therefore less significant. Moreover, once the practice was revealed, Amazon publicly renounced it and began to place advertisements in a clearly recognizable separate category. The lesson was not lost on others. When Google began at roughly the same time as a search engine, it broke with the then-common practice of selling search-result location. When the company later introduced advertised links, it designed its interface to separate out clearly the advertisements from the algorithm-based results, and to give the latter more prominent placement than the former. This does not necessarily mean that any search engine that accepts payments for linking is necessarily bad. A search engine like Overture, which explicitly and publicly returns results ranked according to which, among the sites retrieved, paid Overture the most, has its own value for consumers looking for commercial sites. A transparent, nonmonopolistic option of this sort increases, rather than decreases, the freedom of users to find the information they want and act on it. The problem would be with search engines that mix the two strategies and hide the mix, or with a monopolistic search engine.

Because of the importance of the possibility to work around the owned infrastructure, the degree of competitiveness of any market in such infrastructure is important. Before considering the limits of even competitive markets by comparison to commons, however, it is important to recognize that a concern with autonomy provides a distinct justification for the policy concern with media concentration. To understand the effects of concentration, we can think of freedom from constraint as a dimension of welfare. Just as we have no reason to think that in a concentrated market, total welfare, let alone consumer welfare, will be optimal, we also have no reason to think that a component of welfare - freedom from constraint as a condition to access one's communicative environment - will be optimal. Moreover, when we use a "welfare" calculus as a metaphor for the degree of autonomy users have in the system, we must optimize not total welfare, as we do in economic analysis, but only what in the metaphorical calculus would count as "consumer surplus." In the domain of influence and autonomy, only "consumer surplus" counts as autonomy enhancing. "Producer surplus," the degree of successful imposition of influence on others as a condition of service, translates in an autonomy calculus into control exerted by some people (providers) over others (consumers). It reflects the successful negation of autonomy. The

monopoly case therefore presents a new normative dimension of the well-known critiques of media concentration. Why, however, is this not solely an analysis of media concentration? Why does a competitive market in infrastructure not solve the autonomy deficit of property?

If we make standard assumptions of perfectly competitive markets and apply them to our A-B-D example, one would think that the analysis must change. D no longer has monopoly power. We would presume that the owners of infrastructure would be driven by competition to allocate infrastructure to uses that users value most highly. If one owner "charges" a high price in terms of conditions imposed on users, say to forgo receiving certain kinds of speech uncongenial to the owner, then the users will go to a competitor who does not impose that condition. This standard market response is far from morally irrelevant if one is concerned with autonomy. If, in fact, every individual can choose precisely the package of influence exactions and the cash-to-influence trade-off under which he or she is willing to communicate, then the autonomy deficit that I suggest is created by property rights in communications infrastructure is minimal. If all possible degrees of freedom from the influence of others are available to autonomous individuals, then respecting their choices, including their decisions to subject themselves to the influence of others in exchange for releasing some funds so they are available for other pursuits, respects their autonomy.

Actual competition, however, will not eliminate the autonomy deficit of privately owned communications infrastructure, for familiar reasons. The most familiar constraint on the "market will solve it" hunch is imposed by transaction costs - in particular, information-gathering and negotiation costs. Influence exactions are less easily homogenized than prices expressed in currency. They will therefore be more expensive to eliminate through transactions. Some people value certain kinds of information lobbed at them positively; others negatively. Some people are more immune to suggestion, others less. The content and context of an exaction will have a large effect on its efficacy as a device for affecting the choices of the person subject to its influence, and these could change from communication to communication for the same person, let alone for different individuals. Both users and providers have imperfect information about the users' susceptibility to manipulated information flows; they have imperfect information about the value that each user would place on being free of particular exactions. Obtaining the information necessary to provide a good fit for each consumer's preferences regarding the right influence-to-cash ratio for a given service would be prohibitively expensive. Even if the information were obtained, negotiating the precise cash-to-influence trade-off would be costly. Negotiation also may fail because of strategic behavior. The consumer's ideal outcome is to labor under an exaction that is ineffective. If the consumer can reduce the price by submitting to constraints on communication that would affect an average consumer, but will not change her agenda or subvert her capacity to author her life, she has increased her welfare without compromising her autonomy. The vendor's ideal outcome, however, is that the influence exaction be effective - that it succeed in changing the recipient's preferences or her agenda to fit those of the vendor. The parties, therefore, will hide their true beliefs about whether a particular condition to using proprietary infrastructure is of a type that is likely to be effective at influencing the particular recipient. Under anything less than a hypothetical and practically unattainable perfect market in communications infrastructure services, users of a proprietary infrastructure will face a less-than-perfect menu of influence exactions that they must accept before they can communicate using owned infrastructure.

Adopting a regulatory framework under which all physical means of communication are based on private property rights in the infrastructure will therefore create a cost for users, in terms of autonomy. This cost is the autonomy deficit of exclusive reliance on proprietary models. If ownership of infrastructure is concentrated, or if owners can benefit from exerting political, personal, cultural, or social influence over others who seek access to their infrastructure, they will

impose conditions on use of the infrastructure that will satisfy their will to exert influence. If agents other than owners (advertisers, tobacco companies, the U.S. drug czar) value the ability to influence users of the infrastructure, then the influence-exaction component of the price of using the infrastructure will be sold to serve the interests of these third parties. To the extent that these influence exactions are effective, a pure private-property regime for infrastructure allows owners to constrain the autonomy of users. The owners can do this by controlling and manipulating the users' information environment to shape how they perceive their life choices in ways that make them more likely to act in a manner that the owners prefer.

The traditional progressive or social-democratic response to failures of property-based markets has been administrative regulation. In the area of communications, these responses have taken the form of access regulations - ranging from common carriage to more limited right-of-reply, fairness doctrine-type regulations. Perfect access regulation - in particular, common-carrier obligations - like a perfectly competitive market, could in principle alleviate the autonomy deficit of property. Like markets, however, actual regulation that limits the powers that go with property in infrastructure suffers from a number of limitations. First, the institutional details of the common-carriage regime can skew incentives for what types of communications will be available, and with what degree of freedom. If we learned one thing from the history of American communications policy in the twentieth century, it is that regulated entities are adept at shaping their services, pricing, and business models to take advantage of every weakness in the common-carriage regulatory system. They are even more adept at influencing the regulatory process to introduce lucrative weaknesses into the regulatory system. At present, cable broadband has succeeded in achieving a status almost entirely exempt from access requirements that might mitigate its power to control how the platform is used, and broadband over legacy telephone systems is increasingly winning a parallel status of unregulated semimonopoly. Second, the organization that owns the infrastructure retains the same internal incentives to control content as it would in the absence of common carriage and will do so to the extent that it can sneak by any imperfections in either the carriage regulations or their enforcement. Third, as long as the network is built to run through a central organizational clearinghouse, that center remains a potential point at which regulators can reassert control or delegate to owners the power to prevent unwanted speech by purposefully limiting the scope of the common-carriage requirements.

As a practical matter, then, if all wireless systems are based on property, just like the wired systems are, then wireless will offer some benefits through the introduction of some, albeit imperfect, competition. However, it will not offer the autonomy-enhancing effects that a genuine diversity of constraint can offer. If, on the other hand, policies currently being experimented with in the United States do result in the emergence of a robust, sustainable wireless communications infrastructure, owned and shared by its users and freely available to all under symmetric technical constraints, it will offer a genuinely alternative communications platform. It may be as technically good as the wired platforms for all users and uses, or it may not. Nevertheless, because of its radically distributed capitalization, and its reliance on commons rendered sustainable by equipment-embedded technical protocols, rather than on markets that depend on institutionally created asymmetric power over communications, a commons-based wireless system will offer an infrastructure that operates under genuinely different institutional constraints. Such a system can become an infrastructure of first and last resort for uses that would not fit the constraints of the proprietary market, or for users who find the price-to-influence exaction bundles offered in the market too threatening to their autonomy.

The emerging viability of commons-based strategies for the provisioning of communications, storage, and computation capacity enables us to take a practical, real world look at the autonomy

deficit of a purely property-based communications system. As we compare property to commons, we see that property, by design, introduces a series of legal powers that asymmetrically enable owners of infrastructure to exert influence over users of their systems. This asymmetry is necessary for the functioning of markets. Predictably and systematically, however, it allows one group of actors - owners - to act upon another group of actors - consumers - as objects of manipulation. No single idiom in contemporary culture captures this characteristic better than the term "the market in eyeballs," used to describe the market in advertising slots. Commons, on the other hand, do not rely on asymmetric constraints. They eliminate points of asymmetric control over the resources necessary for effective communication, thereby eliminating the legal bases of the objectification of others. These are not spaces of perfect freedom from all constraints. However, the constraints they impose are substantively different from those generated by either the property system or by an administrative regulatory system. Their introduction alongside proprietary networks therefore diversifies the constraints under which individuals operate. By offering alternative transactional frameworks for alternative information flows, these networks substantially and qualitatively increase the freedom of individuals to perceive the world through their own eyes, and to form their own perceptions of what options are open to them and how they might evaluate alternative courses of action.

Autonomy, Mass Media, and Nonmarket Information Producers

The autonomy deficit of private communications and information systems is a result of the formal structure of property as an institutional device and the role of communications and information systems as basic requirements in the ability of individuals to formulate purposes and plan actions to fit their lives. The gains flow directly from the institutional characteristics of commons. The emergence of the networked information economy makes one other important contribution to autonomy. It qualitatively diversifies the information available to individuals. Information, knowledge, and culture are now produced by sources that respond to a myriad of motivations, rather than primarily the motivation to sell into mass markets. Production is organized in any one of a myriad of productive organizational forms, rather than solely the for-profit business firm. The supplementation of the profit motive and the business organization by other motivations and organizational forms - ranging from individual play to large-scale peer-production projects - provides not only a discontinuously dramatic increase in the number of available information sources but, more significantly, an increase in available information sources that are qualitatively different from others.

Imagine three storytelling societies: the Reds, the Blues, and the Greens. Each society follows a set of customs as to how they live and how they tell stories. Among the Reds and the Blues, everyone is busy all day, and no one tells stories except in the evening. In the evening, in both of these societies, everyone gathers in a big tent, and there is one designated storyteller who sits in front of the audience and tells stories. It is not that no one is allowed to tell stories elsewhere. However, in these societies, given the time constraints people face, if anyone were to sit down in the shade in the middle of the day and start to tell a story, no one else would stop to listen. Among the Reds, the storyteller is a hereditary position, and he or she alone decides which stories to tell. Among the Blues, the storyteller is elected every night by simple majority vote. Every member of the community is eligible to offer him- or herself as that night's storyteller, and every member is eligible to vote. Among the Greens, people tell stories all day, and everywhere. Everyone tells stories. People stop and listen if they wish, sometimes in small groups of two or three, sometimes in very large groups. Stories in each of these societies play a very important role in understanding and evaluating the world. They are the way people describe the world as they know it. They serve as testing grounds to imagine how the world might be, and as a way to work out what is good and

desirable and what is bad and undesirable. The societies are isolated from each other and from any other source of information.

Now consider Ron, Bob, and Gertrude, individual members of the Reds, Blues, and Greens, respectively. Ron's perception of the options open to him and his evaluation of these options are largely controlled by the hereditary storyteller. He can try to contact the storyteller to persuade him to tell different stories, but the storyteller is the figure who determines what stories are told. To the extent that these stories describe the universe of options Ron knows about, the storyteller defines the options Ron has. The storyteller's perception of the range of options largely will determine the size and diversity of the range of options open to Ron. This not only limits the range of known options significantly, but it also prevents Ron from choosing to become a storyteller himself. Ron is subjected to the storyteller's control to the extent that, by selecting which stories to tell and how to tell them, the storyteller can shape Ron's aspirations and actions. In other words, both the freedom to be an active producer and the freedom from the control of another are constrained. Bob's autonomy is constrained not by the storyteller, but by the majority of voters among the Blues. These voters select the storyteller, and the way they choose will affect Bob's access to stories profoundly. If the majority selects only a small group of entertaining, popular, pleasing, or powerful (in some other dimension, like wealth or political power) storytellers, then Bob's perception of the range of options will be only slightly wider than Ron's, if at all. The locus of power to control Bob's sense of what he can and cannot do has shifted. It is not the hereditary storyteller, but rather the majority. Bob can participate in deciding which stories can be told. He can offer himself as a storyteller every night. He cannot, however, decide to become a storyteller independently of the choices of a majority of Blues, nor can he decide for himself what stories he will hear. He is significantly constrained by the preferences of a simple majority. Gertrude is in a very different position. First, she can decide to tell a story whenever she wants to, subject only to whether there is any other Green who wants to listen. She is free to become an active producer except as constrained by the autonomy of other individual Greens. Second, she can select from the stories that any other Green wishes to tell, because she and all those surrounding her can sit in the shade and tell a story. No one person, and no majority, determines for her whether she can or cannot tell a story. No one can unilaterally control whose stories Gertrude can listen to. And no one can determine for her the range and diversity of stories that will be available to her from any other member of the Greens who wishes to tell a story.

The difference between the Reds, on the one hand, and the Blues or Greens, on the other hand, is formal. Among the Reds, only the storyteller may tell the story as a matter of formal right, and listeners only have a choice of whether to listen to this story or to no story at all. Among the Blues and the Greens anyone may tell a story as a matter of formal right, and listeners, as a matter of formal right, may choose from whom they will hear. The difference between the Reds and the Blues, on the one hand, and the Greens, on the other hand, is economic. In the former, opportunities for storytelling are scarce. The social cost is higher, in terms of stories unavailable for hearing, or of choosing one storyteller over another. The difference between the Blues and the Greens, then, is not formal, but practical. The high cost of communication created by the Blues' custom of listening to stories only in the evening, in a big tent, together with everyone else, makes it practically necessary to select "a storyteller" who occupies an evening. Since the stories play a substantive role in individuals' perceptions of how they might live their lives, that practical difference alters the capacity of individual Blues and Greens to perceive a wide and diverse set of options, as well as to exercise control over their perceptions and evaluations of options open for living their lives and to exercise the freedom themselves to be storytellers. The range of stories Bob is likely to listen to, and the degree to which he can choose unilaterally whether he will tell or listen, and to which story, are closer, as a practical matter, to those of Ron than to those of Gertrude.

Gertrude has many more stories and storytelling settings to choose from, and many more instances where she can offer her own stories to others in her society. She, and everyone else in her society, can be exposed to a wider variety of conceptions of how life can and ought to be lived. This wider diversity of perceptions gives her greater choice and increases her ability to compose her own life story out of the more varied materials at her disposal. She can be more self-authored than either Ron or Bob. This diversity replicates, in large measure, the range of perceptions of how one might live a life that can be found among all Greens, precisely because the storytelling customs make every Green a potential storyteller, a potential source of information and inspiration about how one might live one's life.

All this could sound like a morality tale about how wonderfully the market maximizes autonomy. The Greens easily could sound like Greenbacks, rather than like environmentalists staking out public parks as information commons. However, this is not the case in the industrial information economy, where media markets have high entry barriers and large economies of scale. It is costly to start up a television station, not to speak of a network, a newspaper, a cable company, or a movie distribution system. It is costly to produce the kind of content delivered over these systems. Once production costs or the costs of laying a network are incurred, the additional marginal cost of making information available to many users, or of adding users to the network, is much smaller than the initial cost. This is what gives information and cultural products and communications facilities supply-side economies of scale and underlies the industrial model of producing them. The result is that the industrial information economy is better stylized by the Reds and Blues rather than by the Greens. While there is no formal limitation on anyone producing and disseminating information products, the economic realities limit the opportunities for storytelling in the mass-mediated environment and make storytelling opportunities a scarce good. It is very costly to tell stories in the mass-mediated environment. Therefore, most storytellers are commercial entities that seek to sell their stories to the audience. Given the discussion earlier in this chapter, it is fairly straightforward to see how the Greens represent greater freedom to choose to become an active producer of one's own information environment. It is similarly clear that they make it exceedingly difficult for any single actor to control the information flow to any other actor. We can now focus on how the story provides a way of understanding the justification and contours of the third focus of autonomy - respecting policy: the requirement that government not limit the quantity and diversity of information available.

The fact that our mass-mediated environment is mostly commercial makes it more like the Blues than the Reds. These outlets serve the tastes of the majority - expressed in some combination of cash payment and attention to advertising. I do not offer here a full analysis - covered so well by Baker in *Media, Markets, and Democracy* - as to why mass-media markets do not reflect the preferences of their audiences very well. Presented here is a tweak of an older set of analyses of whether monopoly or competition is better in mass-media markets to illustrate the relationship between markets, channels, and diversity of content. In chapter 6, I describe in greater detail the Steiner-Beebe model of diversity and number of channels. For our purposes here, it is enough to note that this model shows how advertiser-supported media tend to program lowest-common-denominator programs, intended to "capture the eyeballs" of the largest possible number of viewers. These media do not seek to identify what viewers intensely want to watch, but tend to clear programs that are tolerable enough to viewers so that they do not switch off their television. The presence or absence of smaller-segment oriented television depends on the shape of demand in an audience, the number of channels available to serve that audience, and the ownership structure. The relationship between diversity of content and diversity of structure or ownership is not smooth. It occurs in leaps. Small increases in the number of outlets continue to serve large clusters of low-intensity preferences - that is, what people find acceptable. A new channel that is

added will more often try to take a bite out of a large pie represented by some lowest-common-denominator audience segment than to try to serve a new niche market. Only after a relatively high threshold number of outlets are reached do advertiser-supported media have sufficient reason to try to capture much smaller and higher-intensity preference clusters - what people are really interested in. The upshot is that if all storytellers in society are profit maximizing and operate in a market, the number of storytellers and venues matters tremendously for the diversity of stories told in a society. It is quite possible to have very active market competition in how well the same narrow set of stories are told, as opposed to what stories are told, even though there are many people who would rather hear different stories altogether, but who are in clusters too small, too poor, or too uncoordinated to persuade the storytellers to change their stories rather than their props.

The networked information economy is departing from the industrial information economy along two dimensions that suggest a radical increase in the number of storytellers and the qualitative diversity of stories told. At the simplest level, the cost of a channel is so low that some publication capacity is becoming available to practically every person in society. Ranging from an e-mail account, to a few megabytes of hosting capacity to host a subscriber's Web site, to space on a peer-to-peer distribution network available for any kind of file (like FreeNet or eDonkey), individuals are now increasingly in possession of the basic means necessary to have an outlet for their stories. The number of channels is therefore in the process of jumping from some infinitesimally small fraction of the population - whether this fraction is three networks or five hundred channels almost does not matter by comparison - to a number of channels roughly equal to the number of users. This dramatic increase in the number of channels is matched by the fact that the low costs of communications and production enable anyone who wishes to tell a story to do so, whether or not the story they tell will predictably capture enough of a paying (or advertising-susceptible) audience to recoup production costs. Self-expression, religious fervor, hobby, community seeking, political mobilization, any one of the many and diverse reasons that might drive us to want to speak to others is now a sufficient reason to enable us to do so in mediated form to people both distant and close. The basic filter of marketability has been removed, allowing anything that emerges out of the great diversity of human experience, interest, taste, and expressive motivation to flow to and from everyone connected to everyone else. Given that all diversity within the industrial information economy needed to flow through the marketability filter, the removal of that filter marks a qualitative increase in the range and diversity of life options, opinions, tastes, and possible life plans available to users of the networked information economy.

The image of everyone being equally able to tell stories brings, perhaps more crisply than any other image, two critical objections to the attractiveness of the networked information economy: quality and cacophony. The problem of quality is easily grasped, but is less directly connected to autonomy. Having many high school plays and pickup basketball games is not the same as having Hollywood movies or the National Basketball Association (NBA). The problem of quality understood in these terms, to the extent that the shift from industrial to networked information production in fact causes it, does not represent a threat to autonomy as much as a welfare cost of making the autonomy-enhancing change. More troubling from the perspective of autonomy is the problem of information overload, which is related to, but distinct from, production quality. The cornucopia of stories out of which each of us can author our own will only enhance autonomy if it does not resolve into a cacophony of meaningless noise. How, one might worry, can a system of information production enhance the ability of an individual to author his or her life, if it is impossible to tell whether this or that particular story or piece of information is credible, or whether it is relevant to the individual's particular experience? Will individuals spend all their time sifting through mounds of inane stories and fairy tales, instead of evaluating which life is best for them based on a small and

manageable set of credible and relevant stories? None of the philosophical accounts of substantive autonomy suggests that there is a linearly increasing relationship between the number of options open to an individual - or in this case, perceivable by an individual - and that person's autonomy. Information overload and decision costs can get in the way of actually living one's autonomously selected life.

The quality problem is often raised in public discussions of the Internet, and takes the form of a question: Where will high-quality information products, like movies, come from? This form of the objection, while common, is underspecified normatively and overstated descriptively. First, it is not at all clear what might be meant by "quality," insofar as it is a characteristic of information, knowledge, and cultural production that is negatively affected by the shift from an industrial to a networked information economy. Chapter 2 explains that information has always been produced in various modalities, not only in market-oriented organizations and certainly not in proprietary strategies. Political theory is not "better" along any interesting dimension when written by someone aiming to maximize her own or her publisher's commercial profits. Most of the commercial, proprietary online encyclopedias are not better than *Wikipedia* along any clearly observable dimension. Moreover, many information and cultural goods are produced on a relational model, rather than a packaged-goods model. The emergence of the digitally networked environment does not much change their economics or sustainability. Professional theatre that depends on live performances is an example, as are musical performances. To the extent, therefore, that the emergence of substantial scope for nonmarket, distributed production in a networked information economy places pressure on "quality," it is quality of a certain kind. The threatened desiderata are those that are uniquely attractive about industrially produced mass-market products. The high-production-cost Hollywood movie or television series are the threatened species. Even that species is not entirely endangered, and the threat varies for different industries, as explained in some detail in chapter 11. Some movies, particularly those currently made for video release only, may well, in fact, recede. However, truly high-production-value movies will continue to have a business model through release windows other than home video distribution. Independently, the pressure on advertising-supported television from multichannel video - cable and satellite - on the other hand, is pushing for more low-cost productions like reality TV. That internal development in mass media, rather than the networked information economy, is already pushing industrial producers toward low-cost, low-quality productions. Moreover, as a large section of chapter 7 illustrates, peer production and nonmarket production are producing desirable public information - news and commentary - that offer qualities central to democratic discourse. Chapter 8 discusses how these two forms of production provide a more transparent and plastic cultural environment - both central to the individual's capacity for defining his or her goals and options. What emerges in the networked information environment, therefore, will not be a system for low-quality amateur mimicry of existing commercial products. What will emerge is space for much more expression, from diverse sources and of diverse qualities. Freedom - the freedom to speak, but also to be free from manipulation and to be cognizant of many and diverse options - inheres in this radically greater diversity of information, knowledge, and culture through which to understand the world and imagine how one could be.

Rejecting the notion that there will be an appreciable loss of quality in some absolute sense does not solve the deeper problem of information overload, or having too much information to be able to focus or act upon it. Having too much information with no real way of separating the wheat from the chaff forms what we might call the Babel objection. Individuals must have access to some mechanism that sifts through the universe of information, knowledge, and cultural moves in order to whittle them down to a manageable and usable scope. The question then becomes whether the networked information economy, given the human need for filtration, actually improves the

information environment of individuals relative to the industrial information economy. There are three elements to the answer: First, as a baseline, it is important to recognize the power that inheres in the editorial function. The extent to which information overload inhibits autonomy relative to the autonomy of an individual exposed to a well-edited information flow depends on how much the editor who whittles down the information flow thereby gains power over the life of the user of the editorial function, and how he or she uses that power. Second, there is the question of whether users can select and change their editor freely, or whether the editorial function is bundled with other communicative functions and sold by service providers among which users have little choice. Finally, there is the understanding that filtration and accreditation are themselves information goods, like any other, and that they too can be produced on a commons-based, nonmarket model, and therefore without incurring the autonomy deficit that a reintroduction of property to solve the Babel objection would impose.

Relevance filtration and accreditation are integral parts of all communications. A communication must be relevant for a given sender to send to a given recipient and relevant for the recipient to receive. Accreditation further filters relevant information for credibility. Decisions of filtration for purposes of relevance and accreditation are made with reference to the values of the person filtering the information, not the values of the person receiving the information. For instance, the editor of a cable network newsmagazine decides whether a given story is relevant to send out. The owner of the cable system decides whether it is, in the aggregate, relevant to its viewers to see that newsmagazine on its system. Only if both so decide, does each viewer get the residual choice of whether to view the story. Of the three decisions that must coincide to mark the newsmagazine as relevant to the viewer, only one is under the control of the individual recipient. And, while the editor's choice might be perceived in some sense as inherent to the production of the information, the cable operator's choice is purely a function of its role as proprietor of the infrastructure. The point to focus on is that the recipient's judgment is dependent on the cable operator's decision as to whether to release the program. The primary benefit of proprietary systems as mechanisms of avoiding the problem of information overload or the Babel objection is precisely the fact that the individual cannot exercise his own judgment as to all the programs that the cable operator - or other commercial intermediary between someone who makes a statement and someone who might receive it - has decided not to release.

As with any flow, control over a necessary passageway or bottleneck in the course of a communication gives the person controlling that point the power to direct the entire flow downstream from it. This power enables the provision of a valuable filtration service, which promises the recipient that he or she will not spend hours gazing at irrelevant materials. However, filtration only enhances the autonomy of users if the editor's notions of relevance and quality resemble those of the sender and the recipient. Imagine a recipient who really wants to be educated about African politics, but also likes sports. Under perfect conditions, he would seek out information on African politics most of the time, with occasional searches for information on sports. The editor, however, makes her money by selling advertising. For her, the relevant information is whatever will keep the viewer's attention most closely on the screen while maintaining a pleasantly acquisitive mood. Given a choice between transmitting information about famine in Sudan, which she worries will make viewers feel charitable rather than acquisitive, and transmitting a football game that has no similar adverse effects, she will prefer the latter. The general point should be obvious. For purposes of enhancing the autonomy of the user, the filtering and accreditation function suffers from an agency problem. To the extent that the values of the editor diverge from those of the user, an editor who selects relevant information based on her values and plans for the users does not facilitate user autonomy, but rather imposes her own preferences regarding what should be relevant to users given her decisions about their life choices.

A parallel effect occurs with accreditation. An editor might choose to treat as credible a person whose views or manner of presentation draw audiences, rather than necessarily the wisest or best-informed of commentators. The wide range in quality of talking heads on television should suffice as an example. The Babel objection may give us good reason to pause before we celebrate the networked information economy, but it does not provide us with reasons to celebrate the autonomy effects of the industrial information economy.

The second component of the response to the Babel objection has to do with the organization of filtration and accreditation in the industrial information economy. The cable operator owns its cable system by virtue of capital investment and (perhaps) expertise in laying cables, hooking up homes, and selling video services. However, it is control over the pipeline into the home that gives it the editorial role in the materials that reach the home. Given the concentrated economics of cable systems, this editorial power is not easy to replace and is not subject to open competition. The same phenomenon occurs with other media that are concentrated and where the information production and distribution functions are integrated with relevance filtration and accreditation: from one-newspaper towns to broadcasters or cable broadband service providers. An edited environment that frees the individual to think about and choose from a small selection of information inputs becomes less attractive when the editor takes on that role as a result of the ownership of carriage media, a large printing press, or copyrights in existing content, rather than as a result of selection by the user as a preferred editor or filter. The existence of an editor means that there is less information for an individual to process. It does not mean that the values according to which the information was pared down are those that the user would have chosen absent the tied relationship between editing and either proprietary content production or carriage.

Finally, and most important, just like any other form of information, knowledge, and culture, relevance and accreditation can be, and are, produced in a distributed fashion. Instead of relying on the judgment of a record label and a DJ of a commercial radio station for what music is worth listening to, users can compare notes as to what they like, and give music to friends whom they think will like it. This is the virtue of music file-sharing systems as distribution systems. Moreover, some of the most interesting experiments in peer production described in chapter 3 are focused on filtration. From the discussions of *Wikipedia* to the moderation and metamoderation scheme of Slashdot, and from the sixty thousand volunteers that make up the Open Directory Project to the PageRank system used by Google, the means of filtering data are being produced within the networked information economy using peer production and the coordinate patterns of nonproprietary production more generally. The presence of these filters provides the most important answer to the Babel objection. The presence of filters that do not depend on proprietary control, and that do not bundle proprietary content production and carriage services with filtering, offers a genuinely distinct approach toward presenting autonomous individuals with a choice among different filters that reflect genuinely diverse motivations and organizational forms of the providers.

Beyond the specific efforts at commons-based accreditation and relevance filtration, we are beginning to observe empirically that patterns of use of the Internet and the World Wide Web exhibit a significant degree of order. In chapter 7, I describe in detail and apply the literature that has explored network topology to the Babel objection in the context of democracy and the emerging networked public sphere, but its basic lesson applies here as well. In brief, the structure of linking on the Internet suggests that, even without quasi-formal collaborative filtering, the coordinate behavior of many autonomous individuals settles on an order that permits us to make sense of the tremendous flow of information that results from universal practical ability to speak and create. We observe the Web developing an order - with high-visibility nodes, and clusters of thickly connected "regions" where groups of Web sites accredit each other by mutual referencing.

The high-visibility Web sites provide points of condensation for informing individual choices, every bit as much as they form points of condensation for public discourse. The enormous diversity of topical and context-dependent clustering, whose content is nonetheless available for anyone to reach from anywhere, provides both a way of slicing through the information and rendering it comprehensible, and a way of searching for new sources of information beyond those that one interacts with as a matter of course. The Babel objection is partly solved, then, by the fact that people tend to congregate around common choices. We do this not as a result of purposeful manipulation, but rather because in choosing whether or not to read something, we probably give some weight to whether or not other people have chosen to read it. Unless one assumes that individual human beings are entirely dissimilar from each other, then the fact that many others have chosen to read something is a reasonable signal that it may be worthwhile for me to read. This phenomenon is both universal - as we see with the fact that Google successfully provides useful ranking by aggregating all judgments around the Web as to the relevance of any given Web site - and recursively present within interest-based and context-based clusters or groups. The clustering and actual degree distribution in the Web suggests, however, that people do not simply follow the herd - they will not read whatever a majority reads. Rather, they will make additional rough judgments about which other people's preferences are most likely to predict their own, or which topics to look in. From these very simple rules - other people share something with me in their tastes, and some sets of other people share more with me than others - we see the Babel objection solved on a distributed model, without anyone exerting formal legal control or practical economic power.

Why, however, is this not a simple reintroduction of heteronomy, of dependence on the judgment of others that subjects individuals to their control? The answer is that, unlike with proprietary filters imposed at bottlenecks or gateways, attention-distribution patterns emerge from many small-scale, independent choices where free choice exists. They are not easily manipulable by anyone. Significantly, the millions of Web sites that do not have high traffic do not "go out of business." As Clay Shirky puts it, while my thoughts about the weekend are unlikely to be interesting to three random users, they may well be interesting, and a basis for conversation, for three of my close friends. The fact that power law distributions of attention to Web sites result from random distributions of interests, not from formal or practical bottlenecks that cannot be worked around, means that whenever an individual chooses to search based on some mechanism other than the simplest, thinnest belief that individuals are all equally similar and dissimilar, a different type of site will emerge as highly visible. Topical sites cluster, unsurprisingly, around topical preference groups; one site does not account for all readers irrespective of their interests. We, as individuals, also go through an iterative process of assigning a likely relevance to the judgments of others. Through this process, we limit the information overload that would threaten to swamp our capacity to know; we diversify the sources of information to which we expose ourselves; and we avoid a stifling dependence on an editor whose judgments we cannot circumvent. We might spend some of our time using the most general, "human interest has some overlap" algorithm represented by Google for some things, but use political common interest, geographic or local interest, hobbyist, subject matter, or the like, to slice the universe of potential others with whose judgments we will choose to affiliate for any given search. By a combination of random searching and purposeful deployment of social mapping - who is likely to be interested in what is relevant to me now - we can solve the Babel objection while subjecting ourselves neither to the legal and market power of proprietors of communications infrastructure or media products nor to the simple judgments of the undifferentiated herd. These observations have the virtue of being not only based on rigorous mathematical and empirical studies, as we see in chapter 7, but also being more consistent with intuitive experience of anyone who has used the Internet for any decent length of time. We do not degenerate into mindless meandering through a cacophonous din. We find things we want quite

well. We stumble across things others suggest to us. When we do go on an unplanned walk, within a very short number of steps we either find something interesting or go back to looking in ways that are more self-conscious and ordered.

The core response to the Babel objection is, then, to accept that filtration is crucial to an autonomous individual. Nonetheless, that acknowledgement does not suggest that the filtration and accreditation systems that the industrial information economy has in fact produced, tied to proprietary control over content production and exchange, are the best means to protect autonomous individuals from the threat of paralysis due to information overload. Property in infrastructure and content affords control that can be used to provide filtration. To that extent, property provides the power for some people to shape the will-formation processes of others. The adoption of distributed information-production systems - both structured as cooperative peer-production enterprises and unstructured coordinate results of individual behavior, like the clustering of preferences around Web sites - does not mean that filtration and accreditation lose their importance. It only means that autonomy is better served when these communicative functions, like others, are available from a nonproprietary, open model of production alongside the proprietary mechanisms of filtration. Being autonomous in this context does not mean that we have to make all the information, read it all, and sift through it all by ourselves. It means that the combination of institutional and practical constraints on who can produce information, who can access it, and who can determine what is worth reading leaves each individual with a substantial role in determining what he shall read, and whose judgment he shall adhere to in sifting through the information environment, for what purposes, and under what circumstances. As always in the case of autonomy for context-bound individuals, the question is the relative role that individuals play, not some absolute, context-independent role that could be defined as being the condition of freedom.

The increasing feasibility of nonmarket, nonproprietary production of information, knowledge, and culture, and of communications and computation capacity holds the promise of increasing the degree of autonomy for individuals in the networked information economy. By removing basic capital and organizational constraints on individual action and effective cooperation, the networked information economy allows individuals to do more for and by themselves, and to form associations with others whose help they require in pursuing their plans. We are beginning to see a shift from the highly constrained roles of employee and consumer in the industrial economy, to more flexible, self-authored roles of user and peer participant in cooperative ventures, at least for some part of life. By providing as commons a set of core resources necessary for perceiving the state of the world, constructing one's own perceptions of it and one's own contributions to the information environment we all occupy, the networked information economy diversifies the set of constraints under which individuals can view the world and attenuates the extent to which users are subject to manipulation and control by the owners of core communications and information systems they rely on. By making it possible for many more diversely motivated and organized individuals and groups to communicate with each other, the emerging model of information production provides individuals with radically different sources and types of stories, out of which we can work to author our own lives. Information, knowledge, and culture can now be produced not only by many more people than could do so in the industrial information economy, but also by individuals and in subjects and styles that could not pass the filter of marketability in the mass-media environment. The result is a proliferation of strands of stories and of means of scanning the universe of potential stories about how the world is and how it might become, leaving individuals with much greater leeway to choose, and therefore a much greater role in weaving their own life tapestry.

Notes

1. Robert Post, "Meiklejohn's Mistake: Individual Autonomy and the Reform of Public Discourse," *University of Colorado Law Review* 64 (1993): 1109, 1130-1132.

2. This conception of property was first introduced and developed systematically by Robert Lee Hale in the 1920s and 1930s, and was more recently integrated with contemporary postmodern critiques of power by Duncan Kennedy, *Sexy Dressing Etc.: Essays on the Power and Politics of Cultural Identity* (Cambridge, MA: Harvard University Press, 1993).

3. White Paper, "Controlling Your Network, A Must for Cable Operators" (1999), http://www.cptech.org/ecom/openaccess/cisco1.html .

4. Data are all based on FCC Report on High Speed Services, Appendix to Fourth 706 Report NOI (Washington, DC: Federal Communications Commission, December 2003).

The Wealth of Networks:
How Social Production Transforms Markets and Freedom
by Yochai Benkler, Yale University Press

Chapter 6
Political Freedom Part 1: The Trouble with Mass Media

Design Characteristics of a Communications Platform for a Liberal Public Platform or a Liberal Public Sphere

The Emergence of the Commercial Mass-Media Platform for the Public Sphere

Basic Critiques of Mass Media

Epigraph

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Chapter 6
Political Freedom Part 1: The Trouble with Mass Media

Modern democracies and mass media have coevolved throughout the twentieth century. The first modern national republics - the early American Republic, the French Republic from the Revolution to the Terror, the Dutch Republic, and the early British parliamentary monarchy - preexisted mass media. They provide us with some model of the shape of the public sphere in a republic without mass media, what Jurgen Habermas called the bourgeois public sphere. However, the expansion of democracies in complex modern societies has largely been a phenomenon of the late nineteenth and twentieth centuries - in particular, the post-World War II years. During this period, the platform of the public sphere was dominated by mass media - print, radio, and television. In authoritarian regimes, these means of mass communication were controlled by the state. In democracies, they operated either under state ownership, with varying degrees of independence from the sitting government, or under private ownership financially dependent on advertising markets. We do not, therefore, have examples of complex modern democracies whose public sphere is built on a platform that is widely distributed and independent of both government control and market demands. The Internet as a technology, and the networked information economy as an organizational and social model of information and cultural production, promise the emergence of a substantial alternative platform for the public sphere. The networked public sphere, as it is currently developing, suggests that it will have no obvious points of control or exertion of influence - either by fiat or by purchase. It seems to invert the mass-media model in that it is driven heavily by what dense clusters of users find intensely interesting and engaging, rather than by what large swathes of them find mildly interesting on average. And it promises to offer a platform for engaged citizens to cooperate and provide observations and opinions, and to serve as a watchdog over society on a peer-production model.

The claim that the Internet democratizes is hardly new. "Everyone a pamphleteer" has been an iconic claim about the Net since the early 1990s. It is a claim that has been subjected to significant critique. What I offer, therefore, in this chapter and the next is not a restatement of the basic case, but a detailed analysis of how the Internet and the emerging networked information economy provide us with distinct improvements in the structure of the public sphere over the mass media. I will also explain and discuss the solutions that have emerged within the networked environment itself to some of the persistent concerns raised about democracy and the Internet: the problems of information overload, fragmentation of discourse, and the erosion of the watchdog function of the media.

For purposes of considering political freedom, I adopt a very limited definition of "public sphere." The term is used in reference to the set of practices that members of a society use to communicate about matters they understand to be of public concern and that potentially require collective action or recognition. Moreover, not even all communications about matters of potential public concern can be said to be part of the public sphere. Communications within self-contained relationships whose boundaries are defined independently of the political processes for collective action are "private," if those communications remain purely internal. Dinner-table conversations, grumblings at a bridge club, or private letters have that characteristic, if they occur in a context where they are not later transmitted across the associational boundaries to others who are not part of the family or the bridge club. Whether these conversations are, or are not, part of the public sphere depends on the actual communications practices in a given society. The same practices can become an initial step in generating public opinion in the public sphere if they are nodes in a network of

communications that do cross associational boundaries. A society with a repressive regime that controls the society-wide communications facilities nonetheless may have an active public sphere if social networks and individual mobility are sufficient to allow opinions expressed within discrete associational settings to spread throughout a substantial portion of the society and to take on political meaning for those who discuss them. The public sphere is, then, a sociologically descriptive category. It is a term for signifying how, if at all, people in a given society speak to each other in their relationship as constituents about what their condition is and what they ought or ought not to do as a political unit. This is a purposefully narrow conception of the public sphere. It is intended to focus on the effects of the networked environment on what has traditionally been understood to be political participation in a republic. I postpone consideration of a broader conception of the public sphere, and of the political nature of who gets to decide meaning and how cultural interpretations of the conditions of life and the alternatives open to a society are created and negotiated in a society until chapter 8.

The practices that define the public sphere are structured by an interaction of culture, organization, institutions, economics, and technical communications infrastructure. The technical platforms of ink and rag paper, handpresses, and the idea of a postal service were equally present in the early American Republic, Britain, and France of the late eighteenth and early nineteenth centuries. However, the degree of literacy, the social practices of newspaper reading, the relative social egalitarianism as opposed to elitism, the practices of political suppression or subsidy, and the extent of the postal system led to a more egalitarian, open public sphere, shaped as a network of smaller-scale local clusters in the United States, as opposed to the more tightly regulated and elitist national and metropolis-centered public spheres of France and Britain. The technical platforms of mass-circulation print and radio were equally available in the Soviet Union and Nazi Germany, in Britain, and in the United States in the 1930s. Again, however, the vastly different political and legal structures of the former created an authoritarian public sphere, while the latter two, both liberal public spheres, differed significantly in the business organization and economic model of production, the legal framework and the cultural practices of reading and listening - leading to the then still elitist overlay on the public sphere in Britain relative to a more populist public sphere in the United States.

Mass media structured the public sphere of the twentieth century in all advanced modern societies. They combined a particular technical architecture, a particular economic cost structure, a limited range of organizational forms, two or three primary institutional models, and a set of cultural practices typified by consumption of finished media goods. The structure of the mass media resulted in a relatively controlled public sphere - although the degree of control was vastly different depending on whether the institutional model was liberal or authoritarian - with influence over the debate in the public sphere heavily tilted toward those who controlled the means of mass communications. The technical architecture was a one-way, hub-and-spoke structure, with unidirectional links to its ends, running from the center to the periphery. A very small number of production facilities produced large amounts of identical copies of statements or communications, which could then be efficiently sent in identical form to very large numbers of recipients. There was no return loop to send observations or opinions back from the edges to the core of the architecture in the same channel and with similar salience to the communications process, and no means within the mass-media architecture for communication among the end points about the content of the exchanges. Communications among the individuals at the ends were shunted to other media-personal communications or telephones - which allowed communications among the ends. However, these edge media were either local or one-to-one. Their social reach, and hence potential political efficacy, was many orders of magnitude smaller than that of the mass media.

The economic structure was typified by high-cost hubs and cheap, ubiquitous, reception-only systems at the ends. This led to a limited range of organizational models available for production: those that could collect sufficient funds to set up a hub. These included: state-owned hubs in most countries; advertising-supported commercial hubs in some of the liberal states, most distinctly in the United States; and, particularly for radio and television, the British Broadcasting Corporation (BBC) model or hybrid models like the Canadian Broadcasting Corporation (CBC) in Canada. The role of hybrid and purely commercial, advertising-supported media increased substantially around the globe outside the United States in the last two to three decades of the twentieth century. Over the course of the century, there also emerged civil-society or philanthropy-supported hubs, like the party presses in Europe, nonprofit publications like *Consumer Reports* (later, in the United States), and, more important, public radio and television. The one-way technical architecture and the mass-audience organizational model underwrote the development of a relatively passive cultural model of media consumption. Consumers (or subjects, in authoritarian systems) at the ends of these systems would treat the communications that filled the public sphere as finished goods. These were to be treated not as moves in a conversation, but as completed statements whose addressees were understood to be passive: readers, listeners, and viewers.

The Internet's effect on the public sphere is different in different societies, depending on what salient structuring components of the existing public sphere its introduction perturbs. In authoritarian countries, it is the absence of a single or manageably small set of points of control that is placing the greatest pressure on the capacity of the regimes to control their public sphere, and thereby to simplify the problem of controlling the actions of the population. In liberal countries, the effect of the Internet operates through its implications for economic cost and organizational form. In both cases, however, the most fundamental and potentially long-standing effect that Internet communications are having is on the cultural practice of public communication. The Internet allows individuals to abandon the idea of the public sphere as primarily constructed of finished statements uttered by a small set of actors socially understood to be "the media" (whether state owned or commercial) and separated from society, and to move toward a set of social practices that see individuals as participating in a debate. Statements in the public sphere can now be seen as invitations for a conversation, not as finished goods. Individuals can work their way through their lives, collecting observations and forming opinions that they understand to be practically capable of becoming moves in a broader public conversation, rather than merely the grist for private musings.

Design Characteristics of a Communications Platform for a Liberal Public Platform or a Liberal Public Sphere

How is private opinion about matters of collective, formal, public action formed? How is private opinion communicated to others in a form and in channels that allow it to be converted into a public, political opinion, and a position worthy of political concern by the formal structures of governance of a society? How, ultimately, is such a political and public opinion converted into formal state action? These questions are central to understanding how individuals in complex contemporary societies, located at great distances from each other and possessing completely different endowments of material, intellectual, social, and formal ties and capabilities, can be citizens of the same democratic polity rather than merely subjects of a more or less responsive authority. In the idealized Athenian agora or New England town hall, the answers are simple and local. All citizens meet in the agora, they speak in a way that all relevant citizens can hear, they argue with each other, and ultimately they also constitute the body that votes and converts the opinion that emerges into a legitimate action of political authority. Of course, even in those small, locally bounded polities, things were never quite so simple. Nevertheless, the idealized version does at least

give us a set of functional characteristics that we might seek in a public sphere: a place where people can come to express and listen to proposals for agenda items - things that ought to concern us as members of a polity and that have the potential to become objects of collective action; a place where we can make and gather statements of fact about the state of our world and about alternative courses of action; where we can listen to opinions about the relative quality and merits of those facts and alternative courses of action; and a place where we can bring our own concerns to the fore and have them evaluated by others.

Understood in this way, the public sphere describes a social communication process. Habermas defines the public sphere as "a network for communicating information and points of view (i.e., opinions expressing affirmative or negative attitudes)"; which, in the process of communicating this information and these points of view, filters and synthesizes them "in such a way that they coalesce into bundles of topically specified public opinions."[/1](#) Taken in this descriptive sense, the public sphere does not relate to a particular form of public discourse that is normatively attractive from some perspective or another. It defines a particular set of social practices that are necessary for the functioning of any complex social system that includes elements of governing human beings. There are authoritarian public spheres, where communications are regimented and controlled by the government in order to achieve acquiescence and to mobilize support, rather than relying solely on force to suppress dissent and opposition. There are various forms of liberal public spheres, constituted by differences in the political and communications systems scattered around liberal democracies throughout the world. The BBC or the state-owned televisions throughout postwar Western European democracies, for example, constituted the public spheres in different ways than did the commercial mass media that dominated the American public sphere. As advertiser-supported mass media have come to occupy a larger role even in places where they were not dominant before the last quarter of the twentieth century, the long American experience with this form provides useful insight globally.

In order to consider the relative advantages and failures of various platforms for a public sphere, we need to define a minimal set of desiderata that such a platform must possess. My point is not to define an ideal set of constraints and affordances of the public sphere that would secure legitimacy or would be most attractive under one conception of democracy or another. Rather, my intention is to define a design question: What characteristics of a communications system and practices are sufficiently basic to be desired by a wide range of conceptions of democracy? With these in hand, we will be able to compare the commercial mass media and the emerging alternatives in the digitally networked environment.

Universal Intake. Any system of government committed to the idea that, in principle, the concerns of all those governed by that system are equally respected as potential proper subjects for political action and that all those governed have a say in what government should do requires a public sphere that can capture the observations of all constituents. These include at least their observations about the state of the world as they perceive and understand it, and their opinions of the relative desirability of alternative courses of action with regard to their perceptions or those of others. It is important not to confuse "universal intake" with more comprehensive ideas, such as that every voice must be heard in actual political debates, or that all concerns deserve debate and answer. Universal intake does not imply these broader requirements. It is, indeed, the role of filtering and accreditation to whittle down what the universal intake function drags in and make it into a manageable set of political discussion topics and interventions. However, the basic requirement of a public sphere is that it must in principle be susceptible to perceiving and considering the issues of anyone who believes that their condition is a matter appropriate for political consideration and collective action. The extent to which that personal judgment about

what the political discourse should be concerned with actually coincides with what the group as a whole will consider in the public sphere is a function of the filtering and accreditation functions.

Filtering for Potential Political Relevance. Not everything that someone considers to be a proper concern for collective action is perceived as such by most other participants in the political debate. A public sphere that has some successful implementation of universal intake must also have a filter to separate out those matters that are plausibly within the domain of organized political action and those that are not. What constitutes the range of plausible political topics is locally contingent, changes over time, and is itself a contested political question, as was shown most obviously by the "personal is political" feminist intellectual campaign. While it left "my dad won't buy me the candy I want" out of the realm of the political, it insisted on treating "my husband is beating me" as critically relevant in political debate. An overly restrictive filtering system is likely to impoverish a public sphere and rob it of its capacity to develop legitimate public opinion. It tends to exclude views and concerns that are in fact held by a sufficiently large number of people, or to affect people in sufficiently salient ways that they turn out, in historical context, to place pressure on the political system that fails to consider them or provide a legitimate answer, if not a solution. A system that is too loose tends to fail because it does not allow a sufficient narrowing of focus to provide the kind of sustained attention and concentration necessary to consider a matter and develop a range of public opinions on it.

Filtering for Accreditation. Accreditation is different from relevance, requires different kinds of judgments, and may be performed in different ways than basic relevance filtering. A statement like "the president has sold out space policy to Martians" is different from "my dad won't buy me the candy I want." It is potentially as relevant as "the president has sold out energy policy to oil companies." What makes the former a subject for entertainment, not political debate, is its lack of credibility. Much of the function of journalistic professional norms is to create and preserve the credibility of the professional press as a source of accreditation for the public at large. Parties provide a major vehicle for passing the filters of both relevance and accreditation. Academia gives its members a source of credibility, whose force (ideally) varies with the degree to which their statements come out of, and pertain to, their core roles as creators of knowledge through their disciplinary constraints. Civil servants in reasonably professional systems can provide a source of accreditation. Large corporations have come to play such a role, though with greater ambiguity. The emerging role of nongovernment organizations (NGOs), very often is intended precisely to preorganize opinion that does not easily pass the relevant public sphere's filters of relevance and accreditation and provide it with a voice that will. Note that accreditation of a move in political discourse is very different from accreditation of a move in, for example, academic discourse, because the objective of each system is different. In academic discourse, the fact that a large number of people hold a particular opinion ("the universe was created in seven days") does not render that opinion credible enough to warrant serious academic discussion. In political discourse, say, about public school curricula, the fact that a large number of people hold the same view and are inclined to have it taught in public schools makes that claim highly relevant and "credible." In other words, it is credible that this could become a political opinion that forms a part of public discourse with the potential to lead to public action.

Filters, both for relevance and accreditation, provide a critical point of control over the debate, and hence are extremely important design elements.

Synthesis of "Public Opinion. The communications system that offers the platform for the public sphere must also enable the synthesis of clusters of individual opinion that are sufficiently close and articulated to form something more than private opinions held by some number of individuals.

How this is done is tricky, and what counts as "public opinion" may vary among different theories of democracy. In deliberative conceptions, this might make requirements of the form of discourse. Civic republicans would focus on open deliberation among people who see their role as deliberating about the common good. Habermas would focus on deliberating under conditions that assure the absence of coercion, while Bruce Ackerman would admit to deliberation only arguments formulated so as to be neutral as among conceptions of the good. In pluralist conceptions, like John Rawls's in *Political Liberalism*, which do not seek ultimately to arrive at a common understanding but instead seek to peaceably clear competing positions as to how we ought to act as a polity, this might mean the synthesis of a position that has sufficient overlap among those who hold it that they are willing to sign on to a particular form of statement in order to get the bargaining benefits of scale as an interest group with a coherent position. That position then comes to the polls and the bargaining table as one that must be considered, overpowered, or bargained with. In any event, the platform has to provide some capacity to synthesize the finely disparate and varied versions of beliefs and positions held by actual individuals into articulated positions amenable for consideration and adoption in the formal political sphere and by a system of government, and to render them in ways that make them sufficiently salient in the overall mix of potential opinions to form a condensation point for collective action.

Independence from Government Control. The core role of the political public sphere is to provide a platform for converting privately developed observations, intuitions, and opinions into public opinions that can be brought to bear in the political system toward determining collective action. One core output of these communications is instructions to the administration sitting in government. To the extent that the platform is dependent on that same sitting government, there is a basic tension between the role of debate in the public sphere as issuing instructions to the executive and the interests of the sitting executive to retain its position and its agenda and have it ratified by the public. This does not mean that the communications system must exclude government from communicating its positions, explaining them, and advocating them. However, when it steps into the public sphere, the locus of the formation and crystallization of public opinion, the sitting administration must act as a participant in explicit conversation, and not as a platform controller that can tilt the platform in its direction.

The Emergence of the Commercial Mass-Media Platform for the Public Sphere

Throughout the twentieth century, the mass media have played a fundamental constitutive role in the construction of the public sphere in liberal democracies. Over this period, first in the United States and later throughout the world, the commercial, advertising-supported form of mass media has become dominant in both print and electronic media. Sometimes, these media have played a role that has drawn admiration as "the fourth estate." Here, the media are seen as a critical watchdog over government processes, and as a major platform for translating the mobilization of social movements into salient, and ultimately actionable, political statements. These same media, however, have also drawn mountains of derision for the power they wield, as well as fail to wield, and for the shallowness of public communication they promote in the normal course of the business of selling eyeballs to advertisers. Nowhere was this clearer than in the criticism of the large role that television came to play in American public culture and its public sphere. Contemporary debates bear the imprint of the three major networks, which in the early 1980s still accounted for 92 percent of television viewers and were turned on and watched for hours a day in typical American homes. These inspired works like Neil Postman's *Amusing Ourselves to Death* or Robert Putnam's claim, in *Bowling Alone*, that television seemed to be the primary identifiable discrete cause of the decline of American civic life. Nevertheless, whether positive or negative, variants of the mass-media model of communications have been dominant throughout the twentieth century, in

both print and electronic media. The mass-media model has been the dominant model of communications in both democracies and their authoritarian rivals throughout the period when democracy established itself, first against monarchies, and later against communism and fascism. To say that mass media were dominant is not to say that only technical systems of remote communications form the platform of the public sphere. As Theda Skocpol and Putnam have each traced in the context of the American and Italian polities, organizations and associations of personal civic involvement form an important platform for public participation. And yet, as both have recorded, these platforms have been on the decline. So "dominant" does not mean sole, but instead means overridingly important in the structuring of the public sphere. It is this dominance, not the very existence, of mass media that is being challenged by the emergence of the networked public sphere.

The roots of the contemporary industrial structure of mass media presage both the attractive and unattractive aspects of the media we see today. Pioneered by the Dutch printers of the seventeenth century, a commercial press that did not need to rely on government grants and printing contracts, or on the church, became a source of a constant flow of heterodox literature and political debate./2 However, a commercial press has always also been sensitive to the conditions of the marketplace - costs, audience, and competition. In seventeenth-century England, the Stationers' Monopoly provided its insiders enough market protection from competitors that its members were more than happy to oblige the Crown with a compliant press in exchange for monopoly. It was only after the demise of that monopoly that a genuinely political press appeared in earnest, only to be met by a combination of libel prosecutions, high stamp taxes, and outright bribery and acquisition by government./3 These, like the more direct censorship and sponsorship relationships that typified the prerevolutionary French press, kept newspapers and gazettes relatively compliant, and their distribution largely limited to elite audiences. Political dissent did not form part of a stable and independent market-based business model. As Paul Starr has shown, the evolution of the British colonies in America was different. While the first century or so of settlement saw few papers, and those mostly "authorized" gazettes, competition began to increase over the course of the eighteenth century. The levels of literacy, particularly in New England, were exceptionally high, the population was relatively prosperous, and the regulatory constraints that applied in England, including the Stamp Tax of 1712, did not apply in the colonies. As second and third newspapers emerged in cities like Boston, Philadelphia, and New York, and were no longer supported by the colonial governments through postal franchises, the public sphere became more contentious. This was now a public sphere whose voices were self-supporting, like Benjamin Franklin's *Pennsylvania Gazette*. The mobilization of much of this press during the revolutionary era, and the broad perception that it played an important role in constituting the American public, allowed the commercial press to continue to play an independent and critical role after the revolution as well, a fate not shared by the brief flowering of the press immediately after the French Revolution. A combination of high literacy and high government tolerance, but also of postal subsidies, led the new United States to have a number and diversity of newspapers unequalled anywhere else, with a higher weekly circulation by 1840 in the 17-million-strong United States than in all of Europe with its population then of 233 million. By 1830, when Tocqueville visited America, he was confronted with a widespread practice of newspaper reading - not only in towns, but in far-flung farms as well, newspapers that were a primary organizing mechanism for political association./4

This widespread development of small-circulation, mostly local, competitive commercial press that carried highly political and associational news and opinion came under pressure not from government, but from the economies of scale of the mechanical press, the telegraph, and the ever-expanding political and economic communities brought together by rail and industrialization. Harold Innis argued more than half a century ago that the increasing costs of mechanical presses,

coupled with the much-larger circulation they enabled and the availability of a flow of facts from around the world through telegraph, reoriented newspapers toward a mass-circulation, relatively low-denominator advertising medium. These internal economies, as Alfred Chandler and, later, James Beniger showed in their work, intersected with the vast increase in industrial output, which in turn required new mechanisms of demand management - in other words, more sophisticated advertising to generate and channel demand. In the 1830s, the *Sun* and *Herald* were published in New York on large-circulation scales, reducing prices to a penny a copy and shifting content from mostly politics and business news to new forms of reporting: petty crimes from the police courts, human-interest stories, and outright entertainment-value hoaxes.[/5] The start-up cost of founding such mass-circulation papers rapidly increased over the second quarter of the nineteenth century, as figure 6.1 illustrates.
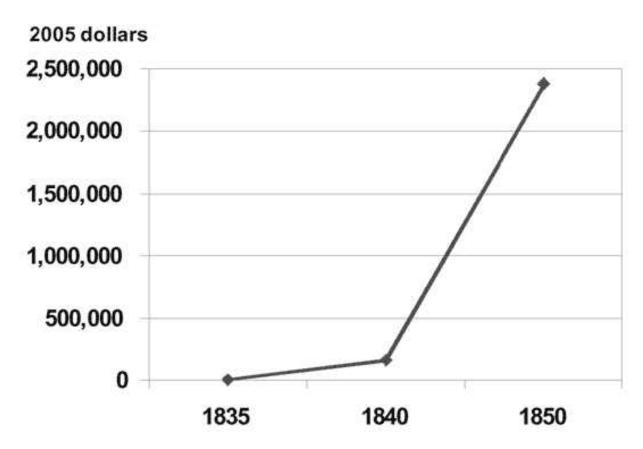
Figure 6.1: Start-up Costs of a Daily Newspaper, 1835-1850 (in 2005 dollars)

James Gordon Bennett founded the *Herald* in 1835, with an investment of five hundred dollars, equal to a little more than $10,400 in 2005 dollars. By 1840, the necessary investment was ten to twenty times greater, between five and ten thousand dollars, or $106,000-$212,000 in 2005 terms. By 1850, that amount had again grown tenfold, to $100,000, about $2.38 million in 2005.[/6] In the span of fifteen years, the costs of starting a newspaper rose from a number that many could conceive of spending for a wide range of motivations using a mix of organizational forms, to something that required a more or less industrial business model to recoup a very substantial financial investment. The new costs reflected mutually reinforcing increases in organizational cost (because of the professionalization of the newspaper publishing model) and the introduction of high-capacity,

higher-cost equipment: electric presses (1839); the Hoe double-cylinder rotary press (1846), which raised output from the five hundred to one thousand sheets per hour of the early steam presses (up from 250 sheets for the handpress) to twelve thousand sheets per hour; and eventually William Bullock's roll-fed rotary press that produced twelve thousand complete newspapers per hour by 1865. The introduction of telegraph and the emergence of news agencies - particularly the Associated Press (AP) in the United States and Reuters in England - completed the basic structure of the commercial printed press. These characteristics - relatively high cost, professional, advertising supported, dependent on access to a comparatively small number of news agencies (which, in the case of the AP, were often used to anticompetitive advantage by their members until the mid-twentieth-century antitrust case) - continued to typify print media. With the introduction of competition from radio and television, these effects tended to lead to greater concentration, with a majority of papers facing no local competition, and an ever-increasing number of papers coming under the joint ownership of a very small number of news publishing houses.

The introduction of radio was the next and only serious potential inflection point, prior to the emergence of the Internet, at which some portion of the public sphere could have developed away from the advertiser-supported mass-media model. In most of Europe, radio followed the path of state-controlled media, with variable degrees of freedom from the executive at different times and places. Britain developed the BBC, a public organization funded by government-imposed levies, but granted sufficient operational freedom to offer a genuine platform for a public sphere, as opposed to a reflection of the government's voice and agenda. While this model successfully developed what is perhaps the gold standard of broadcast journalism, it also grew as a largely elite institution throughout much of the twentieth century. The BBC model of state-based funding and monopoly with genuine editorial autonomy became the basis of the broadcast model in a number of former colonies: Canada and Australia adopted a hybrid model in the 1930s. This included a well-funded public broadcaster, but did not impose a monopoly in its favor, allowing commercial broadcasters to grow alongside it. Newly independent former colonies in the postwar era that became democracies, like India and Israel, adopted the model with monopoly, levy-based funding, and a degree of editorial independence. The most currently visible adoption of a hybrid model based on some state funding but with editorial freedom is Al Jazeera, the Arab satellite station partly funded by the Emir of Qatar, but apparently free to pursue its own editorial policy, whose coverage stands in sharp contrast to that of the state-run broadcasters in the region. In none of these BBC-like places did broadcast diverge from the basic centralized communications model of the mass media, but it followed a path distinct from the commercial mass media. Radio, and later television, was a more tightly controlled medium than was the printed press; its intake, filtering, and synthesis of public discourse were relatively insulated from the pressure of both markets, which typified the American model, and politics, which typified the state-owned broadcasters. These were instead controlled by the professional judgments of their management and journalists, and showed both the high professionalism that accompanied freedom along both those dimensions and the class and professional elite filters that typify those who control the media under that organizational model. The United States took a different path that eventually replicated, extended, and enhanced the commercial, advertiser-supported mass-media model originated in the printed press. This model was to become the template for the development of similar broadcasters alongside the state-owned and independent BBC-model channels adopted throughout much of the rest of the world, and of programming production for newer distribution technologies, like cable and satellite stations. The birth of radio as a platform for the public sphere in the United States was on election night in 1920./7 Two stations broadcast the election returns as their launchpad for an entirely new medium-wireless broadcast to a wide audience. One was the Detroit News amateur station, 8MK, a broadcast that was framed and understood as an internal communication of a technical fraternity - the many amateurs who had been trained in radio communications for World War I and who then

came to form a substantial and engaged technical community. The other was KDKA Pittsburgh, launched by Westinghouse as a bid to create demand for radio receivers of a kind that it had geared up to make during the war. Over the following four or five years, it was unclear which of these two models of communication would dominate the new medium. By 1926, however, the industrial structure that would lead radio to follow the path of commercial, advertiser-supported, concentrated mass media, dependent on government licensing and specializing in influencing its own regulatory oversight process was already in place.

Although this development had its roots in the industrial structure of radio production as it emerged from the first two decades of innovation and businesses in the twentieth century, it was shaped significantly by political-regulatory choices during the 1920s. At the turn of the twentieth century, radio was seen exclusively as a means of wireless telegraphy, emphasizing ship-to-shore and ship-to-ship communications. Although some amateurs experimented with voice programs, broadcast was a mode of point-to-point communications; entertainment was not seen as its function until the 1920s. The first decade and a half of radio in the United States saw rapid innovation and competition, followed by a series of patent suits aimed to consolidate control over the technology. By 1916, the ideal transmitter based on technology available at the time required licenses of patents held by Marconi, AT&T, General Electric (GE), and a few individuals. No licenses were in fact granted. The industry had reached stalemate. When the United States joined the war, however, the navy moved quickly to break the stalemate, effectively creating a compulsory cross-licensing scheme for war production, and brought in Westinghouse, the other major potential manufacturer of vacuum tubes alongside GE, as a participant in the industry. The two years following the war saw intervention by the U.S. government to assure that American radio industry would not be controlled by British Marconi because of concerns in the navy that British control over radio would render the United States vulnerable to the same tactic Britain used against Germany at the start of the war - cutting off all transoceanic telegraph communications. The navy brokered a deal in 1919 whereby a new company was created - the Radio Corporation of America (RCA) - which bought Marconi's American business. By early 1920, RCA, GE, and AT&T entered into a patent cross-licensing model that would allow each to produce for a market segment: RCA would control transoceanic wireless telegraphy, while GE and AT&T's Western Electric subsidiary would make radio transmitters and sell them under the RCA brand. This left Westinghouse with production facilities developed for the war, but shut out of the existing equipment markets by the patent pool. Launching KDKA Pittsburgh was part of its response: Westinghouse would create demand for small receivers that it could manufacture without access to the patents held by the pool. The other part of its strategy consisted of acquiring patents that, within a few months, enabled Westinghouse to force its inclusion in the patent pool, redrawing the market division map to give Westinghouse 40 percent of the receiving equipment market. The first part of Westinghouse's strategy, adoption of broadcasting to generate demand for receivers, proved highly successful and in the long run more important. Within two years, there were receivers in 10 percent of American homes. Throughout the 1920s, equipment sales were big business.

Radio stations, however, were not dominated by the equipment manufacturers, or by anyone else for that matter, in the first few years. While the equipment manufacturers did build powerful stations like KDKA Pittsburgh, WJZ Newark, KYW Chicago (Westinghouse), and WGY Schenectady (GE), they did not sell advertising, but rather made their money from equipment sales. These stations did not, in any meaningful sense of the word, dominate the radio sphere in the first few years of radio, as the networks would indeed come to do within a decade. In November 1921, the first five licenses were issued by the Department of Commerce under the new category of "broadcasting" of "news, lectures, entertainment, etc." Within eight months, the department had issued another 453 licenses. Many of these went to universities, churches, and unions, as well as

local shops hoping to attract business with their broadcasts. Universities, seeing radio as a vehicle for broadening their role, began broadcasting lectures and educational programming. Seventy-four institutes of higher learning operated stations by the end of 1922. The University of Nebraska offered two-credit courses whose lectures were transmitted over the air. Churches, newspapers, and department stores each forayed into this new space, much as we saw the emergence of Web sites for every organization over the course of the mid-1990s. Thousands of amateurs were experimenting with technical and format innovations. While receivers were substantially cheaper than transmitters, it was still possible to assemble and sell relatively cheap transmitters, for local communications, at prices sufficiently low that thousands of individual amateurs could take to the air. At this point in time, then, it was not yet foreordained that radio would follow the mass-media model, with a small number of well-funded speakers and hordes of passive listeners. Within a short period, however, a combination of technology, business practices, and regulatory decisions did in fact settle on the model, comprised of a small number of advertiser-supported national networks, that came to typify the American broadcast system throughout most of the rest of the century and that became the template for television as well.

Herbert Hoover, then secretary of commerce, played a pivotal role in this development. Throughout the first few years after the war, Hoover had positioned himself as the champion of making control over radio a private market affair, allying himself both with commercial radio interests and with the amateurs against the navy and the postal service, each of which sought some form of nationalization of radio similar to what would happen more or less everywhere else in the world. In 1922, Hoover assembled the first of four annual radio conferences, representing radio manufacturers, broadcasters, and some engineers and amateurs. This forum became Hoover's primary stage. Over the next four years, he used its annual meeting to derive policy recommendations, legitimacy, and cooperation for his regulatory action, all without a hint of authority under the Radio Act of 1912. Hoover relied heavily on the rhetoric of public interest and on the support of amateurs to justify his system of private broadcasting coordinated by the Department of Commerce. From 1922 on, however, he followed a pattern that would systematically benefit large commercial broadcasters over small ones; commercial broadcasters over educational and religious broadcasters; and the one-to-many broadcasts over the point-to-point, small-scale wireless telephony and telegraphy that the amateurs were developing. After January 1922, the department inserted a limitation on amateur licenses, excluding from their coverage the broadcast of "weather reports, market reports, music, concerts, speeches, news or similar information or entertainment." This, together with a Department of Commerce order to all amateurs to stop broadcasting at 360 meters (the wave assigned broadcasting), effectively limited amateurs to shortwave radiotelephony and telegraphy in a set of frequencies then thought to be commercially insignificant. In the summer, the department assigned broadcasters, in addition to 360 meters, another band, at 400 meters. Licenses in this Class B category were reserved for transmitters operating at power levels of 500-1,000 watts, who did not use phonograph records. These limitations on Class B licenses made the newly created channel a feasible home only to broadcasters who could afford the much-more-expensive, high-powered transmitters and could arrange for live broadcasts, rather than simply play phonograph records. The success of this new frequency was not immediate, because many receivers could not tune out stations broadcasting at the two frequencies in order to listen to the other. Hoover, failing to move Congress to amend the radio law to provide him with the power necessary to regulate broadcasting, relied on the recommendations of the Second Radio Conference in 1923 as public support for adopting a new regime, and continued to act without legislative authority. He announced that the broadcast band would be divided in three: high-powered (500-1,000 watts) stations serving large areas would have no interference in those large areas, and would not share frequencies. They would transmit on frequencies between 300 and 545 meters. Medium-powered stations served smaller areas without

interference, and would operate at assigned channels between 222 and 300 meters. The remaining low-powered stations would not be eliminated, as the bigger actors wanted, but would remain at 360 meters, with limited hours of operation and geographic reach. Many of these lower-powered broadcasters were educational and religious institutions that perceived Hoover's allocation as a preference for the RCA-GE-AT&T-Westinghouse alliance. Despite his protestations against commercial broadcasting ("If a speech by the President is to be used as the meat in a sandwich of two patent medicine advertisements, there will be no radio left"), Hoover consistently reserved clear channels and issued high-power licenses to commercial broadcasters. The final policy action based on the radio conferences came in 1925, when the Department of Commerce stopped issuing licenses. The result was a secondary market in licenses, in which some religious and educational stations were bought out by commercial concerns. These purchases further gravitated radio toward commercial ownership. The licensing preference for stations that could afford high-powered transmitters, long hours of operation, and compliance with high technical constraints continued after the Radio Act of 1927. As a practical matter, it led to assignment of twenty-one out of the twenty-four clear channel licenses created by the Federal Radio Commission to the newly created network-affiliated stations.

Over the course of this period, tensions also began to emerge within the patent alliance. The phenomenal success of receiver sales tempted Western Electric into that market. In the meantime, AT&T, almost by mistake, began to challenge GE, Westinghouse, and RCA in broadcasting as an outgrowth of its attempt to create a broadcast common-carriage facility. Despite the successes of broadcast and receiver sales, it was not clear in 1922-1923 how the cost of setting up and maintaining stations would be paid for. In England, a tax was levied on radio sets, and its revenue used to fund the BBC. No such proposal was considered in the United States, but the editor of *Radio Broadcast* proposed a national endowed fund, like those that support public libraries and museums, and in 1924, a committee of New York businessmen solicited public donations to fund broadcasters (the response was so pitiful that the funds were returned to their donors). AT&T was the only company to offer a solution. Building on its telephone service experience, it offered radio telephony to the public for a fee. Genuine wireless telephony, even mobile telephony, had been the subject of experimentation since the second decade of radio, but that was not what AT&T offered. In February 1922, AT&T established WEAF in New York, a broadcast station over which AT&T was to provide no programming of its own, but instead would enable the public or program providers to pay on a per-time basis. AT&T treated this service as a form of wireless telephony so that it would fall, under the patent alliance agreements of 1920, under the exclusive control of AT&T. RCA, Westinghouse, and GE could not compete in this area. "Toll broadcasting" was not a success by its own terms. There was insufficient demand for communicating with the public to sustain a full schedule that would justify listeners tuning into the station. As a result, AT&T produced its own programming. In order to increase the potential audience for its transmissions while using its advantage in wired facilities, AT&T experimented with remote transmissions, such as live reports from sports events, and with simultaneous transmissions of its broadcasts by other stations, connected to its New York feed by cable. In its effort to launch toll broadcasting, AT&T found itself by mid-1923 with the first functioning precursor to an advertiser-supported broadcast network.

The alliance members now threatened each other: AT&T threatened to enter into receiver manufacturing and broadcast, and the RCA alliance, with its powerful stations, threatened to adopt "toll broadcasting," or advertiser-supported radio. The patent allies submitted their dispute to an arbitrator, who was to interpret the 1920 agreements, reached at a time of wireless telegraphy, to divide the spoils of the broadcast world of 1924. In late 1924, the arbitrator found for RCA-GE-Westinghouse on almost all issues. Capitalizing on RCA's difficulties with the antitrust

authorities and congressional hearings over aggressive monopolization practices in the receiving set market, however, AT&T countered that if the 1920 agreements meant what the arbitrator said they meant, they were a combination in restraint of trade to which AT&T would not adhere. Bargaining in the shadow of the mutual threats of contract and antitrust actions, the former allies reached a solution that formed the basis of future radio broadcasting. AT&T would leave broadcasting. A new company, owned by RCA, GE, and Westinghouse would be formed, and would purchase AT&T's stations. The new company would enter into a long-term contract with AT&T to provide the long-distance communications necessary to set up the broadcast network that David Sarnoff envisioned as the future of broadcast. This new entity would, in 1926, become the National Broadcasting Company (NBC). AT&T's WEAF station would become the center of one of NBC's two networks, and the division arrived at would thereafter form the basis of the broadcast system in the United States.

By the middle of 1926, then, the institutional and organizational elements that became the American broadcast system were, to a great extent, in place. The idea of government monopoly over broadcasting, which became dominant in Great Britain, Europe, and their former colonies, was forever abandoned. The idea of a private-property regime in spectrum, which had been advocated by commercial broadcasters to spur investment in broadcast, was rejected on the backdrop of other battles over conservation of federal resources. The Radio Act of 1927, passed by Congress in record speed a few months after a court invalidated Hoover's entire regulatory edifice as lacking legal foundation, enacted this framework as the basic structure of American broadcast. A relatively small group of commercial broadcasters and equipment manufacturers took the lead in broadcast development. A governmental regulatory agency, using a standard of "the public good," allocated frequency, time, and power assignments to minimize interference and to resolve conflicts. The public good, by and large, correlated to the needs of commercial broadcasters and their listeners. Later, the broadcast networks supplanted the patent alliance as the primary force to which the Federal Radio Commission paid heed. The early 1930s still saw battles over the degree of freedom that these networks had to pursue their own commercial interests, free of regulation (studied in Robert McChesney's work).[8] By that point, however, the power of the broadcasters was already too great to be seriously challenged. Interests like those of the amateurs, whose romantic pioneering mantle still held strong purchase on the process, educational institutions, and religious organizations continued to exercise some force on the allocation and management of the spectrum. However, they were addressed on the periphery of the broadcast platform, leaving the public sphere to be largely mediated by a tiny number of commercial entities running a controlled, advertiser-supported platform of mass media. Following the settlement around radio, there were no more genuine inflection points in the structure of mass media. Television followed radio, and was even more concentrated. Cable networks and satellite networks varied to some extent, but retained the basic advertiser-supported model, oriented toward luring the widest possible audience to view the advertising that paid for the programming.

Basic Critiques of Mass Media

The cluster of practices that form the mass-media model was highly conducive to social control in authoritarian countries. The hub-and-spoke technical architecture and unidirectional endpoint-reception model of these systems made it very simple to control, by controlling the core - the state-owned television, radio, and newspapers. The high cost of providing high-circulation statements meant that subversive publications were difficult to make and communicate across large distances and to large populations of potential supporters. Samizdat of various forms and channels have existed in most if not all authoritarian societies, but at great disadvantage relative to public communication. The passivity of readers, listeners, and viewers coincided nicely with the role of the

authoritarian public sphere - to manage opinion in order to cause the widest possible willing, or at least quiescent, compliance, and thereby to limit the need for using actual repressive force.

In liberal democracies, the same technical and economic cost characteristics resulted in a very different pattern of communications practices. However, these practices relied on, and took advantage of, some of the very same basic architectural and cost characteristics. The practices of commercial mass media in liberal democracies have been the subject of a vast literature, criticizing their failures and extolling their virtues as a core platform for the liberal public sphere. There have been three primary critiques of these media: First, their intake has been seen as too limited. Too few information collection points leave too many views entirely unexplored and unrepresented because they are far from the concerns of the cadre of professional journalists, or cannot afford to buy their way to public attention. The debates about localism and diversity of ownership of radio and television stations have been the clearest policy locus of this critique in the United States. They are based on the assumption that local and socially diverse ownership of radio stations will lead to better representation of concerns as they are distributed in society. Second, concentrated mass media has been criticized as giving the owners too much power - which they either employ themselves or sell to the highest bidder - over what is said and how it is evaluated. Third, the advertising-supported media needs to attract large audiences, leading programming away from the genuinely politically important, challenging, and engaging, and toward the titillating or the soothing. This critique has emphasized the tension between business interests and journalistic ethics, and the claims that market imperatives and the bottom line lead to shoddy or cowering reporting; quiescence in majority tastes and positions in order to maximize audience; spectacle rather than substantive conversation of issues even when political matters are covered; and an emphasis on entertainment over news and analysis.

Three primary defenses or advantages have also been seen in these media: first is their independence from government, party, or upper-class largesse, particularly against the background of the state-owned media in authoritarian regimes, and given the high cost of production and communication, commercial mass media have been seen as necessary to create a public sphere grounded outside government. Second is the professionalism and large newsrooms that commercial mass media can afford to support to perform the watchdog function in complex societies. Because of their market-based revenues, they can replace universal intake with well-researched observations that citizens would not otherwise have made, and that are critical to a well-functioning democracy. Third, their near-universal visibility and independence enable them to identify important issues percolating in society. They can provide a platform to put them on the public agenda. They can express, filter, and accredit statements about these issues, so that they become well-specified subjects and feasible objects for public debate among informed citizens. That is to say, the limited number of points to which all are tuned and the limited number of "slots" available for speaking on these media form the basis for providing the synthesis required for public opinion and raising the salience of matters of public concern to the point of potential collective action. In the remainder of this chapter, I will explain the criticisms of the commercial mass media in more detail. I then take up in chapter 7 the question of how the Internet in general, and the rise of nonmarket and cooperative individual production in the networked information economy in particular, can solve or alleviate those problems while fulfilling some of the important roles of mass media in democracies today.

Mass Media as a Platform for the Public Sphere

The structure of mass media as a mode of communications imposes a certain set of basic characteristics on the kind of public conversation it makes possible. First, it is always

communication from a small number of people, organized into an even smaller number of distinct outlets, to an audience several orders of magnitude larger, unlimited in principle in its membership except by the production capacity of the media itself - which, in the case of print, may mean the number of copies, and in radio, television, cable, and the like, means whatever physical - reach constraints, if any, are imposed by the technology and business organizational arrangements used by these outlets. In large, complex, modern societies, no one knows everything. The initial function of a platform for the public sphere is one of intake - taking into the system the observations and opinions of as many members of society as possible as potential objects of public concern and consideration. The radical difference between the number of intake points the mass media have and the range and diversity of human existence in large complex societies assures a large degree of information loss at the intake stage. Second, the vast difference between the number of speakers and the number of listeners, and the finished-goods style of mass-media products, imposes significant constraints on the extent to which these media can be open to feedback - that is, to responsive communications that are tied together as a conversation with multiple reciprocal moves from both sides of the conversation. Third, the immense and very loosely defined audience of mass media affects the filtering and synthesis functions of the mass media as a platform for the public sphere. One of the observations regarding the content of newspapers in the late eighteenth to mid-nineteenth centuries was the shift they took as their circulation increased - from party-oriented, based in relatively thick communities of interest and practice, to fact- and sensation-oriented, with content that made thinner requirements on their users in order to achieve broader and more weakly defined readership. Fourth, and finally, because of the high costs of organizing these media, the functions of intake, sorting for relevance, accrediting, and synthesis are all combined in the hands of the same media operators, selected initially for their capacity to pool the capital necessary to communicate the information to wide audiences. While all these functions are necessary for a usable public sphere, the correlation of capacity to pool capital resources with capacity to offer the best possible filtering and synthesis is not obvious. In addition to basic structural constraints that come from the characteristic of a communications modality that can properly be called "mass media," there are also critiques that arise more specifically from the business models that have characterized the commercial mass media over the course of most of the twentieth century. Media markets are relatively concentrated, and the most common business model involves selling the attention of large audiences to commercial advertisers.

Media Concentration: The Power of Ownership and Money

The Sinclair Broadcast Group is one of the largest owners of television broadcast stations in the United States. The group's 2003 Annual Report proudly states in its title, "Our Company. Your Message. 26 Million Households"; that is, roughly one quarter of U.S. households. Sinclair owns and operates or provides programming and sales to sixty-two stations in the United States, including multiple local affiliates of NBC, ABC, CBS, and Fox. In April 2004, ABC News's program *Nightline* dedicated a special program to reading the names of American service personnel who had been killed in the Iraq War. The management of Sinclair decided that its seven ABC affiliates would not air the program, defending its decision because the program "appears to be motivated by a political agenda designed to undermine the efforts of the United States in Iraq."[/9] At the time, the rising number of American casualties in Iraq was already a major factor in the 2004 presidential election campaign, and both ABC's decision to air the program, and Sinclair's decision to refuse to carry it could be seen as interventions by the media in setting the political agenda and contributing to the public debate. It is difficult to gauge the politics of a commercial organization, but one rough proxy is political donations. In the case of Sinclair, 95 percent of the donations made by individuals associated with the company during the 2004 election cycle went to Republicans, while only 5 percent went to Democrats.[/10] Contributions of Disney, on the other hand, the owner of the ABC

network, split about seventy-thirty in favor of contribution to Democrats. It is difficult to parse the extent to which political leanings of this sort are personal to the executives and professional employees who make decisions about programming, and to what extent these are more organizationally self-interested, depending on the respective positions of the political parties on the conditions of the industry's business. In some cases, it is quite obvious that the motives are political. When one looks, for example, at contributions by Disney's film division, they are distributed 100 percent in favor of Democrats. This mostly seems to reflect the large contributions of the Weinstein brothers, who run the semi-independent studio Miramax, which also distributed Michael Moore's politically explosive criticism of the Bush administration, *Fahrenheit 9/11*, in 2004. Sinclair's contributions were aligned with, though more skewed than, those of the National Association of Broadcasters political action committee, which were distributed 61 percent to 39 percent in favor of Republicans. Here the possible motivation is that Republicans have espoused a regulatory agenda at the Federal Communications Commission that allows broadcasters greater freedom to consolidate and to operate more as businesses and less as public trustees.

The basic point is not, of course, to trace the particular politics of one programming decision or another. It is the relative power of those who manage the mass media when it so dominates public discourse as to shape public perceptions and public debate. This power can be brought to bear throughout the components of the platform, from the intake function (what facts about the world are observed) to the filtration and synthesis (the selection of materials, their presentation, and the selection of who will debate them and in what format). These are all central to forming the agenda that the public perceives, choreographing the discussion, the range of opinions perceived and admitted into the conversation, and through these, ultimately, choreographing the perceived consensus and the range of permissible debate. One might think of this as "the Berlusconi effect." Thinking in terms of a particular individual, known for a personal managerial style, who translated the power of control over media into his election as prime minister of his country symbolizes well the concern, but of course does not exhaust the problem, which is both broader and more subtle than the concern with the possibility that mass media will be owned by individuals who would exert total control over these media and translate their control into immediate political power, manufacturing and shaping the appearance of a public sphere, rather than providing a platform for one.

The power of the commercial mass media depends on the degree of concentration in mass-media markets. A million equally watched channels do not exercise power. Concentration is a common word used to describe the power media exercise when there are only few outlets, but a tricky one because it implies two very distinct phenomena. The first is a lack of competition in a market, to a degree sufficient to allow a firm to exercise power over its pricing. This is the antitrust sense. The second, very different concern might be called "mindshare." That is, media is "concentrated" when a small number of media firms play a large role as the channel from and to a substantial majority of readers, viewers, and listeners in a given politically relevant social unit.

If one thinks that commercial firms operating in a market will always "give the audience what it wants" and that what the audience wants is a fully representative cross-section of all observations and opinions relevant to public discourse, then the antitrust sense would be the only one that mattered. A competitive market would force any market actor simply to reflect the range of available opinions actually held in the public. Even by this measure, however, there continue to be debates about how one should define the relevant market and what one is measuring. The more one includes all potential nationally available sources of information, newspapers, magazines, television, radio, satellite, cable, and the like, the less concentrated the market seems. However, as Eli Noam's recent work on local media concentration has argued, treating a tiny television station

on Long Island as equivalent to WCBS in New York severely underrepresents the power of mass media over their audience. Noam offered the most comprehensive analysis currently available of the patterns of concentration where media are actually accessed-locally, where people live - from 1984 to 2001-2002. Most media are consumed locally - because of the cost of national distribution of paper newspapers, and because of the technical and regulatory constraints on nationwide distribution of radio and television. Noam computed two measures of market concentration for each of thirty local markets: the Herfindahl-Hirschman Index (HHI), a standard method used by the Department of Justice to measure market concentration for antitrust purposes; and what he calls a C4 index - that is, the market share of the top four firms in a market, and C1, the share of the top single firm in the market. He found that, based on the HHI index, all the local media markets are highly concentrated. In the standard measure, a market with an index of less than 1,000 is not concentrated, a market with an index of 1,000-1,800 is moderately concentrated, and a market with an index of above 1,800 on the HHI is highly concentrated. Noam found that local radio, which had an index below 1,000 between 1984 and 1992, rose over the course of the following years substantially. Regulatory restrictions were loosened over the course of the 1990s, resulting by the end of the decade in an HHI index measure of 2,400 for big cities, and higher for medium-sized and small markets. And yet, radio is less concentrated than local multichannel television (cable and satellite) with an HHI of 6,300, local magazines with an HHI of 6,859, and local newspapers with an HHI of 7,621. The only form of media whose concentration has declined to less than highly concentrated (HHI 1,714) is local television, as the rise of new networks and local stations' viability on cable has moved us away from the three-network world of 1984. It is still the case, however, that the top four television stations capture 73 percent of the viewers in most markets, and 62 percent in large markets. The most concentrated media in local markets are newspapers, which, except for the few largest markets, operate on a one-newspaper town model. C1 concentration has grown in this area to 83 percent of readership for the leading papers, and an HHI of 7,621.

The degree of concentration in media markets supports the proposition that owners of media can either exercise power over the programming they provide or what they write, or sell their power over programming to those who would like to shape opinions. Even if one were therefore to hold the Pollyannaish view that market-based media in a competitive market would be constrained by competition to give citizens what they need, as Ed Baker put it, there is no reason to think the same in these kinds of highly concentrated markets. As it turns out, a long tradition of scholarship has also developed the claim that even without such high levels of concentration in the antitrust sense, advertiser-supported media markets are hardly good mechanisms for assuring that the contents of the media provide a good reflection of the information citizens need to know as members of a polity, the range of opinions and views about what ought to occupy the public, and what solutions are available to those problems that are perceived and discussed./11 First, we have long known that advertiser-supported media suffer from more or less well-defined failures, purely as market mechanisms, at representing the actual distribution of first-best preferences of audiences. As I describe in more detail in the next section, whether providers in any market structure, from monopoly to full competition, will even try to serve first-best preferences of their audience turns out to be a function of the distribution of actual first-best and second-best preferences, and the number of "channels." Second, there is a systematic analytic problem with defining consumer demand for information. Perfect information is a precondition to an efficient market, not its output. In order for consumers to value information or an opinion fully, they must know it and assimilate it to their own worldview and understanding. However, the basic problem to be solved by media markets is precisely to select which information people will value if they in fact come to know it, so it is impossible to gauge the value of a unit of information before it has been produced, and hence to base production decisions on actual existing user preferences. The result is that, even

if media markets were perfectly competitive, a substantial degree of discretion and influence would remain in the hands of commercial media owners.

The actual cultural practice of mass-media production and consumption is more complex than either the view of "efficient media markets" across the board or the general case against media concentration and commercialism. Many of the relevant companies are public companies, answerable to at least large institutional shareholders, and made up of managements that need not be monolithic in their political alignment or judgment as to the desirability of making political gains as opposed to market share. Unless there is economic or charismatic leadership of the type of a William Randolph Hearst or a Rupert Murdoch, organizations usually have complex structures, with varying degrees of freedom for local editors, reporters, and midlevel managers to tug and pull at the fabric of programming. Different media companies also have different business models, and aim at different market segments. The *New York Times*, *Wall Street Journal*, and *Washington Post* do not aim at the same audience as most daily local newspapers in the United States. They are aimed at elites, who want to buy newspapers that can credibly claim to embody highly professional journalism. This requires separation of editorial from business decisions - at least for some segments of the newspapers that are critical in attracting those readers. The degree to which the Berlusconi effect in its full-blown form of individual or self-consciously directed political power through shaping of the public sphere will apply is not one that can necessarily be answered as a matter of a priori theoretical framework for all mass media. Instead, it is a concern, a tendency, whose actual salience in any given public sphere or set of firms is the product of historical contingency, different from one country to another and one period to another. It will depend on the strategies of particular companies and their relative mindshare in a society. However, it is clear and structurally characteristic of mass media that a society that depends for its public sphere on a relatively small number of actors, usually firms, to provide most of the platform of its public sphere, is setting itself up for, at least, a form of discourse elitism. In other words, those who are on the inside of the media will be able to exert substantially greater influence over the agenda, the shape of the conversation, and through these the outcomes of public discourse, than other individuals or groups in society. Moreover, for commercial organizations, this power could be sold - and as a business model, one should expect it to be. The most direct way to sell influence is explicit political advertising, but just as we see "product placement" in movies as a form of advertising, we see advertiser influence on the content of the editorial materials. Part of this influence is directly substantive and political. Another is the source of the second critique of commercial mass media.

Commercialism, Journalism, and Political Inertness

The second cluster of concerns about the commercial mass media is the degree to which their commercialism undermines their will and capacity to provide a platform for public, politically oriented discourse. The concern is, in this sense, the opposite of the concern with excessive power. Rather than the fear that the concentrated mass media will exercise its power to pull opinion in its owners' interest, the fear is that the commercial interests of these media will cause them to pull content away from matters of genuine political concern altogether. It is typified in a quote offered by Ben Bagdikian, attributed to W. R. Nelson, publisher of the *Kansas City Star* in 1915: "Newspapers are read at the breakfast table and dinner tables. God's great gift to man is appetite. Put nothing in the paper that will destroy it."[12] Examples abound, but the basic analytic structure of the claim is fairly simple and consists of three distinct components. First, advertiser-supported media need to achieve the largest audience possible, not the most engaged or satisfied audience possible. This leads such media to focus on lowest-common-denominator programming and materials that have broad second-best appeal, rather than trying to tailor their programming to the true first-best preferences of well-defined segments of the audience. Second, issues of genuine

public concern and potential political contention are toned down and structured as a performance between iconic representations of large bodies of opinion, in order to avoid alienating too much of the audience. This is the reemergence of spectacle that Habermas identified in *The Transformation of the Public Sphere*. The tendency toward lowest-common-denominator programming translates in the political sphere into a focus on fairly well-defined, iconic views, and to avoidance of genuinely controversial material, because it is easier to lose an audience by offending its members than by being only mildly interesting. The steady structuring of the media as professional, commercial, and one way over 150 years has led to a pattern whereby, when political debate is communicated, it is mostly communicated as performance. Someone represents a party or widely known opinion, and is juxtaposed with others who similarly represent alternative widely known views. These avatars of public opinion then enact a clash of opinion, orchestrated in order to leave the media neutral and free of blame, in the eyes of their viewers, for espousing an offensively partisan view. Third, and finally, this business logic often stands in contradiction to journalistic ethic. While there are niche markets for high-end journalism and strong opinion, outlets that serve those markets are specialized. Those that cater to broader markets need to subject journalistic ethic to business necessity, emphasizing celebrities or local crime over distant famines or a careful analysis of economic policy.

The basic drive behind programming choices in advertising-supported mass media was explored in the context of the problem of "program diversity" and competition. It relies on a type of analysis introduced by Peter Steiner in 1952. The basic model argued that advertiser-supported media are sensitive only to the number of viewers, not the intensity of their satisfaction. This created an odd situation, where competitors would tend to divide among them the largest market segments, and leave smaller slices of the audience unserved, whereas a monopolist would serve each market segment, in order of size, until it ran out of channels. Because it has no incentive to divide all the viewers who want, for example, sitcoms, among two or more stations, a monopolist would program a sitcom on one channel, and the next-most-desired program on the next channel. Two competitors, on the other hand, would both potentially program sitcoms, if dividing those who prefer sitcoms in half still yields a larger total audience size than airing the next-most-desired program. To illustrate this effect with a rather extreme hypothetical example, imagine that we are in a television market of 10 million viewers. Suppose that the distribution of preferences in the audience is as follows: 1,000,000 want to watch sitcoms; 750,000 want sports; 500,000 want local news; 250,000 want action movies; 9,990 are interested in foreign films; and 9,980 want programs on gardening. The stark drop-off between action movies and foreign films and gardening is intended to reflect the fact that the 7.5 million potential viewers who do not fall into one of the first four clusters are distributed in hundreds of small clusters, none commanding more than 10,000 viewers. Before we examine why this extreme assumption is likely correct, let us first see what happens if it were. Table 6.1 presents the programming choices that would typify those of competing channels, based on the number of channels competing and the distribution of preferences in the audience. It reflects the assumptions that each programmer wants to maximize the number of viewers of its channel and that the viewers are equally likely to watch one channel as another if both offer the same type of programming. The numbers in parentheses next to the programming choice represent the number of viewers the programmer can hope to attract given these assumptions, not including the probability that some of the 7.5 million viewers outside the main clusters will also tune in. In this extreme example, one would need a system with more than 250 channels in order to start seeing something other than sitcoms, sports, local news, and action movies. Why, however, is such a distribution likely, or even plausible? The assumption is not intended to represent an actual distribution of what people most prefer to watch. Rather, it reflects the notion that many people have best preferences, fallback preferences, and tolerable options. Their first-best preferences reflect what they really want to watch, and people are highly diverse in

this dimension. Their fallback and tolerable preferences reflect the kinds of things they would be willing to watch if nothing else is available, rather than getting up off the sofa and going to a local café or reading a book. Here represented by sitcoms, sports, and the like, fallback options are more widely shared, even among people whose first-best preferences differ widely, because they represent what people will tolerate before switching, a much less strict requirement than what they really want. This assumption follows Jack Beebe's refinement of Steiner's model. Beebe established that media monopolists would show nothing but common-denominator programs and that competition among broadcasters would begin to serve the smaller preference clusters only if a large enough number of channels were available. Such a model would explain the broad cultural sense of Bruce Springsteen's song, "57 Channels (And Nothin' On)," and why we saw the emergence of channels like Black Entertainment Television, Univision (Spanish channel in the United States), or The History Channel only when cable systems significantly expanded channel capacity, as well as why direct-broadcast satellite and, more recently, digital cable offerings were the first venue for twenty-four-hour-a-day cooking channels and smaller minority-language channels./13

Table 6.1: Distribution of Channels Hypothetical

No. of channels	Programming Available (in thousands of viewers)
1	sitcom (1000)
2	sitcom (1000), sports (750)
3	sitcom (1000 or 500), sports (750), indifferent between sitcoms and local news (500)
4	sitcom (500), sports (750), sitcom (500), local news (500)
5	sitcom (500), sports (375), sitcom (500), local news (500), sports (375)
6	sitcom (333), sports (375), sitcom (333), local news (500), sports (375), sitcom (333)
7	sitcom (333), sports (375), sitcom (333), local news (500), sports (375), sitcom (333), action movies (250)
8	sitcom (333), sports (375), sitcom (333), local news (250), sports (375), sitcom (333), action movies (250), local news (250)
9	sitcom (250), sports (375), sitcom (250), local news (250), sports (375), sitcom (250), action movies (250), local news (250), sitcom (250)
	* * * * * *
250	100 channels of sitcom (10); 75 channels of sports (10); 50 channels of local news (10); 25 channels of action movies (10)
251	100 channels of sitcom (10); 75 channels of sports (10); 50 channels of local news (10); 25 channels of action movies (10); 1 foreign film channel (9.99)
252	100 channels of sitcom (10); 75 channels of sports (10); 50 channels of local news (10); 25 channels of action movies (10); 1 foreign film channel (9.99); 1 gardening channel (9.98)

While this work was developed in the context of analyzing media diversity of offerings, it provides a foundation for understanding the programming choices of all advertiser-supported mass media, including the press, in domains relevant to the role they play as a platform for the public sphere. It provides a framework for understanding, but also limiting, the applicability of the idea that mass media will put nothing in the newspaper that will destroy the reader's appetite. Controversial views and genuinely disturbing images, descriptions, or arguments have a higher likelihood of turning readers, listeners, and viewers away than entertainment, mildly interesting and amusing human-interest stories, and a steady flow of basic crime and courtroom dramas, and similar fare typical of local television newscasts and newspapers. On the other hand, depending on the number of channels, there are clearly market segments for people who are "political junkies," or engaged elites, who can support some small number of outlets aimed at that crowd. The *New York Times* or the *Wall Street Journal* are examples in print, programs like *Meet the Press* or *Nightline* and perhaps channels like CNN and Fox News are examples of the possibility and limitations of this exception to the general entertainment-oriented, noncontroversial, and politically inert style of commercial mass media. The dynamic of programming to the lowest common denominator can, however, iteratively replicate itself even within relatively news- and elite-oriented media outlets. Even among news junkies, larger news outlets must cater relatively to the mainstream of its intended audience. Too strident a position or too probing an inquiry may slice the market segment to which they sell too thin. This is likely what leads to the common criticism, from both the Right and Left, that the same media are too "liberal" and too "conservative," respectively. By contrast, magazines, whose business model can support much lower circulation levels, exhibit a substantially greater will for political engagement and analysis than even the relatively political-readership-oriented, larger-circulation mass media. By definition, however, the media that cater to these niche markets serve only a small segment of the political community. Fox News in the United States appears to be a powerful counterexample to this trend. It is difficult to pinpoint why. The channel likely represents a composite of the Berlusconi effect, the high market segmentation made possible by high-capacity cable systems, the very large market segment of Republicans, and the relatively polarized tone of American political culture since the early 1990s.

The mass-media model as a whole, with the same caveat for niche markets, does not lend itself well to in-depth discussion and dialog. High professionalism can, to some extent, compensate for the basic structural problem of a medium built on the model of a small number of producers transmitting to an audience that is many orders of magnitude larger. The basic problem occurs at the intake and synthesis stages of communication. However diligent they may be, a small number of professional reporters, embedded as they are within social segments that are part of social, economic, and political elites, are a relatively stunted mechanism for intake. If one seeks to collect the wide range of individual observations, experiences, and opinions that make up the actual universe of concerns and opinions of a large public as a basic input into the public sphere, before filtering, the centralized model of mass media provides a limited means of capturing those insights. On the back end of the communication of public discourse, concentrated media of necessity must structure most "participants" in the debate as passive recipients of finished messages and images. That is the core characteristic of mass media: Content is produced prior to transmission in a relatively small number of centers, and when finished is then transmitted to a mass audience, which consumes it. This is the basis of the claim of the role of professional journalism to begin with, separating it from nonprofessional observations of those who consume its products. The result of this basic structure of the media product is that discussion and analysis of issues of common concern is an iconic representation of discussion, a choreographed enactment of public debate. The participants are selected for the fact that they represent well-understood, well-defined positions among those actually prevalent in a population, the images and stories are chosen to represent issues, and the public debate that is actually facilitated (and is supposedly where synthesis of the

opinions in public debate actually happens) is in fact an already presynthesized portrayal of an argument among avatars of relatively large segments of opinion as perceived by the journalists and stagers of the debate. In the United States, this translates into fairly standard formats of "on the left X, on the right Y," or "the Republicans' position" versus "the Democrats' position." It translates into "photo-op" moments of publicly enacting an idea, a policy position, or a state of affairs - whether it is a president landing on an aircraft carrier to represent security and the successful completion of a controversial war, or a candidate hunting with his buddies to represent a position on gun control. It is important to recognize that by describing these characteristics, I am not identifying failures of imagination, thoughtfulness, or professionalism on the part of media organizations. These are simply characteristics of a mass-mediated public sphere; modes of communication that offer the path of least resistance given the characteristics of the production and distribution process of mass media, particularly commercial mass media. There are partial exceptions, as there are to the diversity of content or the emphasis on entertainment value, but these do not reflect what most citizens read, see, or hear. The phenomenon of talk radio and call-in shows represents a very different, but certainly not more reflective form. They represent the pornography and violence of political discourse - a combination of exhibitionism and voyeurism intended to entertain us with opportunities to act out suppressed desires and to glimpse what we might be like if we allowed ourselves more leeway from what it means to be a well-socialized adult.

The two basic critiques of commercial mass media coalesce on the conflict between journalistic ethics and the necessities of commercialism. If professional journalists seek to perform a robust watchdog function, to inform their readers and viewers, and to provoke and explore in depth, then the dynamics of both power and lowest-common-denominator appeal push back. Different organizations, with different degrees of managerial control, editorial independence, internal organizational culture, and freedom from competitive pressures, with different intended market segments, will resolve these tensions differently. A quick reading of the conclusions of some media scholarship, and more commonly, arguments made in public debates over the media, would tend to lump "the media" as a single entity, with a single set of failures. In fact, unsurprisingly, the literature suggests substantial heterogeneity among organizations and media. Television seems to be the worst culprit on the dimension of political inertness. Print media, both magazines and some newspapers, include significant variation in the degree to which they fit these general models of failure.

As we turn now to consider the advantages of the introduction of Internet communications, we shall see how this new model can complement the mass media and alleviate its worst weaknesses. In particular, the discussion focuses on the emergence of the networked information economy and the relatively larger role it makes feasible for nonmarket actors and for radically distributed production of information and culture. One need not adopt the position that the commercial mass media are somehow abusive, evil, corporate-controlled giants, and that the Internet is the ideal Jeffersonian republic in order to track a series of genuine improvements represented by what the new emerging modalities of public communication can do as platforms for the public sphere. Greater access to means of direct individual communications, to collaborative speech platforms, and to nonmarket producers more generally can complement the commercial mass media and contribute to a significantly improved public sphere.

Notes

1. Jurgen Habermas, *Between Facts and Norms, Contributions to Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1996).

2. Elizabeth Eisenstein, *The Printing Press as an Agent of Change* (New York: Cambridge University Press, 1979); Jeremey Popkin, *News and Politics in the Age of Revolution: Jean Luzac's Gazzette de Leyde* (Ithaca, NY: Cornell University Press, 1989).

3. Paul Starr, *The Creation of the Media: Political Origins of Modern Communications* (New York: Basic Books, 2004), 33-46.

4. Starr, *Creation of the Media*, 48-62, 86-87.

5. Starr, *Creation of the Media*, 131-133.

6. Starr, *Creation of the Media*, 135.

7. The following discussion of the birth of radio is adapted from Yochai Benkler, "Overcoming Agoraphobia: Building the Commons of the Digitally Networked Environment," *Harvard Journal of Law and Technology* 11 (Winter 1997-1998): 287. That article provides the detailed support for the description. The major secondary works relied on are Erik Barnouw, *A History of Broadcasting in the United States* (New York: Oxford University Press, 1966-1970); Gleason Archer, *History of Radio to 1926* (New York: Arno Press, 1971); and Philip T. Rosen, *Modern Stentors: Radio Broadcasters and the Federal Government, 1920-1934* (Westport, CT: Greenwood Press, 1980).

8. Robert Waterman McChesney, *Telecommunications, Mass Media, and Democracy: The Battle for the Control of U.S. Broadcasting, 1928-1935* (New York: Oxford University Press, 1993).

9. "Names of U.S. Dead Read on *Nightline*," Associated Press Report, May 1, 2004, http://www.msnbc.msn.com/id/4864247/ .

10. The numbers given here are taken from The Center for Responsive Politics, http://www.opensecrets.org/ , and are based on information released by the Federal Elections Commission.

11. A careful catalog of these makes up the first part of C. Edwin Baker, *Media, Markets, and Democracy* (New York: Cambridge University Press, 2002).

12. Ben H. Bagdikian, *The Media Monopoly*, 5th ed. (Boston: Beacon Press, 1997), 118.

13. Peter O. Steiner, "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting," *The Quarterly Journal of Economics* 66 (1952): 194. The major other contribution in this literature is Jack H. Beebe, "Institutional Structure and Program Choices in Television Markets," *The Quarterly Journal of Economics* 91 (1977): 15. A parallel line of analysis of the relationship between programming and the market structure of broadcasting began with Michael Spence and Bruce Owen, "Television Programming, Monopolistic Competition, and Welfare," *The Quarterly Journal of Economics* 91 (1977): 103. For an excellent review of this literature, see Matthew L. Spitzer, "Justifying Minority Preferences in Broadcasting," *South California Law Review* 64 (1991): 293, 304-319.

The Wealth of Networks:
How Social Production Transforms Markets and Freedom
by Yochai Benkler, Yale University Press

Chapter 7
Political Freedom Part 2:
Emergence of the Networked Public Sphere

Epigraph

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Chapter 7
Political Freedom Part 2:
Emergence of the Networked Public Sphere

The fundamental elements of the difference between the networked information economy and the mass media are network architecture and the cost of becoming a speaker. The first element is the shift from a hub-and-spoke architecture with unidirectional links to the end points in the mass media, to distributed architecture with multidirectional connections among all nodes in the networked information environment. The second is the practical elimination of communications costs as a barrier to speaking across associational boundaries. Together, these characteristics have fundamentally altered the capacity of individuals, acting alone or with others, to be active participants in the public sphere as opposed to its passive readers, listeners, or viewers. For authoritarian countries, this means that it is harder and more costly, though not perhaps entirely impossible, to both be networked and maintain control over their public spheres. China seems to be doing too good a job of this in the middle of the first decade of this century for us to say much more than that it is harder to maintain control, and therefore that at least in some authoritarian regimes, control will be looser. In liberal democracies, ubiquitous individual ability to produce information creates the potential for near-universal intake. It therefore portends significant, though not inevitable, changes in the structure of the public sphere from the commercial mass-media environment. These changes raise challenges for filtering. They underlie some of the critiques of the claims about the democratizing effect of the Internet that I explore later in this chapter. Fundamentally, however, they are the roots of possible change. Beginning with the cost of sending an e-mail to some number of friends or to a mailing list of people interested in a particular subject, to the cost of setting up a Web site or a blog, and through to the possibility of maintaining interactive conversations with large numbers of people through sites like Slashdot, the cost of being a speaker in a regional, national, or even international political conversation is several orders of magnitude lower than the cost of speaking in the mass-mediated environment. This, in turn, leads to several orders of magnitude more speakers and participants in conversation and, ultimately, in the public sphere.

The change is as much qualitative as it is quantitative. The qualitative change is represented in the experience of being a potential speaker, as opposed to simply a listener and voter. It relates to the self-perception of individuals in society and the culture of participation they can adopt. The easy possibility of communicating effectively into the public sphere allows individuals to reorient themselves from passive readers and listeners to potential speakers and participants in a conversation. The way we listen to what we hear changes because of this; as does, perhaps most fundamentally, the way we observe and process daily events in our lives. We no longer need to take these as merely private observations, but as potential subjects for public communication. This change affects the relative power of the media. It affects the structure of intake of observations and views. It affects the presentation of issues and observations for discourse. It affects the way issues are filtered, for whom and by whom. Finally, it affects the ways in which positions are crystallized and synthesized, sometimes still by being amplified to the point that the mass media take them as inputs and convert them into political positions, but occasionally by direct organization of opinion and action to the point of reaching a salience that drives the political process directly.

The basic case for the democratizing effect of the Internet, as seen from the perspective of the mid-1990s, was articulated in an opinion of the U.S. Supreme Court in *Reno v. ACLU*:

The Web is thus comparable, from the readers' viewpoint, to both a vast library including millions of readily available and indexed publications and a sprawling mall offering goods and services. From the publishers' point of view, it constitutes a vast platform from which to address and hear from a world-wide audience of millions of readers, viewers, researchers, and buyers. Any person or organization with a computer connected to the Internet can "publish" information. Publishers include government agencies, educational institutions, commercial entities, advocacy groups, and individuals. . . .

Through the use of chat rooms, any person with a phone line can become a town crier with a voice that resonates farther than it could from any soapbox. Through the use of Web pages, mail exploders, and newsgroups, the same individual can become a pamphleteer. As the District Court found, "the content on the Internet is as diverse as human thought."[1]

The observations of what is different and unique about this new medium relative to those that dominated the twentieth century are already present in the quotes from the Court. There are two distinct types of effects. The first, as the Court notes from "the readers' perspective," is the abundance and diversity of human expression available to anyone, anywhere, in a way that was not feasible in the mass-mediated environment. The second, and more fundamental, is that anyone can be a publisher, including individuals, educational institutions, and nongovernmental organizations (NGOs), alongside the traditional speakers of the mass-media environment-government and commercial entities.

Since the end of the 1990s there has been significant criticism of this early conception of the democratizing effects of the Internet. One line of critique includes variants of the Babel objection: the concern that information overload will lead to fragmentation of discourse, polarization, and the loss of political community. A different and descriptively contradictory line of critique suggests that the Internet is, in fact, exhibiting concentration: Both infrastructure and, more fundamentally, patterns of attention are much less distributed than we thought. As a consequence, the Internet diverges from the mass media much less than we thought in the 1990s and significantly less than we might hope.

I begin the chapter by offering a menu of the core technologies and usage patterns that can be said, as of the middle of the first decade of the twenty-first century, to represent the core Internet-based technologies of democratic discourse. I then use two case studies to describe the social and economic practices through which these tools are implemented to construct the public sphere, and how these practices differ quite radically from the mass-media model. On the background of these stories, we are then able to consider the critiques that have been leveled against the claim that the Internet democratizes. Close examination of the application of networked information economy to the production of the public sphere suggests that the emerging networked public sphere offers significant improvements over one dominated by commercial mass media. Throughout the discussion, it is important to keep in mind that the relevant comparison is always between the public sphere that we in fact had throughout the twentieth century, the one dominated by mass media, that is the baseline for comparison, not the utopian image of the "everyone a pamphleteer" that animated the hopes of the 1990s for Internet democracy. Departures from the naïve utopia are not signs that the Internet does not democratize, after all. They are merely signs that the medium and its analysis are maturing.

Basic Tools of Networked Communication

Analyzing the effect of the networked information environment on public discourse by cataloging the currently popular tools for communication is, to some extent, self-defeating. These will undoubtedly be supplanted by new ones. Analyzing this effect without having a sense of what these tools are or how they are being used is, on the other hand, impossible. This leaves us with the need to catalog what is, while trying to abstract from what is being used to what relationships of information and communication are emerging, and from these to transpose to a theory of the networked information economy as a new platform for the public sphere.

E-mail is the most popular application on the Net. It is cheap and trivially easy to use. Basic e-mail, as currently used, is not ideal for public communications. While it provides a cheap and efficient means of communicating with large numbers of individuals who are not part of one's basic set of social associations, the presence of large amounts of commercial spam and the amount of mail flowing in and out of mailboxes make indiscriminate e-mail distributions a relatively poor mechanism for being heard. E-mails to smaller groups, preselected by the sender for having some interest in a subject or relationship to the sender, do, however, provide a rudimentary mechanism for communicating observations, ideas, and opinions to a significant circle, on an ad hoc basis. Mailing lists are more stable and self-selecting, and therefore more significant as a basic tool for the networked public sphere. Some mailing lists are moderated or edited, and run by one or a small number of editors. Others are not edited in any significant way. What separates mailing lists from most Web-based uses is the fact that they push the information on them into the mailbox of subscribers. Because of their attention limits, individuals restrict their subscriptions, so posting on a mailing list tends to be done by and for people who have self-selected as having a heightened degree of common interest, substantive or contextual. It therefore enhances the degree to which one is heard by those already interested in a topic. It is not a communications model of one-to-many, or few-to-many as broadcast is to an open, undefined class of audience members. Instead, it allows one, or a few, or even a limited large group to communicate to a large but limited group, where the limit is self-selection as being interested or even immersed in a subject.

The World Wide Web is the other major platform for tools that individuals use to communicate in the networked public sphere. It enables a wide range of applications, from basic static Web pages, to, more recently, blogs and various social-software-mediated platforms for large-scale conversations of the type described in chapter 3 - like Slashdot. Static Web pages are the individual's basic "broadcast" medium. They allow any individual or organization to present basic texts, sounds, and images pertaining to their position. They allow small NGOs to have a worldwide presence and visibility. They allow individuals to offer thoughts and commentaries. They allow the creation of a vast, searchable database of information, observations, and opinions, available at low cost for anyone, both to read and write into. This does not yet mean that all these statements are heard by the relevant others to whom they are addressed. Substantial analysis is devoted to that problem, but first let us complete the catalog of tools and information flow structures.

One Web-based tool and an emerging cultural practice around it that extends the basic characteristics of Web sites as media for the political public sphere are Web logs, or blogs. Blogs are a tool and an approach to using the Web that extends the use of Web pages in two significant ways. Technically, blogs are part of a broader category of innovations that make the web "writable." That is, they make Web pages easily capable of modification through a simple interface. They can be modified from anywhere with a networked computer, and the results of writing onto the Web page are immediately available to anyone who accesses the blog to read. This technical change resulted in two divergences from the cultural practice of Web sites in the 1990s. First, they allowed the

evolution of a journal-style Web page, where individual short posts are added to the Web site in short or large intervals. As practice has developed over the past few years, these posts are usually archived chronologically. For many users, this means that blogs have become a form of personal journal, updated daily or so, for their own use and perhaps for the use of a very small group of friends. What is significant about this characteristic from the perspective of the construction of the public sphere is that blogs enable individuals to write to their Web pages in journalism time - that is, hourly, daily, weekly - whereas Web page culture that preceded it tended to be slower moving: less an equivalent of reportage than of the essay. Today, one certainly finds individuals using blog software to maintain what are essentially static Web pages, to which they add essays or content occasionally, and Web sites that do not use blogging technology but are updated daily. The public sphere function is based on the content and cadence - that is, the use practice - not the technical platform.

The second critical innovation of the writable Web in general and of blogs in particular was the fact that in addition to the owner, readers/users could write to the blog. Blogging software allows the person who runs a blog to permit some, all, or none of the readers to post comments to the blog, with or without retaining power to edit or moderate the posts that go on, and those that do not. The result is therefore not only that many more people write finished statements and disseminate them widely, but also that the end product is a weighted conversation, rather than a finished good. It is a conversation because of the common practice of allowing and posting comments, as well as comments to these comments. Blog writers - bloggers - often post their own responses in the comment section or address comments in the primary section. Blog-based conversation is weighted, because the culture and technical affordances of blogging give the owner of the blog greater weight in deciding who gets to post or comment and who gets to decide these questions. Different blogs use these capabilities differently; some opt for broader intake and discussion on the board, others for a more tightly edited blog. In all these cases, however, the communications model or information-flow structure that blogs facilitate is a weighted conversation that takes the form of one or a group of primary contributors/authors, together with some larger number, often many, secondary contributors, communicating to an unlimited number of many readers.

The writable Web also encompasses another set of practices that are distinct, but that are often pooled in the literature together with blogs. These are the various larger-scale, collaborative-content production systems available on the Web, of the type described in chapter 3. Two basic characteristics make sites like Slashdot or Wikipedia different from blogs. First, they are intended for, and used by, very large groups, rather than intended to facilitate a conversation weighted toward one or a small number of primary speakers. Unlike blogs, they are not media for individual or small group expression with a conversation feature. They are intrinsically group communication media. They therefore incorporate social software solutions to avoid deterioration into chaos - peer review, structured posting privileges, reputation systems, and so on. Second, in the case of Wikis, the conversation platform is anchored by a common text. From the perspective of facilitating the synthesis of positions and opinions, the presence of collaborative authorship of texts offers an additional degree of viscosity to the conversation, so that views "stick" to each other, must jostle for space, and accommodate each other. In the process, the output is more easily recognizable as a collective output and a salient opinion or observation than where the form of the conversation is more free-flowing exchange of competing views.

Common to all these Web-based tools - both static and dynamic, individual and cooperative - are linking, quotation, and presentation. It is at the very core of the hypertext markup language (HTML) to make referencing easy. And it is at the very core of a radically distributed network to allow materials to be archived by whoever wants to archive them, and then to be accessible to

whoever has the reference. Around these easy capabilities, the cultural practice has emerged to reference through links for easy transition from your own page or post to the one you are referring to - whether as inspiration or in disagreement. This culture is fundamentally different from the mass-media culture, where sending a five-hundred-page report to millions of users is hard and expensive. In the mass media, therefore, instead of allowing readers to read the report alongside its review, all that is offered is the professional review in the context of a culture that trusts the reviewer. On the Web, linking to original materials and references is considered a core characteristic of communication. The culture is oriented toward "see for yourself." Confidence in an observation comes from a combination of the reputation of the speaker as it has emerged over time, reading underlying sources you believe you have some competence to evaluate for yourself, and knowing that for any given referenced claim or source, there is some group of people out there, unaffiliated with the reviewer or speaker, who will have access to the source and the means for making their disagreement with the speaker's views known. Linking and "see for yourself" represent a radically different and more participatory model of accreditation than typified the mass media.

Another dimension that is less well developed in the United States than it is in Europe and East Asia is mobility, or the spatial and temporal ubiquity of basic tools for observing and commenting on the world we inhabit. Dan Gillmor is clearly right to include these basic characteristics in his book *We the Media*, adding to the core tools of what he describes as a transformation in journalism, short message service (SMS), and mobile connected cameras to mailing lists, Web logs, Wikis, and other tools. The United States has remained mostly a PC-based networked system, whereas in Europe and Asia, there has been more substantial growth in handheld devices, primarily mobile phones. In these domains, SMS - the "e-mail" of mobile phones - and camera phones have become critical sources of information, in real time. In some poor countries, where cell phone minutes remain very (even prohibitively) expensive for many users and where landlines may not exist, text messaging is becoming a central and ubiquitous communication tool. What these suggest to us is a transition, as the capabilities of both systems converge, to widespread availability of the ability to register and communicate observations in text, audio, and video, wherever we are and whenever we wish. Drazen Pantic tells of how listeners of Internet-based Radio B-92 in Belgrade reported events in their neighborhoods after the broadcast station had been shut down by the Milosevic regime. Howard Rheingold describes in *Smart Mobs* how citizens of the Philippines used SMS to organize real-time movements and action to overthrow their government. In a complex modern society, where things that matter can happen anywhere and at any time, the capacities of people armed with the means of recording, rendering, and communicating their observations change their relationship to the events that surround them. Whatever one sees and hears can be treated as input into public debate in ways that were impossible when capturing, rendering, and communicating were facilities reserved to a handful of organizations and a few thousands of their employees.

Networked Information Economy Meets the Public Sphere

The networked public sphere is not made of tools, but of social production practices that these tools enable. The primary effect of the Internet on the public sphere in liberal societies relies on the information and cultural production activity of emerging nonmarket actors: individuals working alone and cooperatively with others, more formal associations like NGOs, and their feedback effect on the mainstream media itself. These enable the networked public sphere to moderate the two major concerns with commercial mass media as a platform for the public sphere: (1) the excessive power it gives its owners, and (2) its tendency, when owners do not dedicate their media to exert power, to foster an inert polity. More fundamentally, the social practices of information and discourse allow a very large number of actors to see themselves as potential contributors to public discourse and as potential actors in political arenas, rather than mostly passive recipients of

mediated information who occasionally can vote their preferences. In this section, I offer two detailed stories that highlight different aspects of the effects of the networked information economy on the construction of the public sphere. The first story focuses on how the networked public sphere allows individuals to monitor and disrupt the use of mass-media power, as well as organize for political action. The second emphasizes in particular how the networked public sphere allows individuals and groups of intense political engagement to report, comment, and generally play the role traditionally assigned to the press in observing, analyzing, and creating political salience for matters of public interest. The case studies provide a context both for seeing how the networked public sphere responds to the core failings of the commercial, mass-media-dominated public sphere and for considering the critiques of the Internet as a platform for a liberal public sphere.

Our first story concerns Sinclair Broadcasting and the 2004 U.S. presidential election. It highlights the opportunities that mass-media owners have to exert power over the public sphere, the variability within the media itself in how this power is used, and, most significant for our purposes here, the potential corrective effect of the networked information environment. At its core, it suggests that the existence of radically decentralized outlets for individuals and groups can provide a check on the excessive power that media owners were able to exercise in the industrial information economy.

Sinclair, which owns major television stations in a number of what were considered the most competitive and important states in the 2004 election - including Ohio, Florida, Wisconsin, and Iowa - informed its staff and stations that it planned to preempt the normal schedule of its sixty-two stations to air a documentary called *Stolen Honor: The Wounds That Never Heal*, as a news program, a week and a half before the elections./2 The documentary was reported to be a strident attack on Democratic candidate John Kerry's Vietnam War service. One reporter in Sinclair's Washington bureau, who objected to the program and described it as "blatant political propaganda," was promptly fired./3 The fact that Sinclair owns stations reaching one quarter of U.S. households, that it used its ownership to preempt local broadcast schedules, and that it fired a reporter who objected to its decision, make this a classic "Berlusconi effect" story, coupled with a poster-child case against media concentration and the ownership of more than a small number of outlets by any single owner. The story of Sinclair's plans broke on Saturday, October 9, 2004, in the *Los Angeles Times*. Over the weekend, "official" responses were beginning to emerge in the Democratic Party. The Kerry campaign raised questions about whether the program violated election laws as an undeclared "in-kind" contribution to the Bush campaign. By Tuesday, October 12, the Democratic National Committee announced that it was filing a complaint with the Federal Elections Commission (FEC), while seventeen Democratic senators wrote a letter to the chairman of the Federal Communications Commission (FCC), demanding that the commission investigate whether Sinclair was abusing the public trust in the airwaves. Neither the FEC nor the FCC, however, acted or intervened throughout the episode.

Alongside these standard avenues of response in the traditional public sphere of commercial mass media, their regulators, and established parties, a very different kind of response was brewing on the Net, in the blogosphere. On the morning of October 9, 2004, the *Los Angeles Times* story was blogged on a number of political blogs - Josh Marshall on talkingpointsmemo.com, Chris Bower on MyDD.com, and Markos Moulitsas on dailyKos.com. By midday that Saturday, October 9, two efforts aimed at organizing opposition to Sinclair were posted in the dailyKos and MyDD. A "boycottSinclair" site was set up by one individual, and was pointed to by these blogs. Chris Bowers on MyDD provided a complete list of Sinclair stations and urged people to call the stations and threaten to picket and boycott. By Sunday, October 10, the dailyKos posted a list of national

advertisers with Sinclair, urging readers to call them. On Monday, October 11, MyDD linked to that list, while another blog, theleftcoaster.com, posted a variety of action agenda items, from picketing affiliates of Sinclair to suggesting that readers oppose Sinclair license renewals, providing a link to the FCC site explaining the basic renewal process and listing public-interest organizations to work with. That same day, another individual, Nick Davis, started a Web site, BoycottSBG.com, on which he posted the basic idea that a concerted boycott of local advertisers was the way to go, while another site, stopsinclair.org, began pushing for a petition. In the meantime, TalkingPoints published a letter from Reed Hundt, former chairman of the FCC, to Sinclair, and continued finding tidbits about the film and its maker. Later on Monday, TalkingPoints posted a letter from a reader who suggested that stockholders of Sinclair could bring a derivative action. By 5:00 a.m. on the dawn of Tuesday, October 12, however, TalkingPoints began pointing toward Davis's database on BoycottSBG.com. By 10:00 that morning, Marshall posted on TalkingPoints a letter from an anonymous reader, which began by saying: "I've worked in the media business for 30 years and I guarantee you that sales is what these local TV stations are all about. They don't care about license renewal or overwhelming public outrage. They care about sales only, so only local advertisers can affect their decisions." This reader then outlined a plan for how to watch and list all local advertisers, and then write to the sales managers - not general managers - of the local stations and tell them which advertisers you are going to call, and then call those. By 1:00 p.m. Marshall posted a story of his own experience with this strategy. He used Davis's database to identify an Ohio affiliate's local advertisers. He tried to call the sales manager of the station, but could not get through. He then called the advertisers. The post is a "how to" instruction manual, including admonitions to remember that the advertisers know nothing of this, the story must be explained, and accusatory tones avoided, and so on. Marshall then began to post letters from readers who explained with whom they had talked - a particular sales manager, for example - and who were then referred to national headquarters. He continued to emphasize that advertisers were the right addressees. By 5:00 p.m. that same Tuesday, Marshall was reporting more readers writing in about experiences, and continued to steer his readers to sites that helped them to identify their local affiliate's sales manager and their advertisers./4

By the morning of Wednesday, October 13, the boycott database already included eight hundred advertisers, and was providing sample letters for users to send to advertisers. Later that day, BoycottSBG reported that some participants in the boycott had received reply e-mails telling them that their unsolicited e-mail constituted illegal spam. Davis explained that the CAN-SPAM Act, the relevant federal statute, applied only to commercial spam, and pointed users to a law firm site that provided an overview of CAN-SPAM. By October 14, the boycott effort was clearly bearing fruit. Davis reported that Sinclair affiliates were threatening advertisers who cancelled advertisements with legal action, and called for volunteer lawyers to help respond. Within a brief period, he collected more than a dozen volunteers to help the advertisers. Later that day, another blogger at grassrootsnation.com had set up a utility that allowed users to send an e-mail to all advertisers in the BoycottSBG database. By the morning of Friday, October 15, Davis was reporting more than fifty advertisers pulling ads, and three or four mainstream media reports had picked up the boycott story and reported on it. That day, an analyst at Lehman Brothers issued a research report that downgraded the expected twelve-month outlook for the price of Sinclair stock, citing concerns about loss of advertiser revenue and risk of tighter regulation. Mainstream news reports over the weekend and the following week systematically placed that report in context of local advertisers pulling their ads from Sinclair. On Monday, October 18, the company's stock price dropped by 8 percent (while the S&P 500 rose by about half a percent). The following morning, the stock dropped a further 6 percent, before beginning to climb back, as Sinclair announced that it would not show *Stolen Honor*, but would provide a balanced program with only portions of the documentary and one that would include arguments on the other side. On that day, the company's

stock price had reached its lowest point in three years. The day after the announced change in programming decision, the share price bounced back to where it had been on October 15. There were obviously multiple reasons for the stock price losses, and Sinclair stock had been losing ground for many months prior to these events. Nonetheless, as figure 7.1 demonstrates, the market responded quite sluggishly to the announcements of regulatory and political action by the Democratic establishment earlier in the week of October 12, by comparison to the precipitous decline and dramatic bounce-back surrounding the market projections that referred to advertising loss. While this does not prove that the Web-organized, blog-driven and -facilitated boycott was the determining factor, as compared to fears of formal regulatory action, the timing strongly suggests that the efficacy of the boycott played a very significant role.
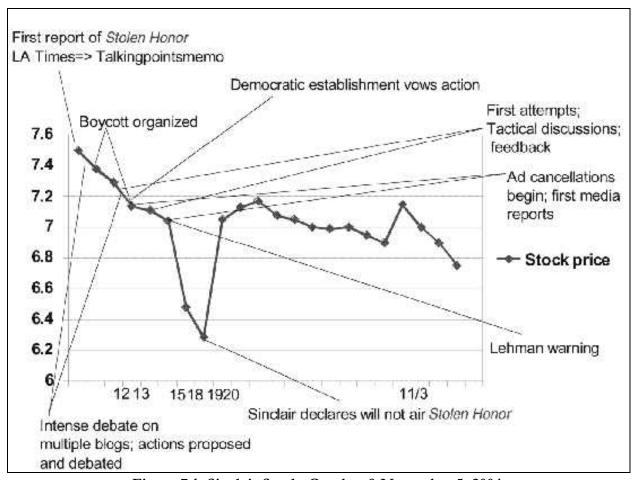
Figure 7.1: Sinclair Stock, October 8-November 5, 2004

The first lesson of the Sinclair *Stolen Honor* story is about commercial mass media themselves. The potential for the exercise of inordinate power by media owners is not an imaginary concern. Here was a publicly traded firm whose managers supported a political party and who planned to use their corporate control over stations reaching one quarter of U.S. households, many in swing states, to put a distinctly political message in front of this large audience. We also learn, however, that in the absence of monopoly, such decisions do not determine what everyone sees or hears, and that other mass-media outlets will criticize each other under these conditions. This criticism alone, however, cannot stop a determined media owner from trying to exert its influence in the public sphere, and if placed as Sinclair was, in locations with significant political weight, such intervention could have

substantial influence. Second, we learn that the new, network-based media can exert a significant counterforce. They offer a completely new and much more widely open intake basin for insight and commentary. The speed with which individuals were able to set up sites to stake out a position, to collect and make available information relevant to a specific matter of public concern, and to provide a platform for others to exchange views about the appropriate political strategy and tactics was completely different from anything that the economics and organizational structure of mass media make feasible. The third lesson is about the internal dynamics of the networked public sphere. Filtering and synthesis occurred through discussion, trial, and error. Multiple proposals for action surfaced, and the practice of linking allowed most anyone interested who connected to one of the nodes in the network to follow quotations and references to get a sense of the broad range of proposals. Different people could coalesce on different modes of action - 150,000 signed the petition on stopsinclair.org, while others began to work on the boycott. Setting up the mechanism was trivial, both technically and as a matter of cost - something a single committed individual could choose to do. Pointing and adoption provided the filtering, and feedback about the efficacy, again distributed through a system of cross-references, allowed for testing and accreditation of this course of action. High-visibility sites, like Talkingpointsmemo or the dailyKos, offered transmissions hubs that disseminated information about the various efforts and provided a platform for interest-group-wide tactical discussions. It remains ambiguous to what extent these dispersed loci of public debate still needed mass-media exposure to achieve broad political salience. BoycottSBG.com received more than three hundred thousand unique visitors during its first week of operations, and more than one million page views. It successfully coordinated a campaign that resulted in real effects on advertisers in a large number of geographically dispersed media markets. In this case, at least, mainstream media reports on these efforts were few, and the most immediate "transmission mechanism" of their effect was the analyst's report from Lehman, not the media. It is harder to judge the extent to which those few mainstream media reports that did appear featured in the decision of the analyst to credit the success of the boycott efforts. The fact that mainstream media outlets may have played a role in increasing the salience of the boycott does not, however, take away from the basic role played by these new mechanisms of bringing information and experience to bear on a broad public conversation combined with a mechanism to organize political action across many different locations and social contexts.

Our second story focuses not on the new reactive capacity of the networked public sphere, but on its generative capacity. In this capacity, it begins to outline the qualitative change in the role of individuals as potential investigators and commentators, as active participants in defining the agenda and debating action in the public sphere. This story is about Diebold Election Systems (one of the leading manufacturers of electronic voting machines and a subsidiary of one of the foremost ATM manufacturers in the world, with more than $2 billion a year in revenue), and the way that public criticism of its voting machines developed. It provides a series of observations about how the networked information economy operates, and how it allows large numbers of people to participate in a peer-production enterprise of news gathering, analysis, and distribution, applied to a quite unsettling set of claims. While the context of the story is a debate over electronic voting, that is not what makes it pertinent to democracy. The debate could have centered on any corporate and government practice that had highly unsettling implications, was difficult to investigate and parse, and was largely ignored by mainstream media. The point is that the networked public sphere did engage, and did successfully turn something that was not a matter of serious public discussion to a public discussion that led to public action.

Electronic voting machines were first used to a substantial degree in the United States in the November 2002 elections. Prior to, and immediately following that election, there was sparse mass-media coverage of electronic voting machines. The emphasis was mostly on the newness,

occasional slips, and the availability of technical support staff to help at polls. An *Atlanta Journal-Constitution* story, entitled "Georgia Puts Trust in Electronic Voting, Critics Fret about Absence of Paper Trails,"[/5] is not atypical of coverage at the time, which generally reported criticism by computer engineers, but conveyed an overall soothing message about the efficacy of the machines and about efforts by officials and companies to make sure that all would be well. The *New York Times* report of the Georgia effort did not even mention the critics.[/6] The *Washington Post* reported on the fears of failure with the newness of the machines, but emphasized the extensive efforts that the manufacturer, Diebold, was making to train election officials and to have hundreds of technicians available to respond to failure.[/7] After the election, the *Atlanta Journal-Constitution* reported that the touch-screen machines were a hit, burying in the text any references to machines that highlighted the wrong candidates or the long lines at the booths, while the *Washington Post* highlighted long lines in one Maryland county, but smooth operation elsewhere. Later, the *Post* reported a University of Maryland study that surveyed users and stated that quite a few needed help from election officials, compromising voter privacy.[/8] Given the centrality of voting mechanisms for democracy, the deep concerns that voting irregularities determined the 2000 presidential elections, and the sense that voting machines would be a solution to the "hanging chads" problem (the imperfectly punctured paper ballots that came to symbolize the Florida fiasco during that election), mass-media reports were remarkably devoid of any serious inquiry into how secure and accurate voting machines were, and included a high quotient of soothing comments from election officials who bought the machines and executives of the manufacturers who sold them. No mass-media outlet sought to go behind the claims of the manufacturers about their machines, to inquire into their security or the integrity of their tallying and transmission mechanisms against vote tampering. No doubt doing so would have been difficult. These systems were protected as trade secrets. State governments charged with certifying the systems were bound to treat what access they had to the inner workings as confidential. Analyzing these systems requires high degrees of expertise in computer security. Getting around these barriers is difficult. However, it turned out to be feasible for a collection of volunteers in various settings and contexts on the Net.

In late January 2003, Bev Harris, an activist focused on electronic voting machines, was doing research on Diebold, which has provided more than 75,000 voting machines in the United States and produced many of the machines used in Brazil's purely electronic voting system. Harris had set up a whistle-blower site as part of a Web site she ran at the time, blackboxvoting.com. Apparently working from a tip, Harris found out about an openly available site where Diebold stored more than forty thousand files about how its system works. These included specifications for, and the actual code of, Diebold's machines and vote-tallying system. In early February 2003, Harris published two initial journalistic accounts on an online journal in New Zealand, Scoop.com - whose business model includes providing an unedited platform for commentators who wish to use it as a platform to publish their materials. She also set up a space on her Web site for technically literate users to comment on the files she had retrieved. In early July of that year, she published an analysis of the results of the discussions on her site, which pointed out how access to the Diebold open site could have been used to affect the 2002 election results in Georgia (where there had been a tightly contested Senate race). In an editorial attached to the publication, entitled "Bigger than Watergate," the editors of Scoop claimed that what Harris had found was nothing short of a mechanism for capturing the U.S. elections process. They then inserted a number of lines that go to the very heart of how the networked information economy can use peer production to play the role of watchdog:

> We can now reveal for the first time the location of a complete online copy of the original data set. As we anticipate attempts to prevent the distribution of this information we encourage supporters of democracy to make copies of these files and to make them available on websites and file sharing networks: http://users.actrix.co.nz/dolly/. As many of

the files are zip password protected you may need some assistance in opening them, we have found that the utility available at the following URL works well: http://www.lostpassword.com. Finally some of the zip files are partially damaged, but these too can be read by using the utility at: http://www.zip-repair.com/. At this stage in this inquiry we do not believe that we have come even remotely close to investigating all aspects of this data; i.e., there is no reason to believe that the security flaws discovered so far are the only ones. Therefore we expect many more discoveries to be made. We want the assistance of the online computing community in this enterprise and we encourage you to file your findings at the forum HERE [providing link to forum].

A number of characteristics of this call to arms would have been simply infeasible in the mass-media environment. They represent a genuinely different mind-set about how news and analysis are produced and how censorship and power are circumvented. First, the ubiquity of storage and communications capacity means that public discourse can rely on "see for yourself" rather than on "trust me." The first move, then, is to make the raw materials available for all to see. Second, the editors anticipated that the company would try to suppress the information. Their response was not to use a counterweight of the economic and public muscle of a big media corporation to protect use of the materials. Instead, it was widespread distribution of information - about where the files could be found, and about where tools to crack the passwords and repair bad files could be found - matched with a call for action: get these files, copy them, and store them in many places so they cannot be squelched. Third, the editors did not rely on large sums of money flowing from being a big media organization to hire experts and interns to scour the files. Instead, they posed a challenge to whoever was interested - there are more scoops to be found, this is important for democracy, good hunting!! Finally, they offered a platform for integration of the insights on their own forum. This short paragraph outlines a mechanism for radically distributed storage, distribution, analysis, and reporting on the Diebold files.

As the story unfolded over the next few months, this basic model of peer production of investigation, reportage, analysis, and communication indeed worked. It resulted in the decertification of some of Diebold's systems in California, and contributed to a shift in the requirements of a number of states, which now require voting machines to produce a paper trail for recount purposes. The first analysis of the Diebold system based on the files Harris originally found was performed by a group of computer scientists at the Information Security Institute at Johns Hopkins University and released as a working paper in late July 2003. The Hopkins Report, or Rubin Report as it was also named after one of its authors, Aviel Rubin, presented deep criticism of the Diebold system and its vulnerabilities on many dimensions. The academic credibility of its authors required a focused response from Diebold. The company published a line-by-line response. Other computer scientists joined in the debate. They showed the limitations and advantages of the Hopkins Report, but also where the Diebold response was adequate and where it provided implicit admission of the presence of a number of the vulnerabilities identified in the report. The report and comments to it sparked two other major reports, commissioned by Maryland in the fall of 2003 and later in January 2004, as part of that state's efforts to decide whether to adopt electronic voting machines. Both studies found a wide range of flaws in the systems they examined and required modifications (see figure 7.2).
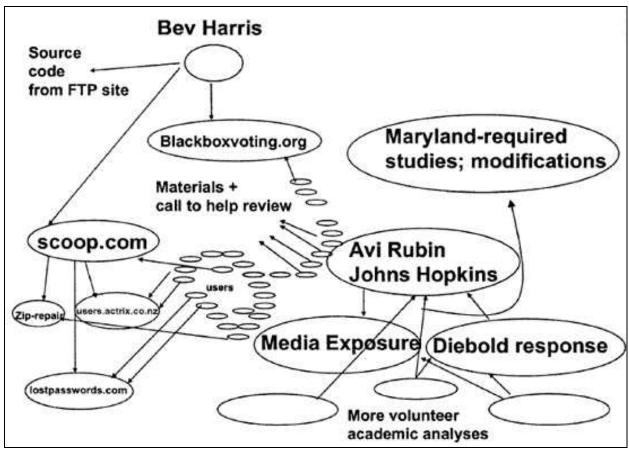
Figure 7.2: Analysis of the Diebold Source Code Materials

Meanwhile, trouble was brewing elsewhere for Diebold. In early August 2003, someone provided *Wired* magazine with a very large cache containing thousands of internal e-mails of Diebold. *Wired* reported that the e-mails were obtained by a hacker, emphasizing this as another example of the laxity of Diebold's security. However, the magazine provided neither an analysis of the e-mails nor access to them. Bev Harris, the activist who had originally found the Diebold materials, on the other hand, received the same cache, and posted the e-mails and memos on her site. Diebold's response was to threaten litigation. Claiming copyright in the e-mails, the company demanded from Harris, her Internet service provider, and a number of other sites where the materials had been posted, that the e-mails be removed. The e-mails were removed from these sites, but the strategy of widely distributed replication of data and its storage in many different topological and organizationally diverse settings made Diebold's efforts ultimately futile. The protagonists from this point on were college students. First, two students at Swarthmore College in Pennsylvania, and quickly students in a number of other universities in the United States, began storing the e-mails and scouring them for evidence of impropriety. In October 2003, Diebold proceeded to write to the universities whose students were hosting the materials. The company invoked provisions of the Digital Millennium Copyright Act that require Web-hosting companies to remove infringing materials when copyright owners notify them of the presence of these materials on their sites. The universities obliged, and required the students to remove the materials from their sites. The students, however, did not disappear quietly into the night. On October 21, 2003, they launched a multipronged campaign of what they described as "electronic civil disobedience." First, they kept moving the files from one student to another's machine, encouraging students around the country to resist the efforts to eliminate the material. Second, they injected the materials into FreeNet, the

anticensorship peer-to-peer publication network, and into other peer-to-peer file-sharing systems, like eDonkey and BitTorrent. Third, supported by the Electronic Frontier Foundation, one of the primary civil-rights organizations concerned with Internet freedom, the students brought suit against Diebold, seeking a judicial declaration that their posting of the materials was privileged. They won both the insurgent campaign and the formal one. As a practical matter, the materials remained publicly available throughout this period. As a matter of law, the litigation went badly enough for Diebold that the company issued a letter promising not to sue the students. The court nonetheless awarded the students damages and attorneys' fees because it found that Diebold had "knowingly and materially misrepresented" that the publication of the e-mail archive was a copyright violationin its letters to the Internet service providers./9

Central from the perspective of understanding the dynamics of the networked public sphere is not, however, the court case - it was resolved almost a year later, after most of the important events had already unfolded - but the efficacy of the students' continued persistent publication in the teeth of the cease-and-desist letters and the willingness of the universities to comply. The strategy of replicating the files everywhere made it impracticable to keep the documents from the public eye. And the public eye, in turn, scrutinized. Among the things that began to surface as users read the files were internal e-mails recognizing problems with the voting system, with the security of the FTP site from which Harris had originally obtained the specifications of the voting systems, and e-mail that indicated that the machines implemented in California had been "patched" or updated after their certification. That is, the machines actually being deployed in California were at least somewhat different from the machines that had been tested and certified by the state. This turned out to have been a critical find.

California had a Voting Systems Panel within the office of the secretary of state that reviewed and certified voting machines. On November 3, 2003, two weeks after the students launched their electronic disobedience campaign, the agenda of the panel's meeting was to include a discussion of proposed modifications to one of Diebold's voting systems. Instead of discussing the agenda item, however, one of the panel members made a motion to table the item until the secretary of state had an opportunity to investigate, because "It has come to our attention that some very disconcerting information regarding this item [*sic*] and we are informed that this company, Diebold, may have installed uncertified software in at least one county before it was certified."/10 The source of the information is left unclear in the minutes. A later report in *Wired* cited an unnamed source in the secretary of state's office as saying that somebody within the company had provided this information. The timing and context, however, suggest that it was the revelation and discussion of the e-mail memoranda online that played that role. Two of the members of the public who spoke on the record mention information from within the company. One specifically mentions the information gleaned from company e-mails. In the next committee meeting, on December 16, 2003, one member of the public who was in attendance specifically referred to the e-mails on the Internet, referencing in particular a January e-mail about upgrades and changes to the certified systems. By that December meeting, the independent investigation by the secretary of state had found systematic discrepancies between the systems actually installed and those tested and certified by the state. The following few months saw more studies, answers, debates, and the eventual decertification of many of the Diebold machines installed in California (see figures 7.3a and 7.3b).
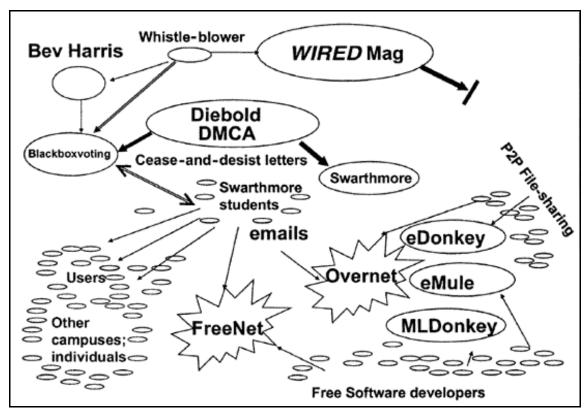
Figure 7.3a: Diebold Internal E-mails Discovery and Distribution
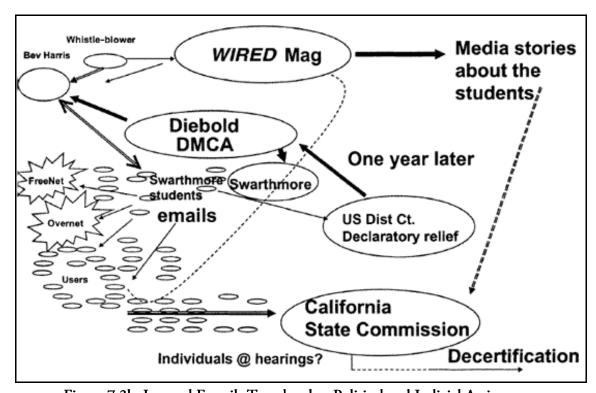
Figure 7.3b: Internal E-mails Translated to Political and Judicial Action

The structure of public inquiry, debate, and collective action exemplified by this story is fundamentally different from the structure of public inquiry and debate in the mass-media-dominated public sphere of the twentieth century. The initial investigation and analysis was done by a committed activist, operating on a low budget and with no financing from a media company. The output of this initial inquiry was not a respectable analysis by a major player in the public debate. It was access to raw materials and initial observations about them, available to start a conversation. Analysis then emerged from a widely distributed process undertaken by Internet users of many different types and abilities. In this case, it included academics studying electronic voting systems, activists, computer systems practitioners, and mobilized students. When the pressure from a well-financed corporation mounted, it was not the prestige and money of a *Washington Post* or a *New York Times* that protected the integrity of the information and its availability for public scrutiny. It was the radically distributed cooperative efforts of students and peer-to-peer network users around the Internet. These efforts were, in turn, nested in other communities of cooperative production - like the free software community that developed some of the applications used to disseminate the e-mails after Swarthmore removed them from the students' own site. There was no single orchestrating power - neither party nor professional commercial media outlet. There was instead a series of uncoordinated but mutually reinforcing actions by individuals in different settings and contexts, operating under diverse organizational restrictions and affordances, to expose, analyze, and distribute criticism and evidence for it. The networked public sphere here does not rely on advertising or capturing large audiences to focus its efforts. What became salient for the public agenda and shaped public discussion was what intensely engaged active participants, rather than what kept the moderate attention of large groups of passive viewers. Instead of the lowest-common-denominator focus typical of commercial mass media, each individual and group can - and, indeed, most likely will - focus precisely on what is most intensely interesting to its participants. Instead of iconic representation built on the scarcity of time slots and space on the air or on the page, we see the emergence of a "see for yourself" culture. Access to underlying documents and statements, and to the direct expression of the opinions of others, becomes a central part of the medium.

Critiques of the Claims that the Internet has Democratizing Effects

It is common today to think of the 1990s, out of which came the Supreme Court's opinion in *Reno v. ACLU*, as a time of naïve optimism about the Internet, expressing in political optimism the same enthusiasm that drove the stock market bubble, with the same degree of justifiability. An ideal liberal public sphere did not, in fact, burst into being from the Internet, fully grown like Athena from the forehead of Zeus. The detailed criticisms of the early claims about the democratizing effects of the Internet can be characterized as variants of five basic claims:

1.*Information overload.* A basic problem created when everyone can speak is that there will be too many statements, or too much information. Too many observations and too many points of view make the problem of sifting through them extremely difficult, leading to an unmanageable din. This overall concern, a variant of the Babel objection, underlies three more specific arguments: that money will end up dominating anyway, that there will be fragmentation of discourse, and that fragmentation of discourse will lead to its polarization.

a.*Money will end up dominating anyway.* A point originally raised by Eli Noam is that in this explosively large universe, getting attention will be as difficult as getting your initial message out in the mass-media context, if not more so. The same means that dominated the capacity to speak in the mass-media environment - money - will dominate the capacity to be heard on the Internet, even if it no longer controls the capacity to speak.

b.*Fragmentation of attention and discourse.* A point raised most explicitly by Cass Sunstein in *Republic.com* is that the ubiquity of information and the absence of the mass media as condensation points will impoverish public discourse by fragmenting it. There will be no public sphere. Individuals will view the world through millions of personally customized windows that will offer no common ground for political discourse or action, except among groups of highly similar individuals who customize their windows to see similar things.

c.*Polarization.* A descriptively related but analytically distinct critique of Sunstein's was that the fragmentation would lead to polarization. When information and opinions are shared only within groups of like-minded participants, he argued, they tend to reinforce each other's views and beliefs without engaging with alternative views or seeing the concerns and critiques of others. This makes each view more extreme in its own direction and increases the distance between positions taken by opposing camps.

2.*Centralization of the Internet.* A second-generation criticism of the democratizing effects of the Internet is that it turns out, in fact, not to be as egalitarian or distributed as the 1990s conception had suggested. First, there is concentration in the pipelines and basic tools of communications. Second, and more intractable to policy, even in an open network, a high degree of attention is concentrated on a few top sites - a tiny number of sites are read by the vast majority of readers, while many sites are never visited by anyone. In this context, the Internet is replicating the mass-media model, perhaps adding a few channels, but not genuinely changing anything structural.

Note that the concern with information overload is in direct tension with the second-generation concerns. To the extent that the concerns about Internet concentration are correct, they suggest that the information overload is not a deep problem. Sadly, from the perspective of democracy, it turns out that according to the concentration concern, there are few speakers to which most people listen, just as in the mass-media environment. While this means that the supposed benefits of the networked public sphere are illusory, it also means that the information overload concerns about what happens when there is no central set of speakers to whom most people listen are solved in much the same way that the mass-media model deals with the factual diversity of information, opinion, and observations in large societies - by consigning them to public oblivion. The response to both sets of concerns will therefore require combined consideration of a series of questions: To what extent are the claims of concentration correct? How do they solve the information overload problem? To what extent does the observed concentration replicate the mass-media model?

3.*Centrality of commercial mass media to the Fourth Estate function.* The importance of the press to the political process is nothing new. It earned the press the nickname "the Fourth Estate" (a reference to the three estates that made up the prerevolutionary French Estates-General, the clergy, nobility, and townsmen), which has been in use for at least a hundred and fifty years. In American free speech theory, the press is often described as fulfilling "the watchdog function," deriving from the notion that the public representatives must be watched over to assure they do the public's business faithfully. In the context of the Internet, the concern, most clearly articulated by Neil Netanel, has been that in the modern complex societies in which we live, commercial mass media are critical for preserving the watchdog function of the media. Big, sophisticated, well-funded government and corporate market actors have enormous resources at their disposal to act as they please and to avoid scrutiny and democratic control. Only similarly big, powerful, independently funded media organizations, whose basic market roles are to observe and criticize other large organizations, can match these established elite organizational actors. Individuals and collections of volunteers talking to each other may be nice, but they cannot seriously replace well-funded, economically and politically powerful media.

4.*Authoritarian countries can use filtering and monitoring to squelch Internet use.* A distinct set of claims and their critiques have to do with the effects of the Internet on authoritarian countries. The critique is leveled at a basic belief supposedly, and perhaps actually, held by some cyber-libertarians, that with enough access to Internet tools freedom will burst out everywhere. The argument is that China, more than any other country, shows that it is possible to allow a population access to the Internet - it is now home to the second-largest national population of Internet users - and still control that use quite substantially.

5.*Digital divide.* While the Internet may increase the circle of participants in the public sphere, access to its tools is skewed in favor of those who already are well-off in society - in terms of wealth, race, and skills. I do not respond to this critique in this chapter. First, in the United States, this is less stark today than it was in the late 1990s. Computers and Internet connections are becoming cheaper and more widely available in public libraries and schools. As they become more central to life, they seem to be reaching higher penetration rates, and growth rates among underrepresented groups are higher than the growth rate among the highly represented groups. The digital divide with regard to basic access within advanced economies is important as long as it persists, but seems to be a transitional problem. Moreover, it is important to recall that the democratizing effects of the Internet must be compared to democracy in the context of mass media, not in the context of an idealized utopia. Computer literacy and skills, while far from universal, are much more widely distributed than the skills and instruments of mass-media production. Second, I devote chapter 9 to the question of how and why the emergence specifically of nonmarket production provides new avenues for substantial improvements in equality of access to various desiderata that the market distributes unevenly, both within advanced economies and globally, where the maldistribution is much more acute. While the digital divide critique can therefore temper our enthusiasm for how radical the change represented by the networked information economy may be in terms of democracy, the networked information economy is itself an avenue for alleviating maldistribution.

The remainder of this chapter is devoted to responding to these critiques, providing a defense of the claim that the Internet can contribute to a more attractive liberal public sphere. As we work through these objections, we can develop a better understanding of how the networked information economy responds to or overcomes the particular systematic failures of mass media as platforms for the public sphere. Throughout this analysis, it is comparison of the attractiveness of the networked public sphere to that baseline - the mass-media-dominated public sphere - not comparison to a nonexistent ideal public sphere or to the utopia of "everyone a pamphleteer," that should matter most to our assessment of its democratic promise.

Is the Internet Too Chaotic, Too Concentrated, or Neither?

The first-generation critique of the claims that the Internet democratizes focused heavily on three variants of the information overload or Babel objection. The basic descriptive proposition that animated the Supreme Court in *Reno v. ACLU* was taken as more or less descriptively accurate: Everyone would be equally able to speak on the Internet. However, this basic observation was then followed by a descriptive or normative explanation of why this development was a threat to democracy, or at least not much of a boon. The basic problem that is diagnosed by this line of critique is the problem of attention. When everyone can speak, the central point of failure becomes the capacity to be heard - who listens to whom, and how that question is decided. Speaking in a medium that no one will actually hear with any reasonable likelihood may be psychologically satisfying, but it is not a move in a political conversation. Noam's prediction was, therefore, that there would be a reconcentration of attention: money would reemerge in this environment as a major determinant of the capacity to be heard, certainly no less, and perhaps even more so, than it

was in the mass-media environment./11 Sunstein's theory was different. He accepted Nicholas Negroponte's prediction that people would be reading "The Daily Me," that is, that each of us would create highly customized windows on the information environment that would be narrowly tailored to our unique combination of interests. From this assumption about how people would be informed, he spun out two distinct but related critiques. The first was that discourse would be fragmented. With no six o'clock news to tell us what is on the public agenda, there would be no public agenda, just a fragmented multiplicity of private agendas that never coalesce into a platform for political discussion. The second was that, in a fragmented discourse, individuals would cluster into groups of self-reinforcing, self-referential discussion groups. These types of groups, he argued from social scientific evidence, tend to render their participants' views more extreme and less amenable to the conversation across political divides necessary to achieve reasoned democratic decisions.

Extensive empirical and theoretical studies of actual use patterns of the Internet over the past five to eight years has given rise to a second-generation critique of the claim that the Internet democratizes. According to this critique, attention is much more concentrated on the Internet than we thought a few years ago: a tiny number of sites are highly linked, the vast majority of "speakers" are not heard, and the democratic potential of the Internet is lost. If correct, these claims suggest that Internet use patterns solve the problem of discourse fragmentation that Sunstein was worried about. Rather than each user reading a customized and completely different "newspaper," the vast majority of users turn out to see the same sites. In a network with a small number of highly visible sites that practically everyone reads, the discourse fragmentation problem is resolved. Because they are seen by most people, the polarization problem too is solved - the highly visible sites are not small-group interactions with homogeneous viewpoints. While resolving Sunstein's concerns, this pattern is certainly consistent with Noam's prediction that money would have to be paid to reach visibility, effectively replicating the mass-media model. While centralization would resolve the Babel objection, it would do so only at the expense of losing much of the democratic promise of the Net.

Therefore, we now turn to the question: Is the Internet in fact too chaotic or too concentrated to yield a more attractive democratic discourse than the mass media did? I suggest that neither is the case. At the risk of appearing a chimera of Goldilocks and Pangloss, I argue instead that the observed use of the network exhibits an order that is not too concentrated and not too chaotic, but rather, if not "just right," at least structures a networked public sphere more attractive than the mass-media-dominated public sphere.

There are two very distinct types of claims about Internet centralization. The first, and earlier, has the familiar ring of media concentration. It is the simpler of the two, and is tractable to policy. The second, concerned with the emergent patterns of attention and linking on an otherwise open network, is more difficult to explain and intractable to policy. I suggest, however, that it actually stabilizes and structures democratic discourse, providing a better answer to the fears of information overload than either the mass media or any efforts to regulate attention to matters of public concern.

The media-concentration type argument has been central to arguments about the necessity of open access to broadband platforms, made most forcefully over the past few years by Lawrence Lessig. The argument is that the basic instrumentalities of Internet communications are subject to concentrated markets. This market concentration in basic access becomes a potential point of concentration of the power to influence the discourse made possible by access. Eli Noam's recent work provides the most comprehensive study currently available of the degree of market concentration in media industries. It offers a bleak picture./12 Noam looked at markets in basic

infrastructure components of the Internet: Internet backbones, Internet service providers (ISPs), broadband providers, portals, search engines, browser software, media player software, and Internet telephony. Aggregating across all these sectors, he found that the Internet sector defined in terms of these components was, throughout most of the period from 1984 to 2002, concentrated according to traditional antitrust measures. Between 1992 and 1998, however, this sector was "highly concentrated" by the Justice Department's measure of market concentration for antitrust purposes. Moreover, the power of the top ten firms in each of these markets, and in aggregate for firms that had large market segments in a number of these markets, shows that an ever-smaller number of firms were capturing about 25 percent of the revenues in the Internet sector. A cruder, but consistent finding is the FCC's, showing that 96 percent of homes and small offices get their broadband access either from their incumbent cable operator or their incumbent local telephone carrier./13 It is important to recognize that these findings are suggesting potential points of failure for the networked information economy. They are not a critique of the democratic potential of the networked public sphere, but rather show us how we could fail to develop it by following the wrong policies.

The risk of concentration in broadband access services is that a small number of firms, sufficiently small to have economic power in the antitrust sense, will control the markets for the basic instrumentalities of Internet communications. Recall, however, that the low cost of computers and the open-ended architecture of the Internet protocol itself are the core enabling facts that have allowed us to transition from the mass-media model to the networked information model. As long as these basic instrumentalities are open and neutral as among uses, and are relatively cheap, the basic economics of nonmarket production described in part I should not change. Under competitive conditions, as technology makes computation and communications cheaper, a well-functioning market should ensure that outcome. Under oligopolistic conditions, however, there is a threat that the network will become too expensive to be neutral as among market and nonmarket production. If basic upstream network connections, server space, and up-to-date reading and writing utilities become so expensive that one needs to adopt a commercial model to sustain them, then the basic economic characteristic that typifies the networked information economy - the relatively large role of nonproprietary, nonmarket production - will have been reversed. However, the risk is not focused solely or even primarily on explicit pricing. One of the primary remaining scarce resources in the networked environment is user time and attention. As chapter 5 explained, owners of communications facilities can extract value from their users in ways that are more subtle than increasing price. In particular, they can make some sites and statements easier to reach and see - more prominently displayed on the screen, faster to load - and sell that relative ease to those who are willing to pay./14 In that environment, nonmarket sites are systematically disadvantaged irrespective of the quality of their content.

The critique of concentration in this form therefore does not undermine the claim that the networked information economy, if permitted to flourish, will improve the democratic public sphere. It underscores the threat of excessive monopoly in infrastructure to the sustainability of the networked public sphere. The combination of observations regarding market concentration and an understanding of the importance of a networked public sphere to democratic societies suggests that a policy intervention is possible and desirable. Chapter 11 explains why the relevant intervention is to permit substantial segments of the core common infrastructure - the basic physical transport layer of wireless or fiber and the software and standards that run communications - to be produced and provisioned by users and managed as a commons.

On Power Law Distributions, Network Topology, and Being Heard

A much more intractable challenge to the claim that the networked information economy will democratize the public sphere emerges from observations of a set or phenomena that characterize the Internet, the Web, the blogosphere, and, indeed, most growing networks. In order to extract information out of the universe of statements and communications made possible by the Internet, users are freely adopting practices that lead to the emergence of a new hierarchy. Rather than succumb to the "information overload" problem, users are solving it by congregating in a small number of sites. This conclusion is based on a new but growing literature on the likelihood that a Web page will be linked to by others. The distribution of that probability turns out to be highly skew. That is, there is a tiny probability that any given Web site will be linked to by a huge number of people, and a very large probability that for a given Web site only one other site, or even no site, will link to it. This fact is true of large numbers of very different networks described in physics, biology, and social science, as well as in communications networks. If true in this pure form about Web usage, this phenomenon presents a serious theoretical and empirical challenge to the claim that Internet communications of the sorts we have seen here meaningfully decentralize democratic discourse. It is not a problem that is tractable to policy. We cannot as a practical matter force people to read different things than what they choose to read; nor should we wish to. If users avoid information overload by focusing on a small subset of sites in an otherwise open network that allows them to read more or less whatever they want and whatever anyone has written, policy interventions aimed to force a different pattern would be hard to justify from the perspective of liberal democratic theory.

The sustained study of the distribution of links on the Internet and the Web is relatively new - only a few years old. There is significant theoretical work in a field of mathematics called graph theory, or network topology, on power law distributions in networks, on skew distributions that are not pure power law, and on the mathematically related small-worlds phenomenon in networks. The basic intuition is that, if indeed a tiny minority of sites gets a large number of links, and the vast majority gets few or no links, it will be very difficult to be seen unless you are on the highly visible site. Attention patterns make the open network replicate mass media. While explaining this literature over the next few pages, I show that what is in fact emerging is very different from, and more attractive than, the mass-media-dominated public sphere.

While the Internet, the Web, and the blogosphere are indeed exhibiting much greater order than the freewheeling, "everyone a pamphleteer" image would suggest, this structure does not replicate a mass-media model. We are seeing a newly shaped information environment, where indeed few are read by many, but clusters of moderately read sites provide platforms for vastly greater numbers of speakers than were heard in the mass-media environment. Filtering, accreditation, synthesis, and salience are created through a system of peer review by information affinity groups, topical or interest based. These groups filter the observations and opinions of an enormous range of people, and transmit those that pass local peer review to broader groups and ultimately to the polity more broadly, without recourse to market-based points of control over the information flow. Intense interest and engagement by small groups that share common concerns, rather than lowest-common-denominator interest in wide groups that are largely alienated from each other, is what draws attention to statements and makes them more visible. This makes the emerging networked public sphere more responsive to intensely held concerns of a much wider swath of the population than the mass media were capable of seeing, and creates a communications process that is more resistant to corruption by money.

In what way, first, is attention concentrated on the Net? We are used to seeing probability distributions that describe social phenomena following a Gaussian distribution: where the mean and the median are the same and the probabilities fall off symmetrically as we describe events that are

farther from the median. This is the famous Bell Curve. Some phenomena, however, observed initially in Pareto's work on income distribution and Zipf's on the probability of the use of English words in text and in city populations, exhibit completely different probability distributions. These distributions have very long "tails" - that is, they are characterized by a very small number of very high-yield events (like the number of words that have an enormously high probability of appearing in a randomly chosen sentence, like "the" or "to") and a very large number of events that have a very low probability of appearing (like the probability that the word "probability" or "blogosphere" will appear in a randomly chosen sentence). To grasp intuitively how unintuitive such distributions are to us, we could think of radio humorist Garrison Keillor's description of the fictitious Lake Wobegon, where "all the children are above average." That statement is amusing because we assume intelligence follows a normal distribution. If intelligence were distributed according to a power law, most children there would actually be below average - the median is well below the mean in such distributions (see figure 7.4). Later work by Herbert Simon in the 1950s, and by Derek de Solla Price in the 1960s, on cumulative advantage in scientific citations/15 presaged an emergence at the end of the 1990s of intense interest in power law characterizations of degree distributions, or the number of connections any point in a network has to other points, in many kinds of networks - from networks of neurons and axons, to social networks and communications and information networks.
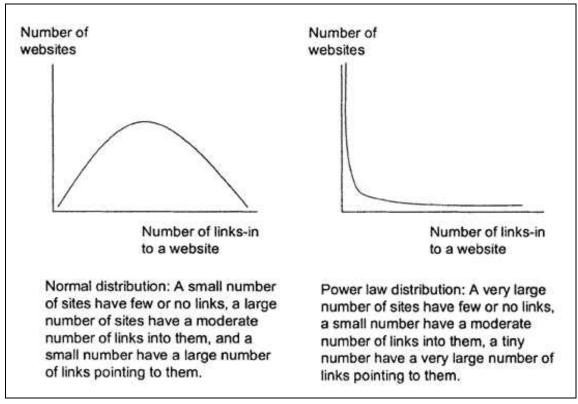
Figure 7.4: Illustration of How Normal Distribution and Power Law Distribution Would Differ in Describing How Many Web Sites Have Few or Many Links Pointing at Them

The Internet and the World Wide Web offered a testable setting, where large-scale investigation could be done automatically by studying link structure (who is linked-in to and by whom, who links out and to whom, how these are related, and so on), and where the practical applications of better understanding were easily articulated - such as the design of better search engines. In 1999, Albert-László Barabási and Reka Albert published a paper in *Science* showing that a variety of networked phenomena have a predictable topology: The distribution of links into and out of nodes on the network follows a power law. There is a very low probability that any vertex, or node, in the network will be very highly connected to many others, and a very large probability that a very large number of nodes will be connected only very loosely, or perhaps not at all. Intuitively, a lot of Web sites link to information that is located on Yahoo!, while very few link to any randomly selected individual's Web site. Barabási and Albert hypothesized a mechanism for this distribution to evolve, which they called "preferential attachment." That is, new nodes prefer to attach to already well-attached nodes. Any network that grows through the addition of new nodes, and in which nodes preferentially attach to nodes that are already well attached, will eventually exhibit this distribution./16 In other words, the rich get richer. At the same time, two computer scientists, Lada Adamic and Bernardo Huberman, published a study in *Nature* that identified the presence of power law distributions in the number of Web pages in a given site. They hypothesized not that new nodes preferentially attach to old ones, but that each site has an intrinsically different growth rate, and that new sites are formed at an exponential rate./17 The intrinsically different growth rates could be interpreted as quality, interest, or perhaps investment of money in site development and marketing. They showed that on these assumptions, a power law distribution would emerge. Since the publication of these articles we have seen an explosion of theoretical and empirical literature on graph theory, or the structure and growth of networks, and particularly on link structure in the World Wide Web. It has consistently shown that the number of links into and out of Web sites follows power laws and that the exponent (the exponential factor that determines that the drop-off between the most linked-to site and the second most linked-to site, and the third, and so on, will be so dramatically rapid, and how rapid it is) for inlinks is roughly 2.1 and for outlinks 2.7.

If one assumes that most people read things by either following links, or by using a search engine, like Google, that heavily relies on counting inlinks to rank its results, then it is likely that the number of visitors to a Web page, and more recently, the number of readers of blogs, will follow a similarly highly skew distribution. The implication for democracy that comes most immediately to mind is dismal. While, as the Supreme Court noted with enthusiasm, on the Internet everyone can be a pamphleteer or have their own soapbox, the Internet does not, in fact, allow individuals to be heard in ways that are substantially more effective than standing on a soapbox in a city square. Many Web pages and blogs will simply go unread, and will not contribute to a more engaged polity. This argument was most clearly made in Barabási's popularization of his field, *Linked*: "The most intriguing result of our Web-mapping project was the *complete* absence of democracy, fairness, and egalitarian values on the Web. We learned that the topology of the Web prevents us from seeing anything but a mere handful of the billion documents out there."/18

The stories offered in this chapter and throughout this book present a puzzle for this interpretation of the power law distribution of links in the network as re-creating a concentrated medium. The success of Nick Davis's site, BoycottSBG, would be a genuine fluke. The probability that such a site could be established on a Monday, and by Friday of the same week would have had three hundred thousand unique visitors and would have orchestrated a successful campaign, is so small as to be negligible. The probability that a completely different site, StopSinclair.org, of equally network-obscure origins, would be established on the very same day and also successfully catch the attention of enough readers to collect 150,000 signatures on a petition to protest Sinclair's broadcast, rather than wallowing undetected in the mass of self-published angry commentary, is

practically insignificant. And yet, intuitively, it seems unsurprising that a large population of individuals who are politically mobilized on the same side of the political map and share a political goal in the public sphere - using a network that makes it trivially simple to set up new points of information and coordination, tell each other about them, and reach and use them from anywhere - would, in fact, inform each other and gather to participate in a political demonstration. We saw that the boycott technique that Davis had designed his Web site to facilitate was discussed on TalkingPoints - a site near the top of the power law distribution of political blogs - but that it was a proposal by an anonymous individual who claimed to know what makes local affiliates tick, not of TalkingPoints author Josh Marshall. By midweek, after initially stoking the fires of support for Davis's boycott, Marshall had stepped back, and Davis's site became the clearing point for reports, tactical conversations, and mobilization. Davis not only was visible, but rather than being drowned out by the high-powered transmitter, TalkingPoints, his relationship with the high-visibility site was part of his success. This story alone cannot, of course, "refute" the power law distribution of network links, nor is it offered as a refutation. It does, however, provide a context for looking more closely at the emerging understanding of the topology of the Web, and how it relates to the fears of concentration of the Internet, and the problems of information overload, discourse fragmentation, and the degree to which money will come to dominate such an unstructured and wide-open environment. It suggests a more complex story than simply "the rich get richer" and "you might speak, but no one will hear you." In this case, the topology of the network allowed rapid emergence of a position, its filtering and synthesis, and its rise to salience. Network topology helped facilitate all these components of the public sphere, rather than undermined them. We can go back to the mathematical and computer science literature to begin to see why.

Within two months of the publication of Barabási and Albert's article, Adamic and Huberman had published a letter arguing that, if Barabási and Albert were right about preferential attachment, then older sites should systematically be among those that are at the high end of the distribution, while new ones will wallow in obscurity. The older sites are already attached, so newer sites would preferentially attach to the older sites. This, in turn, would make them even more attractive when a new crop of Web sites emerged and had to decide which sites to link to. In fact, however, Adamic and Huberman showed that there is no such empirical correlation among Web sites. They argued that their mechanism - that nodes have intrinsic growth rates that are different - better describes the data. In their response, Barabási and Albert showed that on their data set, the older nodes are actually more connected in a way that follows a power law, but only on average - that is to say, the average number of connections of a class of older nodes related to the average number of links to a younger class of nodes follows a power law. This argued that their basic model was sound, but required that they modify their equations to include something similar to what Huberman and Adamic had proposed - an intrinsic growth factor for each node, as well as the preferential connection of new nodes to established nodes.[/19] This modification is important because it means that not every new node is doomed to be unread relative to the old ones, only that on average they are much less likely to be read. It makes room for rapidly growing new nodes, but does not theorize what might determine the rate of growth. It is possible, for example, that money could determine growth rates: In order to be seen, new sites or statements would have to spend money to gain visibility and salience. As the BoycottSBG and Diebold stories suggest, however, as does the Lott story described later in this chapter, there are other ways of achieving immediate salience. In the case of BoycottSBG, it was providing a solution that resonated with the political beliefs of many people and was useful to them for their expression and mobilization. Moreover, the continued presence of preferential attachment suggests that noncommercial Web sites that are already highly connected because of the time they were introduced (like the Electronic Frontier Foundation), because of their internal attractiveness to large communities (like Slashdot), or because of their

salience to the immediate interests of users (like BoycottSBG), will have persistent visibility even in the face of large infusions of money by commercial sites.

Developments in network topology theory and its relationship to the structure of the empirically mapped real Internet offer a map of the networked information environment that is indeed quite different from the naïve model of "everyone a pamphleteer." To the limited extent that these findings have been interpreted for political meaning, they have been seen as a disappointment - the real world, as it turns out, does not measure up to anything like that utopia. However, that is the wrong baseline. There never has been a complex, large modern democracy in which everyone could speak and be heard by everyone else. The correct baseline is the one-way structure of the commercial mass media. The normatively relevant descriptive questions are whether the networked public sphere provides broader intake, participatory filtering, and relatively incorruptible platforms for creating public salience. I suggest that it does. Four characteristics of network topology structure the Web and the blogosphere in an ordered, but nonetheless meaningfully participatory form. First, at a microlevel, sites cluster - in particular, topically and interest-related sites link much more heavily to each other than to other sites. Second, at a macrolevel, the Web and the blogosphere have giant, strongly connected cores - "areas" where 20-30 percent of all sites are highly and redundantly interlinked; that is, tens or hundreds of millions of sites, rather than ten, fifty, or even five hundred television stations. That pattern repeats itself in smaller subclusters as well. Third, as the clusters get small enough, the obscurity of sites participating in the cluster diminishes, while the visibility of the superstars remains high, forming a filtering and transmission backbone for universal intake and local filtering. Fourth and finally, the Web exhibits "small-world" phenomena, making most Web sites reachable through shallow paths from most other Web sites. I will explain each of these below, as well as how they interact to form a reasonably attractive image of the networked public sphere.

First, links are not smoothly distributed throughout the network. Sites cluster into densely linked "regions" or communities of interest. Computer scientists have looked at clustering from the perspective of what topical or other correlated characteristics describe these relatively high-density interconnected regions of nodes. What they found was perhaps entirely predictable from an intuitive perspective of the network users, but important as we try to understand the structure of information flow on the Web. Web sites cluster into topical and social/organizational clusters. Early work done in the IBM Almaden Research Center on how link structure could be used as a search technique showed that by mapping densely interlinked sites without looking at content, one could find communities of interest that identify very fine-grained topical connections, such as Australian fire brigades or Turkish students in the United States.[20] A later study out of the NEC Research Institute more formally defined the interlinking that would identify a "community" as one in which the nodes were more densely connected to each other than they were to nodes outside the cluster by some amount. The study also showed that topically connected sites meet this definition. For instance, sites related to molecular biology clustered with each other - in the sense of being more interlinked with each other than with off-topic sites - as did sites about physics and black holes.[21] Lada Adamic and Natalie Glance recently showed that liberal political blogs and conservative political blogs densely interlink with each other, mostly pointing within each political leaning but with about 15 percent of links posted by the most visible sites also linking across the political divide.[22] Physicists analyze clustering as the property of transitivity in networks: the increased probability that if node A is connected to node B, and node B is connected to node C, that node A also will be connected to node C, forming a triangle. Newman has shown that the clustering coefficient of a network that exhibits power law distribution of connections or degrees - that is, its tendency to cluster - is related to the exponent of the distribution. At low exponents, below 2.333, the clustering coefficient becomes high. This explains analytically the empirically observed high level of clustering on the Web, whose exponent for inlinks has been empirically shown to be 2.1.[23]

Second, at a macrolevel and in smaller subclusters, the power law distribution does not resolve into everyone being connected in a mass-media model relationship to a small number of major "backbone" sites. As early as 1999, Broder and others showed that a very large number of sites occupy what has been called a giant, strongly connected core./24 That is, nodes within this core are heavily linked and interlinked, with multiple redundant paths among them. Empirically, as of 2001, this structure was comprised of about 28 percent of nodes. At the same time, about 22 percent of nodes had links into the core, but were not linked to from it - these may have been new sites, or relatively lower-interest sites. The same proportion of sites was linked-to from the core, but did not link back to it - these might have been ultimate depositories of documents, or internal organizational sites. Finally, roughly the same proportion of sites occupied "tendrils" or "tubes" that cannot reach, or be reached from, the core. Tendrils can be reached from the group of sites that link into the strongly connected core or can reach into the group that can be connected to from the core. Tubes connect the inlinking sites to the outlinked sites without going through the core. About 10 percent of sites are entirely isolated. This structure has been called a "bow tie" - with a large core and equally sized in- and outflows to and from that core (see figure 7.5).
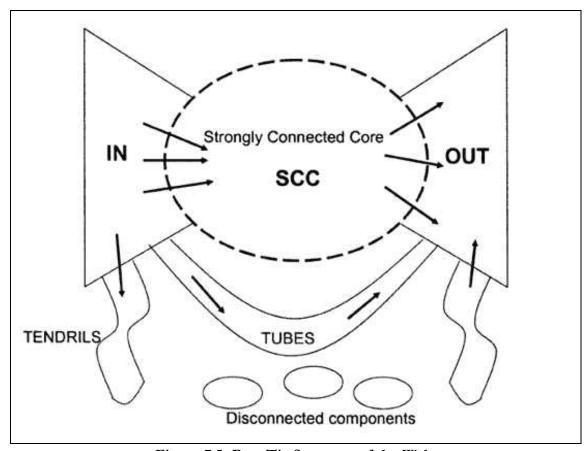
Figure 7.5: Bow Tie Structure of the Web

One way of interpreting this structure as counterdemocratic is to say: This means that half of all Web sites are not reachable from the other half - the "IN," "tendrils," and disconnected portions cannot be reached from any of the sites in SCC and OUT. This is indeed disappointing from the "everyone a pamphleteer" perspective. On the other hand, one could say that half of all Web pages, the SCC and OUT components, *are* reachable from IN and SCC. That is, hundreds of millions of pages are reachable from hundreds of millions of potential entry points. This represents a very different intake function and freedom to speak in a way that is potentially accessible to others than a five-hundred-channel, mass-media model. More significant yet, Dill and others showed that the bow tie structure appears not only at the level of the Web as a whole, but repeats itself within clusters. That is, the Web appears to show characteristics of self-similarity, up to a point - links within clusters also follow a power law distribution and cluster, and have a bow tie structure of similar proportions to that of the overall Web. Tying the two points about clustering and the presence of a strongly connected core, Dill and his coauthors showed that what they called "thematically unified clusters," such as geographically or content-related groupings of Web sites, themselves exhibit these strongly connected cores that provided a thematically defined navigational backbone to the Web. It is not that one or two major sites were connected to by all thematically related sites; rather, as at the network level, on the order of 25-30 percent were highly interlinked, and another 25 percent were reachable from within the strongly connected core./25 Moreover, when the data was pared down to treat only the home page, rather than each Web page within a single site as a distinct "node" (that is, everything that came under www.foo.com was treated as one node, as opposed to the usual method where www.foo.com, www.foo.com/nonsuch, and www.foo.com/somethingelse are each treated as a separate node), fully 82 percent of the nodes were in the strongly connected core, and an additional 13 percent were reachable from the SCC as the OUT group.

Third, another finding of Web topology and critical adjustment to the basic Barabási and Albert model is that when the topically or organizationally related clusters become small enough - on the order of hundreds or even low thousands of Web pages - they no longer follow a pure power law distribution. Instead, they follow a distribution that still has a very long tail - these smaller clusters still have a few genuine "superstars" - but the body of the distribution is substantially more moderate: beyond the few superstars, the shape of the link distribution looks a little more like a normal distribution. Instead of continuing to drop off exponentially, many sites exhibit a moderate degree of connectivity. Figure 7.6 illustrates how a hypothetical distribution of this sort would differ both from the normal and power law distributions illustrated in figure 7.4. David Pennock and others, in their paper describing these empirical findings, hypothesized a uniform component added to the purely exponential original Barabási and Albert model. This uniform component could be random (as they modeled it), but might also stand for quality of materials, or level of interest in the site by participants in the smaller cluster. At large numbers of nodes, the exponent dominates the uniform component, accounting for the pure power law distribution when looking at the Web as a whole, or even at broadly defined topics. In smaller clusters of sites, however, the uniform component begins to exert a stronger pull on the distribution. The exponent keeps the long tail intact, but the uniform component accounts for a much more moderate body. Many sites will have dozens, or even hundreds of links. The Pennock paper looked at sites whose number was reduced by looking only at sites of certain organizations - universities or public companies. Chakrabarti and others later confirmed this finding for topical clusters as well. That is, when they looked at small clusters of topically related sites, the distribution of links still has a long tail for a small number of highly connected sites in every topic, but the body of the distribution diverges from a power law distribution, and represents a substantial proportion of sites that are moderately linked./26 Even more specifically, Daniel Drezner and Henry Farrell reported that the Pennock modification better describes distribution of links specifically to and among political blogs./27
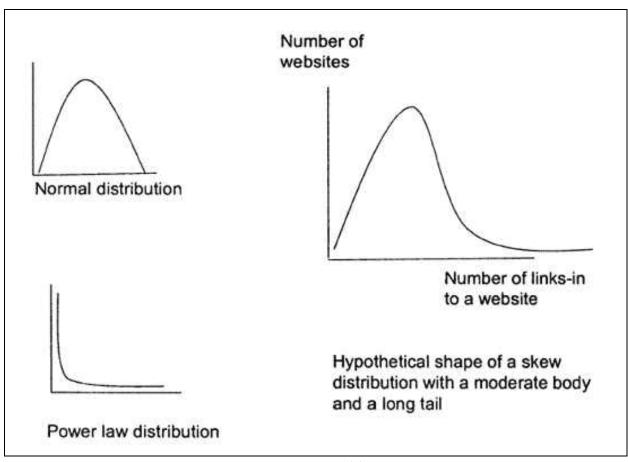
Figure 7.6: Illustration of a Skew Distribution That Does Not Follow a Power Law

These findings are critical to the interpretation of the distribution of links as it relates to human attention and communication. There is a big difference between a situation where no one is looking at any of the sites on the low end of the distribution, because everyone is looking only at the superstars, and a situation where dozens or hundreds of sites at the low end are looking at each other, as well as at the superstars. The former leaves all but the very few languishing in obscurity, with no one to look at them. The latter, as explained in more detail below, offers a mechanism for topically related and interest-based clusters to form a peer-reviewed system of filtering, accreditation, and salience generation. It gives the long tail on the low end of the distribution heft (and quite a bit of wag).

The fourth and last piece of mapping the network as a platform for the public sphere is called the "small-worlds effect." Based on Stanley Milgram's sociological experiment and on mathematical models later proposed by Duncan Watts and Steven Strogatz, both theoretical and empirical work has shown that the number of links that must be traversed from any point in the network to any other point is relatively small./28 Fairly shallow "walks" - that is, clicking through three or four layers of links - allow a user to cover a large portion of the Web.

What is true of the Web as a whole turns out to be true of the blogosphere as well, and even of the specifically political blogosphere. Early 2003 saw increasing conversations in the blogosphere about the emergence of an "A-list," a number of highly visible blogs that were beginning to seem more like mass media than like blogs. In two blog-based studies, Clay Shirky and then Jason Kottke

published widely read explanations of how the blogosphere was simply exhibiting the power law characteristics common on the Web.[/29](#) The emergence in 2003 of discussions of this sort in the blogosphere is, it turns out, hardly surprising. In a time-sensitive study also published in 2003, Kumar and others provided an analysis of the network topology of the blogosphere. They found that it was very similar to that of the Web as a whole - both at the macro- and microlevels. Interestingly, they found that the strongly connected core only developed after a certain threshold, in terms of total number of nodes, had been reached, and that it began to develop extensively only in 2001, reached about 20 percent of all blogs in 2002, and continued to grow rapidly. They also showed that what they called the "community" structure - the degree of clustering or mutual pointing within groups - was high, an order of magnitude more than a random graph with a similar power law exponent would have generated. Moreover, the degree to which a cluster is active or inactive, highly connected or not, changes over time. In addition to time-insensitive superstars, there are also flare-ups of connectivity for sites depending on the activity and relevance of their community of interest. This latter observation is consistent with what we saw happen for BoycottSBG.com. Kumar and his collaborators explained these phenomena by the not-too-surprising claim that bloggers link to each other based on topicality - that is, their judgment of the quality and relevance of the materials - not only on the basis of how well connected they are already.[/30](#)

This body of literature on network topology suggests a model for how order has emerged on the Internet, the World Wide Web, and the blogosphere. The networked public sphere allows hundreds of millions of people to publish whatever and whenever they please without disintegrating into an unusable cacophony, as the first-generation critics argued, and it filters and focuses attention without re-creating the highly concentrated model of the mass media that concerned the second-generation critique. We now know that the network at all its various layers follows a degree of order, where some sites are vastly more visible than most. This order is loose enough, however, and exhibits a sufficient number of redundant paths from an enormous number of sites to another enormous number, that the effect is fundamentally different from the small number of commercial professional editors of the mass media.

Individuals and individual organizations cluster around topical, organizational, or other common features. At a sufficiently fine-grained degree of clustering, a substantial proportion of the clustered sites are moderately connected, and each can therefore be a point of intake that will effectively transmit observations or opinions within and among the users of that topical or interest-based cluster. Because even in small clusters the distribution of links still has a long tail, these smaller clusters still include high-visibility nodes. These relatively high-visibility nodes can serve as points of transfer to larger clusters, acting as an attention backbone that transmits information among clusters. Subclusters within a general category - such as liberal and conservative blogs clustering within the broader cluster of political blogs - are also interlinked, though less densely than within-cluster connectivity. The higher level or larger clusters again exhibit a similar feature, where higher visibility nodes can serve as clearinghouses and connectivity points among clusters and across the Web. These are all highly connected with redundant links within a giant, strongly connected core - comprising more than a quarter of the nodes in any given level of cluster. The small-worlds phenomenon means that individual users who travel a small number of different links from similar starting points within a cluster cover large portions of the Web and can find diverse sites. By then linking to them on their own Web sites, or giving them to others by e-mail or blog post, sites provide multiple redundant paths open to many users to and from most statements on the Web. High-visibility nodes amplify and focus on given statements, and in this regard, have greater power in the information environment they occupy. However, there is sufficient redundancy of paths through high-visibility nodes that no single node or small collection of nodes

can control the flow of information in the core and around the Web. This is true both at the level of the cluster and at the level of the Web as a whole.

The result is an ordered system of intake, filtering, and synthesis that can in theory emerge in networks generally, and empirically has been shown to have emerged on the Web. It does not depend on single points of control. It avoids the generation of a din through which no voice can be heard, as the fears of fragmentation predicted. And, while money may be useful in achieving visibility, the structure of the Web means that money is neither necessary nor sufficient to grab attention - because the networked information economy, unlike its industrial predecessor, does not offer simple points of dissemination and control for purchasing assured attention. What the network topology literature allows us to do, then, is to offer a richer, more detailed, and empirically supported picture of how the network can be a platform for the public sphere that is structured in a fundamentally different way than the mass-media model. The problem is approached through a self-organizing principle, beginning with communities of interest on smallish scales, practices of mutual pointing, and the fact that, with freedom to choose what to see and who to link to, with some codependence among the choices of individuals as to whom to link, highly connected points emerge even at small scales, and continue to be replicated with ever-larger visibility as the clusters grow. Without forming or requiring a formal hierarchy, and without creating single points of control, each cluster generates a set of sites that offer points of initial filtering, in ways that are still congruent with the judgments of participants in the highly connected small cluster. The process is replicated at larger and more general clusters, to the point where positions that have been synthesized "locally" and "regionally" can reach Web-wide visibility and salience. It turns out that we are not intellectual lemmings. We do not use the freedom that the network has made possible to plunge into the abyss of incoherent babble. Instead, through iterative processes of cooperative filtering and "transmission" through the high visibility nodes, the low-end thin tail turns out to be a peer-produced filter and transmission medium for a vastly larger number of speakers than was imaginable in the mass-media model.

The effects of the topology of the network are reinforced by the cultural forms of linking, e-mail lists, and the writable Web. The network topology literature treats every page or site as a node. The emergence of the writable Web, however, allows each node to itself become a cluster of users and posters who, collectively, gain salience as a node. Slashdot is "a node" in the network as a whole, one that is highly linked and visible. Slashdot itself, however, is a highly distributed system for peer production of observations and opinions about matters that people who care about information technology and communications ought to care about. Some of the most visible blogs, like the dailyKos, are cooperative blogs with a number of authors. More important, the major blogs receive input - through posts or e-mails - from their users. Recall, for example, that the original discussion of a Sinclair boycott that would focus on local advertisers arrived on TalkingPoints through an e-mail comment from a reader. Talkingpoints regularly solicits and incorporates input from and research by its users. The cultural practice of writing to highly visible blogs with far greater ease than writing a letter to the editor and with looser constraints on what gets posted makes these nodes themselves platforms for the expression, filtering, and synthesis of observations and opinions. Moreover, as Drezner and Farrell have shown, blogs have developed cultural practices of mutual citation - when one blogger finds a source by reading another, the practice is to link to the original blog, not only directly to the underlying source. Jack Balkin has argued that the culture of linking more generally and the "see for yourself" culture also significantly militate against fragmentation of discourse, because users link to materials they are commenting on, even in disagreement.

Our understanding of the emerging structure of the networked information environment, then, provides the basis for a response to the family of criticisms of the first generation claims that the Internet democratizes. Recall that these criticisms, rooted in the problem of information overload, or the Babel objection, revolved around three claims. The first claim was that the Internet would result in a fragmentation of public discourse. The clustering of topically related sites, such as politically oriented sites, and of communities of interest, the emergence of high-visibility sites that the majority of sites link to, and the practices of mutual linking show quantitatively and qualitatively what Internet users likely experience intuitively. While there is enormous diversity on the Internet, there are also mechanisms and practices that generate a common set of themes, concerns, and public knowledge around which a public sphere can emerge. Any given site is likely to be within a very small number of clicks away from a site that is visible from a very large number of other sites, and these form a backbone of common materials, observations, and concerns. All the findings of power law distribution of linking, clustering, and the presence of a strongly connected core, as well as the linking culture and "see for yourself," oppose the fragmentation prediction. Users self-organize to filter the universe of information that is generated in the network. This self-organization includes a number of highly salient sites that provide a core of common social and cultural experiences and knowledge that can provide the basis for a common public sphere, rather than a fragmented one.

The second claim was that fragmentation would cause polarization. Because like-minded people would talk only to each other, they would tend to amplify their differences and adopt more extreme versions of their positions. Given that the evidence demonstrates there is no fragmentation, in the sense of a lack of a common discourse, it would be surprising to find higher polarization because of the Internet. Moreover, as Balkin argued, the fact that the Internet allows widely dispersed people with extreme views to find each other and talk is not a failure for the liberal public sphere, though it may present new challenges for the liberal state in constraining extreme action. Only polarization of discourse in society as a whole can properly be considered a challenge to the attractiveness of the networked public sphere. However, the practices of linking, "see for yourself," or quotation of the position one is criticizing, and the widespread practice of examining and criticizing the assumptions and assertions of one's interlocutors actually point the other way, militating against polarization. A potential counterargument, however, was created by the most extensive recent study of the political blogosphere. In that study, Adamic and Glance showed that only about 10 percent of the links on any randomly selected political blog linked to a site across the ideological divide. The number increased for the "A-list" political blogs, which linked across the political divide about 15 percent of the time. The picture that emerges is one of distinct "liberal" and "conservative" spheres of conversation, with very dense links within, and more sparse links between them. On one interpretation, then, although there are salient sites that provide a common subject matter for discourse, actual conversations occur in distinct and separate spheres - exactly the kind of setting that Sunstein argued would lead to polarization. Two of the study's findings, however, suggest a different interpretation. The first was that there was still a substantial amount of cross-divide linking. One out of every six or seven links in the top sites on each side of the divide linked to the other side in roughly equal proportions (although conservatives tended to link slightly more overall - both internally and across the divide). The second was, that in an effort to see whether the more closely interlinked conservative sites therefore showed greater convergence "on message," Adamic and Glance found that greater interlinking did not correlate with less diversity in external (outside of the blogosphere) reference points.[31] Together, these findings suggest a different interpretation. Each cluster of more or less like-minded blogs tended to read each other and quote each other much more than they did the other side. This operated not so much as an echo chamber as a forum for working out of observations and interpretations internally, among like-minded people. Many of these initial statements or inquiries die because the community finds them uninteresting or fruitless.

Some reach greater salience, and are distributed through the high-visibility sites throughout the community of interest. Issues that in this form reached political salience became topics of conversation and commentary across the divide. This is certainly consistent with both the BoycottSBG and Diebold stories, where we saw a significant early working out of strategies and observations before the criticism reached genuine political salience. There would have been no point for opponents to link to and criticize early ideas kicked around within the community, like opposing Sinclair station renewal applications. Only after a few days, when the boycott was crystallizing, would opponents have reason to point out the boycott effort and discuss it. This interpretation also well characterizes the way in which the Trent Lott story described later in this chapter began percolating on the liberal side of the blogosphere, but then migrated over to the center-right.

The third claim was that money would reemerge as the primary source of power brokerage because of the difficulty of getting attention on the Net. Descriptively, it shares a prediction with the second-generation claims: Namely, that the Internet will centralize discourse. It differs in the mechanism of concentration: it will not be the result of an emergent property of large-scale networks, but rather of an old, tried-and-true way of capturing the political arena - money. But the peer-production model of filtering and discussion suggests that the networked public sphere will be substantially less corruptible by money. In the interpretation that I propose, filtering for the network as a whole is done as a form of nested peer-review decisions, beginning with the speaker's closest information affinity group. Consistent with what we have been seeing in more structured peer-production projects like *Wikipedia*, Slashdot, or free software, communities of interest use clustering and mutual pointing to peer produce the basic filtering mechanism necessary for the public sphere to be effective and avoid being drowned in the din of the crowd. The nested structure of the Web, whereby subclusters form relatively dense higher-level clusters, which then again combine into even higher-level clusters, and in each case, have a number of high-end salient sites, allows for the statements that pass these filters to become globally salient in the relevant public sphere. This structure, which describes the analytic and empirical work on the Web as a whole, fits remarkably well as a description of the dynamics we saw in looking more closely at the success of the boycott on Sinclair, as well as the successful campaign to investigate and challenge Diebold's voting machines.

The peer-produced structure of the attention backbone suggests that money is neither necessary nor sufficient to attract attention in the networked public sphere (although nothing suggests that money has become irrelevant to political attention given the continued importance of mass media). It renders less surprising Howard Dean's strong campaign for the Democratic presidential primaries in 2003 and the much more stable success of MoveOn.org since the late 1990s. These suggest that attention on the network has more to do with mobilizing the judgments, links, and cooperation of large bodies of small-scale contributors than with applying large sums of money. There is no obvious broadcast station that one can buy in order to assure salience. There are, of course, the highly visible sites, and they do offer a mechanism of getting your message to large numbers of people. However, the degree of engaged readership, interlinking, and clustering suggests that, in fact, being exposed to a certain message in one or a small number of highly visible places accounts for only a small part of the range of "reading" that gets done. More significantly, it suggests that reading, as opposed to having a conversation, is only part of what people do in the networked environment. In the networked public sphere, receiving information or getting out a finished message are only parts, and not necessarily the most important parts, of democratic discourse. The central desideratum of a political campaign that is rooted in the Internet is the capacity to engage users to the point that they become effective participants in a conversation and an effort; one that they have a genuine stake in and that is linked to a larger, society-wide debate. This engagement is

not easily purchased, nor is it captured by the concept of a well-educated public that receives all the information it needs to be an informed citizenry. Instead, it is precisely the varied modes of participation in small-, medium-, and large-scale conversations, with varied but sustained degrees of efficacy, that make the public sphere of the networked environment different, and more attractive, than was the mass-media-based public sphere.

The networked public sphere is not only more resistant to control by money, but it is also less susceptible to the lowest-common-denominator orientation that the pursuit of money often leads mass media to adopt. Because communication in peer-produced media starts from an intrinsic motivation - writing or commenting about what one cares about - it begins with the opposite of lowest common denominator. It begins with what irks you, the contributing peer, individually, the most. This is, in the political world, analogous to Eric Raymond's claim that every free or open-source software project begins with programmers with an itch to scratch - something directly relevant to their lives and needs that they want to fix. The networked information economy, which makes it possible for individuals alone and in cooperation with others to scour the universe of politically relevant events, to point to them, and to comment and argue about them, follows a similar logic. This is why one freelance writer with lefty leanings, Russ Kick, is able to maintain a Web site, The Memory Hole, with documents that he gets by filing Freedom of Information Act requests. In April 2004, Kick was the first to obtain the U.S. military's photographs of the coffins of personnel killed in Iraq being flown home. No mainstream news organization had done so, but many published the photographs almost immediately after Kick had obtained them. Like free software, like Davis and the bloggers who participated in the debates over the Sinclair boycott, or the students who published the Diebold e-mails, the decision of what to publish does not start from a manager's or editor's judgment of what would be relevant and interesting to many people without being overly upsetting to too many others. It starts with the question: What do I care about most now?

To conclude, we need to consider the attractiveness of the networked public sphere not from the perspective of the mid-1990s utopianism, but from the perspective of how it compares to the actual media that have dominated the public sphere in all modern democracies. The networked public sphere provides an effective nonmarket alternative for intake, filtering, and synthesis outside the market-based mass media. This nonmarket alternative can attenuate the influence over the public sphere that can be achieved through control over, or purchase of control over, the mass media. It offers a substantially broader capture basin for intake of observations and opinions generated by anyone with a stake in the polity, anywhere. It appears to have developed a structure that allows for this enormous capture basin to be filtered, synthesized, and made part of a polity-wide discourse. This nested structure of clusters of communities of interest, typified by steadily increasing visibility of superstar nodes, allows for both the filtering and salience to climb up the hierarchy of clusters, but offers sufficient redundant paths and interlinking to avoid the creation of a small set of points of control where power can be either directly exercised or bought.

There is, in this story, an enormous degree of contingency and factual specificity. That is, my claims on behalf of the networked information economy as a platform for the public sphere are not based on general claims about human nature, the meaning of liberal discourse, context-independent efficiency, or the benevolent nature of the technology we happen to have stumbled across at the end of the twentieth century. They are instead based on, and depend on the continued accuracy of, a description of the economics of fabrication of computers and network connections, and a description of the dynamics of linking in a network of connected nodes. As such, my claim is not that the Internet inherently liberates. I do not claim that commons-based production of information, knowledge, and culture will win out by some irresistible progressive force. That is what

makes the study of the political economy of information, knowledge, and culture in the networked environment directly relevant to policy. The literature on network topology suggests that, as long as there are widely distributed capabilities to publish, link, and advise others about what to read and link to, networks enable intrinsic processes that allow substantial ordering of the information. The pattern of information flow in such a network is more resistant to the application of control or influence than was the mass-media model. But things can change. Google could become so powerful on the desktop, in the e-mail utility, and on the Web, that it will effectively become a supernode that will indeed raise the prospect of a reemergence of a mass-media model. Then the politics of search engines, as Lucas Introna and Helen Nissenbaum called it, become central. The zeal to curb peer-to-peer file sharing of movies and music could lead to a substantial redesign of computing equipment and networks, to a degree that would make it harder for end users to exchange information of their own making. Understanding what we will lose if such changes indeed warp the topology of the network, and through it the basic structure of the networked public sphere, is precisely the object of this book as a whole. For now, though, let us say that the networked information economy as it has developed to this date has a capacity to take in, filter, and synthesize observations and opinions from a population that is orders of magnitude larger than the population that was capable of being captured by the mass media. It has done so without re-creating identifiable and reliable points of control and manipulation that would replicate the core limitation of the mass-media model of the public sphere - its susceptibility to the exertion of control by its regulators, owners, or those who pay them.

Who Will Play the Watchdog Function?

A distinct critique leveled at the networked public sphere as a platform for democratic politics is the concern for who will fill the role of watchdog. Neil Netanel made this argument most clearly. His concern was that, perhaps freedom of expression for all is a good thing, and perhaps we could even overcome information overflow problems, but we live in a complex world with powerful actors. Government and corporate power is large, and individuals, no matter how good their tools, cannot be a serious alternative to a well-funded, independent press that can pay investigative reporters, defend lawsuits, and generally act like the *New York Times* and the *Washington Post* when they published the Pentagon Papers in the teeth of the Nixon administration's resistance, providing some of the most damning evidence against the planning and continued prosecution of the war in Vietnam. Netanel is cognizant of the tensions between the need to capture large audiences and sell advertising, on the one hand, and the role of watchdog, on the other. He nonetheless emphasizes that the networked public sphere cannot investigate as deeply or create the public salience that the mass media can. These limitations make commercial mass media, for all their limitations, necessary for a liberal public sphere.

This diagnosis of the potential of the networked public sphere underrepresents its productive capacity. The Diebold story provides in narrative form a detailed response to each of the concerns. The problem of voting machines has all the characteristics of an important, hard subject. It stirs deep fears that democracy is being stolen, and is therefore highly unsettling. It involves a difficult set of technical judgments about the functioning of voting machines. It required exposure and analysis of corporate-owned materials in the teeth of litigation threats and efforts to suppress and discredit the criticism. At each juncture in the process, the participants in the critique turned iteratively to peer production and radically distributed methods of investigation, analysis, distribution, and resistance to suppression: the initial observations of the whistle-blower or the hacker; the materials made available on a "see for yourself" and "come analyze this and share your insights" model; the distribution by students; and the fallback option when their server was shut down of replication around the network. At each stage, a peer-production solution was interposed

in place of where a well-funded, high-end mass-media outlet would have traditionally applied funding in expectation of sales of copy. And it was only after the networked public sphere developed the analysis and debate that the mass media caught on, and then only gingerly.

The Diebold case was not an aberration, but merely a particularly rich case study of a much broader phenomenon, most extensively described in Dan Gilmore's *We the Media*. The basic production modalities that typify the networked information economy are now being applied to the problem of producing politically relevant information. In 2005, the most visible example of application of the networked information economy - both in its peer-production dimension and more generally by combining a wide range of nonproprietary production models - to the watchdog function of the media is the political blogosphere. The founding myth of the blogosphere's journalistic potency was built on the back of then Senate majority leader Trent Lott. In 2002, Lott had the indiscretion of saying, at the one-hundredth-birthday party of Republican Senator Strom Thurmond, that if Thurmond had won his Dixiecrat presidential campaign, "we wouldn't have had all these problems over all these years." Thurmond had run on a segregationist campaign, splitting from the Democratic Party in opposition to Harry Truman's early civil rights efforts, as the post-World War II winds began blowing toward the eventual demise of formal, legal racial segregation in the United States. Few positions are taken to be more self-evident in the national public morality of early twenty-first-century America than that formal, state-imposed, racial discrimination is an abomination. And yet, the first few days after the birthday party at which Lott made his statement saw almost no reporting on the statement. ABC News and the *Washington Post* made small mention of it, but most media outlets reported merely on a congenial salute and farewell celebration of the Senate's oldest and longest-serving member. Things were different in the blogosphere. At first liberal blogs, and within three days conservative bloggers as well, began to excavate past racist statements by Lott, and to beat the drums calling for his censure or removal as Senate leader. Within about a week, the story surfaced in the mainstream media, became a major embarrassment, and led to Lott's resignation as Senate majority leader about a week later. A careful case study of this event leaves it unclear why the mainstream media initially ignored the story./32 It may have been that the largely social event drew the wrong sort of reporters. It may have been that reporters and editors who depend on major Washington, D.C., players were reluctant to challenge Lott. Perhaps they thought it rude to emphasize this indiscretion, or too upsetting to us all to think of just how close to the surface thoughts that we deem abominable can lurk. There is little disagreement that the day after the party, the story was picked up and discussed by Marshall on TalkingPoints, as well as by another liberal blogger, Atrios, who apparently got it from a post on Slate's "Chatterbox," which picked it up from ABC News's own *The Note*, a news summary made available on the television network's Web site. While the mass media largely ignored the story, and the two or three mainstream reporters who tried to write about it were getting little traction, bloggers were collecting more stories about prior instances where Lott's actions tended to suggest support for racist causes. Marshall, for example, found that Lott had filed a 1981 amicus curiae brief in support of Bob Jones University's effort to retain its tax-exempt status. The U.S. government had rescinded that status because the university practiced racial discrimination - such as prohibiting interracial dating. By Monday of the following week, four days after the remarks, conservative bloggers like Glenn Reynolds on Instapundit, Andrew Sullivan, and others were calling for Lott's resignation. It is possible that, absent the blogosphere, the story would still have flared up. There were two or so mainstream reporters still looking into the story. Jesse Jackson had come out within four days of the comment and said Lott should resign as majority leader. Eventually, when the mass media did enter the fray, its coverage clearly dominated the public agenda and its reporters uncovered materials that helped speed Lott's exit. However, given the short news cycle, the lack of initial interest by the media, and the large time lag between the event itself and when the media actually took the subject up, it seems likely that without the intervention of the blogosphere, the

story would have died. What happened instead is that the cluster of political blogs - starting on the Left but then moving across the Left-Right divide - took up the subject, investigated, wrote opinions, collected links and public interest, and eventually captured enough attention to make the comments a matter of public importance. Free from the need to appear neutral and not to offend readers, and free from the need to keep close working relationships with news subjects, bloggers were able to identify something that grated on their sensibilities, talk about it, dig deeper, and eventually generate a substantial intervention into the public sphere. That intervention still had to pass through the mass media, for we still live in a communications environment heavily based on those media. However, the new source of insight, debate, and eventual condensation of effective public opinion came from within the networked information environment.

The point is not to respond to the argument with a litany of anecdotes. The point is that the argument about the commercial media's role as watchdog turns out to be a familiar argument - it is the same argument that was made about software and supercomputers, encyclopedias and immersive entertainment scripts. The answer, too, is by now familiar. Just as the World Wide Web can offer a platform for the emergence of an enormous and effective almanac, just as free software can produce excellent software and peer production can produce a good encyclopedia, so too can peer production produce the public watchdog function. In doing so, clearly the unorganized collection of Internet users lacks some of the basic tools of the mass media: dedicated full-time reporters; contacts with politicians who need media to survive, and therefore cannot always afford to stonewall questions; or public visibility and credibility to back their assertions. However, network-based peer production also avoids the inherent conflicts between investigative reporting and the bottom line - its cost, its risk of litigation, its risk of withdrawal of advertising from alienated corporate subjects, and its risk of alienating readers. Building on the wide variation and diversity of knowledge, time, availability, insight, and experience, as well as the vast communications and information resources on hand for almost anyone in advanced economies, we are seeing that the watchdog function too is being peer produced in the networked information economy.

Note that while my focus in this chapter has been mostly the organization of public discourse, both the Sinclair and the Diebold case studies also identify characteristics of distributed political action. We see collective action emerging from the convergence of independent individual actions, with no hierarchical control like that of a political party or an organized campaign. There may be some coordination and condensation points - like BoycottSBG.com or blackboxvoting.org. Like other integration platforms in peer-production systems, these condensation points provide a critical function. They do not, however, control the process. One manifestation of distributed coordination for political action is something Howard Rheingold has called "smart mobs" - large collections of individuals who are able to coordinate real-world action through widely distributed information and communications technology. He tells of the "People Power II" revolution in Manila in 2001, where demonstrations to oust then president Estrada were coordinated spontaneously through extensive text messaging.[/33] Few images in the early twenty-first century can convey this phenomenon more vividly than the demonstrations around the world on February 15, 2003. Between six and ten million protesters were reported to have gone to the streets of major cities in about sixty countries in opposition to the American-led invasion of Iraq. There had been no major media campaign leading up to the demonstrations - though there was much media attention to them later. There had been no organizing committee. Instead, there was a network of roughly concordant actions, none controlling the other, all loosely discussing what ought to be done and when. MoveOn.org in the United States provides an example of a coordination platform for a network of politically mobilized activities. It builds on e-mail and Web-based media to communicate opportunities for political action to those likely to be willing and able to take it. Radically distributed, network-based solutions to the problems of political mobilization rely on the same characteristics as networked

information production more generally: extensive communications leading to concordant and cooperative patterns of behavior without the introduction of hierarchy or the interposition of payment.

Using Networked Communication to Work Around Authoritarian Control

The Internet and the networked public sphere offer a different set of potential benefits, and suffer a different set of threats, as a platform for liberation in authoritarian countries. State-controlled mass-media models are highly conducive to authoritarian control. Because they usually rely on a small number of technical and organizational points of control, mass media offer a relatively easy target for capture and control by governments. Successful control of such universally visible media then becomes an important tool of information manipulation, which, in turn, eases the problem of controlling the population. Not surprisingly, capture of the national television and radio stations is invariably an early target of coups and revolutions. The highly distributed networked architecture of the Internet makes it harder to control communications in this way.

The case of Radio B92 in Yugoslavia offers an example. B92 was founded in 1989, as an independent radio station. Over the course of the 1990s, it developed a significant independent newsroom broadcast over the station itself, and syndicated through thirty affiliated independent stations. B92 was banned twice after the NATO bombing of Belgrade, in an effort by the Milosevic regime to control information about the war. In each case, however, the station continued to produce programming, and distributed it over the Internet from a server based in Amsterdam. The point is a simple one. Shutting down a broadcast station is simple. There is one transmitter with one antenna, and police can find and hold it. It is much harder to shut down all connections from all reporters to a server and from the server back into the country wherever a computer exists.

This is not to say that the Internet will of necessity in the long term lead all authoritarian regimes to collapse. One option open to such regimes is simply to resist Internet use. In 2003, Burma, or Myanmar, had 28,000 Internet users out of a population of more than 42 million, or one in fifteen hundred, as compared, for example, to 6 million out of 65 million in neighboring Thailand, or roughly one in eleven. Most countries are not, however, willing to forgo the benefits of connectivity to maintain their control. Iran's population of 69 million includes 4.3 million Internet users, while China has about 80 million users, second only to the United States in absolute terms, out of a population of 1.3 billion. That is, both China and Iran have a density of Internet users of about one in sixteen.[/34] Burma's negligible level of Internet availability is a compound effect of low gross domestic product (GDP) per capita and government policies. Some countries with similar GDP levels still have levels of Internet users in the population that are two orders of magnitude higher: Cameroon (1 Internet user for every 27 residents), Moldova (1 in 30), and Mongolia (1 in 55). Even very large poor countries have several times more users per population than Myanmar: like Pakistan (1 in 100), Mauritania (1 in 300), and Bangladesh (1 in 580). Lawrence Solum and Minn Chung outline how Myanmar achieves its high degree of control and low degree of use.[/35] Myanmar has only one Internet service provider (ISP), owned by the government. The government must authorize anyone who wants to use the Internet or create a Web page within the country. Some of the licensees, like foreign businesses, are apparently permitted and enabled only to send e-mail, while using the Web is limited to security officials who monitor it. With this level of draconian regulation, Myanmar can avoid the liberating effects of the Internet altogether, at the cost of losing all its economic benefits. Few regimes are willing to pay that price.

Introducing Internet communications into a society does not, however, immediately and automatically mean that an open, liberal public sphere emerges. The Internet is technically harder to

control than mass media. It increases the cost and decreases the efficacy of information control. However, a regime willing and able to spend enough money and engineering power, and to limit its population's access to the Internet sufficiently, can have substantial success in controlling the flow of information into and out of its country. Solum and Chung describe in detail one of the most extensive and successful of these efforts, the one that has been conducted by China - home to the second-largest population of Internet users in the world, whose policies controlled use of the Internet by two out of every fifteen Internet users in the world in 2003. In China, the government holds a monopoly over all Internet connections going into and out of the country. It either provides or licenses the four national backbones that carry traffic throughout China and connect it to the global network. ISPs that hang off these backbones are licensed, and must provide information about the location and workings of their facilities, as well as comply with a code of conduct. Individual users must register and provide information about their machines, and the many Internet cafes are required to install filtering software that will filter out subversive sites. There have been crackdowns on Internet cafes to enforce these requirements. This set of regulations has replicated one aspect of the mass-medium model for the Internet - it has created a potential point of concentration or centralization of information flow that would make it easier to control Internet use. The highly distributed production capabilities of the networked information economy, however, as opposed merely to the distributed carriage capability of the Internet, mean that more must be done at this bottleneck to squelch the flow of information and opinion than would have to be done with mass media. That "more" in China has consisted of an effort to employ automatic filters - some at the level of the cybercafe or the local ISP, some at the level of the national backbone networks. The variability of these loci and their effects is reflected in partial efficacy and variable performance for these mechanisms. The most extensive study of the efficacy of these strategies for controlling information flows over the Internet to China was conducted by Jonathan Zittrain and Ben Edelman. From servers within China, they sampled about two hundred thousand Web sites and found that about fifty thousand were unavailable at least once, and close to nineteen thousand were unavailable on two distinct occasions. The blocking patterns seemed to follow mass-media logic - BBC News was consistently unavailable, as CNN and other major news sites often were; the U.S. court system official site was unavailable. However, Web sites that provided similar information - like those that offered access to all court cases but were outside the official system - were available. The core Web sites of human rights organizations or of Taiwan and Tibet-related organizations were blocked, and about sixty of the top one hundred results for "Tibet" on Google were blocked. What is also apparent from their study, however, and confirmed by Amnesty International's reports on Internet censorship in China, is that while censorship is significant, it is only partially effective.[/36](#) The Amnesty report noted that Chinese users were able to use a variety of techniques to avoid the filtering, such as the use of proxy servers, but even Zittrain and Edelman, apparently testing for filtering as experienced by unsophisticated or compliant Internet users in China, could access many sites that would, on their face, seem potentially destabilizing.

This level of censorship may indeed be effective enough for a government negotiating economic and trade expansion with political stability and control. It suggests, however, limits of the ability of even a highly dedicated government to control the capacity of Internet communications to route around censorship and to make it much easier for determined users to find information they care about, and to disseminate their own information to others. Iran's experience, with a similar level of Internet penetration, emphasizes the difficulty of maintaining control of Internet publication.[/37](#) Iran's network emerged from 1993 onward from the university system, quite rapidly complemented by commercial ISPs. Because deployment and use of the Internet preceded its regulation by the government, its architecture is less amenable to centralized filtering and control than China's. Internet access through university accounts and cybercafes appears to be substantial, and until the

past three or four years, had operated free of the crackdowns and prison terms suffered by opposition print publications and reporters. The conservative branches of the regime seem to have taken a greater interest in suppressing Internet communications since the publication of imprisoned Ayatollah Montazeri's critique of the foundations of the Islamic state on the Web in December 2000. While the original Web site, montazeri.com, seems to have been eliminated, the site persists as montazeri.ws, using a Western Samoan domain name, as do a number of other Iranian publications. There are now dozens of chat rooms, blogs, and Web sites, and e-mail also seems to be playing an increasing role in the education and organization of an opposition. While the conservative branches of the Iranian state have been clamping down on these forms, and some bloggers and Web site operators have found themselves subject to the same mistreatment as journalists, the efficacy of these efforts to shut down opposition seems to be limited and uneven.

Media other than static Web sites present substantially deeper problems for regimes like those of China and Iran. Scanning the text of e-mail messages of millions of users who can encrypt their communications with widely available tools creates a much more complex problem. Ephemeral media like chat rooms and writable Web tools allow the content of an Internet communication or Web site to be changed easily and dynamically, so that blocking sites becomes harder, while coordinating moves to new sites to route around blocking becomes easier. At one degree of complexity deeper, the widely distributed architecture of the Net also allows users to build censorship-resistant networks by pooling their own resources. The pioneering example of this approach is Freenet, initially developed in 1999-2000 by Ian Clarke, an Irish programmer fresh out of a degree in computer science and artificial intelligence at Edinburgh University. Now a broader free-software project, Freenet is a peer-to-peer application specifically designed to be censorship resistant. Unlike the more famous peer-to-peer network developed at the time - Napster - Freenet was not intended to store music files on the hard drives of users. Instead, it stores bits and pieces of publications, and then uses sophisticated algorithms to deliver the documents to whoever seeks them, in encrypted form. This design trades off easy availability for a series of security measures that prevent even the owners of the hard drives on which the data resides - or government agents that search their computers - from knowing what is on their hard drive or from controlling it. As a practical matter, if someone in a country that prohibits certain content but enables Internet connections wants to publish content - say, a Web site or blog - safely, they can inject it into the Freenet system. The content will be encrypted and divided into little bits and pieces that are stored in many different hard drives of participants in the network. No single computer will have all the information, and shutting down any given computer will not make the information unavailable. It will continue to be accessible to anyone running the Freenet client. Freenet indeed appears to be used in China, although the precise scope is hard to determine, as the network is intended to mask the identity and location of both readers and publishers in this system. The point to focus on is not the specifics of Freenet, but the feasibility of constructing user-based censorship-resistant storage and retrieval systems that would be practically impossible for a national censorship system to identify and block subversive content.

To conclude, in authoritarian countries, the introduction of Internet communications makes it harder and more costly for governments to control the public sphere. If these governments are willing to forgo the benefits of Internet connectivity, they can avoid this problem. If they are not, they find themselves with less control over the public sphere. There are, obviously, other means of more direct repression. However, control over the mass media was, throughout most of the twentieth century, a core tool of repressive governments. It allowed them to manipulate what the masses of their populations knew and believed, and thus limited the portion of the population that the government needed to physically repress to a small and often geographically localized group. The efficacy of these techniques of repression is blunted by adoption of the Internet and the

emergence of a networked information economy. Low-cost communications, distributed technical and organizational structure, and ubiquitous presence of dynamic authorship tools make control over the public sphere difficult, and practically never perfect.

Toward a Networked Public Sphere

The first generation of statements that the Internet democratizes was correct but imprecise. The Internet does restructure public discourse in ways that give individuals a greater say in their governance than the mass media made possible. The Internet does provide avenues of discourse around the bottlenecks of older media, whether these are held by authoritarian governments or by media owners. But the mechanisms for this change are more complex than those articulated in the past. And these more complex mechanisms respond to the basic critiques that have been raised against the notion that the Internet enhances democracy.

Part of what has changed with the Internet is technical infrastructure. Network communications do not offer themselves up as easily for single points of control as did the mass media. While it is possible for authoritarian regimes to try to retain bottlenecks in the Internet, the cost is higher and the efficacy lower than in mass-media-dominated systems. While this does not mean that introduction of the Internet will automatically result in global democratization, it does make the work of authoritarian regimes harder. In liberal democracies, the primary effect of the Internet runs through the emergence of the networked information economy. We are seeing the emergence to much greater significance of nonmarket, individual, and cooperative peer-production efforts to produce universal intake of observations and opinions about the state of the world and what might and ought to be done about it. We are seeing the emergence of filtering, accreditation, and synthesis mechanisms as part of network behavior. These rely on clustering of communities of interest and association and highlighting of certain sites, but offer tremendous redundancy of paths for expression and accreditation. These practices leave no single point of failure for discourse: no single point where observations can be squelched or attention commanded - by fiat or with the application of money. Because of these emerging systems, the networked information economy is solving the information overload and discourse fragmentation concerns without reintroducing the distortions of the mass-media model. Peer production, both long-term and organized, as in the case of Slashdot, and ad hoc and dynamically formed, as in the case of blogging or the Sinclair or Diebold cases, is providing some of the most important functionalities of the media. These efforts provide a watchdog, a source of salient observations regarding matters of public concern, and a platform for discussing the alternatives open to a polity.

In the networked information environment, everyone is free to observe, report, question, and debate, not only in principle, but in actual capability. They can do this, if not through their own widely read blog, then through a cycle of mailing lists, collective Web-based media like Slashdot, comments on blogs, or even merely through e-mails to friends who, in turn, have meaningful visibility in a smallish-scale cluster of sites or lists. We are witnessing a fundamental change in how individuals can interact with their democracy and experience their role as citizens. Ideal citizens need not be seen purely as trying to inform themselves about what others have found, so that they can vote intelligently. They need not be limited to reading the opinions of opinion makers and judging them in private conversations. They are no longer constrained to occupy the role of mere readers, viewers, and listeners. They can be, instead, participants in a conversation. Practices that begin to take advantage of these new capabilities shift the locus of content creation from the few professional journalists trolling society for issues and observations, to the people who make up that society. They begin to free the public agenda setting from dependence on the judgments of managers, whose job it is to assure that the maximum number of readers, viewers, and listeners are

sold in the market for eyeballs. The agenda thus can be rooted in the life and experience of individual participants in society - in their observations, experiences, and obsessions. The network allows all citizens to change their relationship to the public sphere. They no longer need be consumers and passive spectators. They can become creators and primary subjects. It is in this sense that the Internet democratizes.

Notes

1. *Reno v. ACLU*, 521 U.S. 844, 852-853, and 896-897 (1997).

2. Elizabeth Jensen, "Sinclair Fires Journalist After Critical Comments," *Los Angeles Times*, October 19, 2004.

3. Jensen, "Sinclair Fires Journalist"; Sheridan Lyons, "Fired Reporter Tells Why He Spoke Out," *Baltimore Sun*, October 29, 2004.

4. The various posts are archived and can be read, chronologically, at http://www.talkingpointsmemo.com/archives/week_2004_10_10.php .

5. Duane D. Stanford, *Atlanta Journal-Constitution*, October 31, 2002, 1A.

6. Katherine Q. Seelye, "The 2002 Campaign: The States; Georgia About to Plunge into Touch-Screen Voting," *New York Times*, October 30, 2002, A22.

7. Edward Walsh, "Election Day to Be Test of Voting Process," *Washington Post*, November 4, 2002, A1.

8. *Washington Post*, December 12, 2002.

9. *Online Policy Group v. Diebold, Inc.*, 337 F. Supp. 2d 1195 (2004).

10. California Secretary of State Voting Systems Panel, Meeting Minutes, November 3, 2003, http://www.ss.ca.gov/elections/vsp_min_110303.pdf .

11. Eli Noam, "Will the Internet Be Bad for Democracy?" (November 2001), http://www.citi.columbia.edu/elinoam/articles/int_bad_dem.htm .

12. Eli Noam, "The Internet Still Wide, Open, and Competitive?" Paper presented at The Telecommunications Policy Research Conference, September 2003, http://www.tprc.org/papers/2003/200/noam_TPRC2003.pdf .

13. Federal Communications Commission, Report on High Speed Services, December 2003.

14. See Eszter Hargittai, "The Changing Online Landscape: From Free-For-All to Commercial Gatekeeping," http://www.eszter.com/research/pubs/hargittai-onlinelandscape.pdf .

15. Derek de Solla Price, "Networks of Scientific Papers," *Science* 149 (1965): 510; Herbert Simon, "On a Class of Skew Distribution Function," *Biometrica* 42 (1955): 425-440, reprinted in Herbert

Simon, *Models of Man Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting* (New York: Garland, 1957).

16. Albert-Lászio Barabási and Reka Albert, "Emergence of Scaling in Random Networks," *Science* 286 (1999): 509.

17. Bernardo Huberman and Lada Adamic, "Growth Dynamics of the World Wide Web," *Nature* 401 (1999): 131.

18. Albert-Lászio Barabási, *Linked, How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life* (New York: Penguin, 2003), 56-57. One unpublished quantitative study showed specifically that the skewness holds for political Web sites related to various hot-button political issues in the United States-like abortion, gun control, or the death penalty. A small fraction of the Web sites discussing these issues account for the large majority of links into them. Matthew Hindman, Kostas Tsioutsiouliklis, and Judy Johnson, " 'Googlearchy': How a Few Heavily Linked Sites Dominate Politics on the Web," July 28, 2003, http://www.princeton.edu/~mhindman/googlearchy--hindman.pdf .

19. Lada Adamic and Bernardo Huberman, "Power Law Distribution of the World Wide Web," *Science* 287 (2000): 2115.

20. Ravi Kumar et al., "Trawling the Web for Emerging Cyber-Communities," *WWW8/Computer Networks* 31, nos. 11-16 (1999): 1481-1493.

21. Gary W. Flake et al., "Self-Organization and Identification of Web Communities," *IEEE Computer* 35, no. 3 (2002): 66-71. Another paper that showed significant internal citations within topics was Soumen Chakrabati et al., "The Structure of Broad Topics on the Web," WWW2002, Honolulu, HI, May 7-11, 2002.

22. Lada Adamic and Natalie Glance, "The Political Blogosphere and the 2004 Election: Divided They Blog," March 1, 2005, http://www.blogpulse.com/papers/2005/AdamicGlanceBlogWWW.pdf .

23. M.E.J. Newman, "The Structure and Function of Complex Networks," *Society for Industrial and Applied Mathematics Review* 45, section 4.2.2 (2003): 167-256; S. N. Dorogovstev and J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford: Oxford University Press, 2003).

24. This structure was first described by Andrei Broder et al., "Graph Structure of the Web," paper presented at www9 conference (1999), http://www.almaden.ibm.com/webfountain/resources/GraphStructureintheWeb.pdf . It has since been further studied, refined, and substantiated in various studies.

25. Dill et al., "Self-Similarity in the Web" (San Jose, CA: IBM Almaden Research Center, 2001); S. N. Dorogovstev and J.F.F. Mendes, *Evolution of Networks*.

26. Soumen Chakrabarti et al., "The Structure of Broad Topics on the Web," WWW2002, Honolulu, HI, May 7-11, 2002.

27. Daniel W. Drezner and Henry Farrell, "The Power and Politics of Blogs" (July 2004), http://www.danieldrezner.com/research/blogpaperfinal.pdf .

28. D. J. Watts and S. H. Strogatz, "Collective Dynamics of 'Small World' Networks," *Nature* 393 (1998): 440-442; D. J. Watts, *Small Worlds: The Dynamics of Networks Between Order and Randomness* (Princeton, NJ: Princeton University Press, 1999).

29. Clay Shirky, "Power Law, Weblogs, and Inequality" (February 8, 2003), http://www.shirky.com/writings/powerlaw_weblog.htm ; Jason Kottke, "Weblogs and Power Laws" (February 9, 2003), http://www.kottke.org/03/02/weblogs-and-power-laws .

30. Ravi Kumar et al., "On the Bursty Evolution of Blogspace," Proceedings of WWW2003, May 20-24, 2003, http://www2003.org/cdrom/papers/refereed/p477/p477-kumar/p477-kumar.htm .

31. Both of these findings are consistent with even more recent work by Hargittai, E., J. Gallo and S. Zehnder, "Mapping the Political Blogosphere: An Analysis of Large-Scale Online Political Discussions," 2005. Poster presented at the International Communication Association meetings, New York.

32. Harvard Kennedy School of Government, Case Program: " 'Big Media' Meets 'Bloggers': Coverage of Trent Lott's Remarks at Strom Thurmond's Birthday Party," http://www.ksg.harvard.edu/presspol/Research_Publications/Case_Studies/1731_0.pdf .

33. Howard Rheingold, *Smart Mobs, The Next Social Revolution* (Cambridge, MA: Perseus Publishing, 2002).

34. Data taken from *CIA World Fact Book* (Washington, DC: Central Intelligence Agency, 2004).

35. Lawrence Solum and Minn Chung, "The Layers Principle: Internet Architecture and the Law" (working paper no. 55, University of San Diego School of Law, Public Law and Legal Theory, June 2003).

36. Amnesty International, People's Republic of China, State Control of the Internet in China (2002).

37. A synthesis of news-based accounts is Babak Rahimi, "Cyberdissent: The Internet in Revolutionary Iran," *Middle East Review of International Affairs* 7, no. 3 (2003).

Chapter 8
Cultural Freedom: A Culture Both Plastic and Critical

Gone with the Wind

There was a land of Cavaliers and Cotton Fields called the Old South. Here in this pretty world, Gallantry took its last bow. Here was the last ever to be seen of Knights and their Ladies Fair, of Master and of Slave. Look for it only in books, for it is no more than a dream remembered, a Civilization gone with the wind.

-MGM (1939) film adaptation of Margaret Mitchell's novel (1936)

Strange Fruit

Southern trees bear strange fruit,
Blood on the leaves and blood at the root,
Black bodies swinging in the southern breeze,
Strange fruit hanging from the poplar trees.

Pastoral scene of the gallant south,
The bulging eyes and the twisted mouth,
Scent of magnolias, sweet and fresh,
Then the sudden smell of burning flesh.

Here is the fruit for the crows to pluck,
For the rain to gather, for the wind to suck,
For the sun to rot, for the trees to drop,
Here is a strange and bitter crop.

-Billie Holiday (1939)
from lyrics by Abel Meeropol (1937)

In 1939, *Gone with the Wind* reaped seven Oscars, while Billie Holiday's song reached number 16 on the charts, even though Columbia Records refused to release it: Holiday had to record it with a small company that was run out of a storefront in midtown Manhattan. On the eve of the second reconstruction era, which was to overhaul the legal framework of race relations over the two decades beginning with the desegregation of the armed forces in the late 1940s and culminating with the civil rights acts passed between 1964-1968, the two sides of the debate over desegregation and the legacy of slavery were minting new icons through which to express their most basic beliefs about the South and its peculiar institutions. As the following three decades unfolded and the South was gradually forced to change its ways, the cultural domain continued to work out the meaning of race relations in the United States and the history of slavery. The actual slogging of regulation of discrimination, implementation of desegregation and later affirmative action, and the more local politics of hiring and firing were punctuated throughout this period by salient iconic retellings of the stories of race relations in the United States, from *Guess Who's Coming to Dinner?* to *Roots*. The point of this chapter, however, is not to discuss race relations, but to understand culture and cultural production in terms of political theory. *Gone with the Wind* and *Strange Fruit* or *Guess Who's Coming to Dinner?* offer us intuitively accessible instances of a much broader and more basic

characteristic of human understanding and social relations. Culture, shared meaning, and symbols are how we construct our views of life across a wide range of domains - personal, political, and social. How culture is produced is therefore an essential ingredient in structuring how freedom and justice are perceived, conceived, and pursued. In the twentieth century, Hollywood and the recording industry came to play a very large role in this domain. The networked information economy now seems poised to attenuate that role in favor of a more participatory and transparent cultural production system.

Cultural freedom occupies a position that relates to both political freedom and individual autonomy, but is synonymous with neither. The root of its importance is that none of us exist outside of culture. As individuals and as political actors, we understand the world we occupy, evaluate it, and act in it from within a set of understandings and frames of meaning and reference that we share with others. What institutions and decisions are considered "legitimate" and worthy of compliance or participation; what courses of action are attractive; what forms of interaction with others are considered appropriate - these are all understandings negotiated from within a set of shared frames of meaning. How those frames of meaning are shaped and by whom become central components of the structure of freedom for those individuals and societies that inhabit it and are inhabited by it. They define the public sphere in a much broader sense than we considered in the prior chapters.

The networked information economy makes it possible to reshape both the "who" and the "how" of cultural production relative to cultural production in the twentieth century. It adds to the centralized, market-oriented production system a new framework of radically decentralized individual and cooperative nonmarket production. It thereby affects the ability of individuals and groups to participate in the production of the cultural tools and frameworks of human understanding and discourse. It affects the way we, as individuals and members of social and political clusters, interact with culture, and through it with each other. It makes culture more transparent to its inhabitants. It makes the process of cultural production more participatory, in the sense that more of those who live within a culture can actively participate in its creation. We are seeing the possibility of an emergence of a new popular culture, produced on the folk-culture model and inhabited actively, rather than passively consumed by the masses. Through these twin characteristics - transparency and participation - the networked information economy also creates greater space for critical evaluation of cultural materials and tools. The practice of producing culture makes us all more sophisticated readers, viewers, and listeners, as well as more engaged makers.

Throughout the twentieth century, the making of widely shared images and symbols was a concentrated practice that went through the filters of Hollywood and the recording industry. The radically declining costs of manipulating video and still images, audio, and text have, however, made culturally embedded criticism and broad participation in the making of meaning much more feasible than in the past. Anyone with a personal computer can cut and mix files, make their own files, and publish them to a global audience. This is not to say that cultural bricolage, playfulness, and criticism did not exist before. One can go to the avant-garde movement, but equally well to African-Brazilian culture or to Our Lady of Guadalupe to find them. Even with regard to television, that most passive of electronic media, John Fiske argued under the rubric of "semiotic democracy" that viewers engage in creative play and meaning making around the TV shows they watch. However, the technical characteristics of digital information technology, the economics of networked information production, and the social practices of networked discourse qualitatively change the role individuals can play in cultural production.

The practical capacity individuals and noncommercial actors have to use and manipulate cultural artifacts today, playfully or critically, far outstrips anything possible in television, film, or recorded music, as these were organized throughout the twentieth century. The diversity of cultural moves and statements that results from these new opportunities for creativity vastly increases the range of cultural elements accessible to any individual. Our ability, therefore, to navigate the cultural environment and make it our own, both through creation and through active selection and attention, has increased to the point of making a qualitative difference. In the academic law literature, Niva Elkin Koren wrote early about the potential democratization of "meaning making processes," William Fisher about "semiotic democracy," and Jack Balkin about a "democratic culture." Lessig has explored the generative capacity of the freedom to create culture, its contribution to creativity itself. These efforts revolve around the idea that there is something normatively attractive, from the perspective of "democracy" as a liberal value, about the fact that anyone, using widely available equipment, can take from the existing cultural universe more or less whatever they want, cut it, paste it, mix it, and make it their own - equally well expressing their adoration as their disgust, their embrace of certain images as their rejection of them.

Building on this work, this chapter seeks to do three things: First, I claim that the modalities of cultural production and exchange are a proper subject for normative evaluation within a broad range of liberal political theory. Culture is a social-psychological-cognitive fact of human existence. Ignoring it, as rights-based and utilitarian versions of liberalism tend to do, disables political theory from commenting on central characteristics of a society and its institutional frameworks. Analyzing the attractiveness of any given political institutional system without considering how it affects cultural production, and through it the production of the basic frames of meaning through which individual and collective self-determination functions, leaves a large hole in our analysis. Liberal political theory needs a theory of culture and agency that is viscous enough to matter normatively, but loose enough to give its core foci - the individual and the political system - room to be effective independently, not as a mere expression or extension of culture. Second, I argue that cultural production in the form of the networked information economy offers individuals a greater participatory role in making the culture they occupy, and makes this culture more transparent to its inhabitants. This descriptive part occupies much of the chapter. Third, I suggest the relatively straightforward conclusion of the prior two observations. From the perspective of liberal political theory, the kind of open, participatory, transparent folk culture that is emerging in the networked environment is normatively more attractive than was the industrial cultural production system typified by Hollywood and the recording industry.

A nine-year-old girl searching Google for Barbie will quite quickly find links to AdiosBarbie.com, to the Barbie Liberation Organization (BLO), and to other, similarly critical sites interspersed among those dedicated to selling and playing with the doll. The contested nature of the doll becomes publicly and everywhere apparent, liberated from the confines of feminist-criticism symposia and undergraduate courses. This simple Web search represents both of the core contributions of the networked information economy. First, from the perspective of the searching girl, it represents a new transparency of cultural symbols. Second, from the perspective of the participants in AdiosBarbie or the BLO, the girl's use of their site completes their own quest to participate in making the cultural meaning of Barbie. The networked information environment provides an outlet for contrary expression and a medium for shaking what we accept as cultural baseline assumptions. Its radically decentralized production modes provide greater freedom to participate effectively in defining the cultural symbols of our day. These characteristics make the networked environment attractive from the perspectives of both personal freedom of expression and an engaged and self-aware political discourse.

We cannot, however, take for granted that the technological capacity to participate in the cultural conversation, to mix and make our own, will translate into the freedom to do so. The practices of cultural and countercultural creation are at the very core of the battle over the institutional ecology of the digital environment. The tension is perhaps not new or unique to the Internet, but its salience is now greater. The makers of the 1970s comic strip *Air Pirates* already found their comics confiscated when they portrayed Mickey and Minnie and Donald and Daisy in various compromising countercultural postures. Now, the ever-increasing scope and expanse of copyright law and associated regulatory mechanisms, on the one hand, and of individual and collective nonmarket creativity, on the other hand, have heightened the conflict between cultural freedom and the regulatory framework on which the industrial cultural production system depends. As Lessig, Jessica Litman, and Siva Vaidhyanathan have each portrayed elegantly and in detail, the copyright industries have on many dimensions persuaded both Congress and courts that individual, nonmarket creativity using the cultural outputs of the industrial information economy is to be prohibited. As we stand today, freedom to play with the cultural environment is nonetheless preserved in the teeth of the legal constraints, because of the high costs of enforcement, on the one hand, and the ubiquity and low cost of the means to engage in creative cultural bricolage, on the other hand. These social, institutional, and technical facts still leave us with quite a bit of unauthorized creative expression. These facts, however, are contingent and fragile. Chapter 11 outlines in some detail the long trend toward the creation of ever-stronger legal regulation of cultural production, and in particular, the enclosure movement that began in the 1970s and gained steam in the mid-1990s. A series of seemingly discrete regulatory moves threatens the emerging networked folk culture. Ranging from judicial interpretations of copyright law to efforts to regulate the hardware and software of the networked environment, we are seeing a series of efforts to restrict nonmarket use of twentieth-century cultural materials in order to preserve the business models of Hollywood and the recording industry. These regulatory efforts threaten the freedom to participate in twenty-first-century cultural production, because current creation requires taking and mixing the twentieth-century cultural materials that make up who we are as culturally embedded beings. Here, however, I focus on explaining how cultural participation maps onto the project of liberal political theory, and why the emerging cultural practices should be seen as attractive within that normative framework. I leave development of the policy implications to part III.

Cultural Freedom in Liberal Political Theory

Utilitarian and rights-based liberal political theories have an awkward relationship to culture. Both major strains of liberal theory make a certain set of assumptions about the autonomous individuals with which they are concerned. Individuals are assumed to be rational and knowledgeable, at least about what is good for them. They are conceived of as possessing a capacity for reason and a set of preferences prior to engagement with others. Political theory then proceeds to concern itself with political structures that respect the autonomy of individuals with such characteristics. In the political domain, this conception of the individual is easiest to see in pluralist theories, which require institutions for collective decision making that clear what are treated as already-formed preferences of individuals or voluntary groupings.

Culture represents a mysterious category for these types of liberal political theories. It is difficult to specify how it functions in terms readily amenable to a conception of individuals whose rationality and preferences for their own good are treated as though they preexist and are independent of society. A concept of culture requires some commonly held meaning among these individuals. Even the simplest intuitive conception of what culture might mean would treat this common frame of meaning as the result of social processes that preexist any individual, and partially structure what it is that individuals bring to the table as they negotiate their lives together, in society or in a polity.

Inhabiting a culture is a precondition to any interpretation of what is at stake in any communicative exchange among individuals. A partly subconscious, lifelong dynamic social process of becoming and changing as a cultural being is difficult to fold into a collective decision-making model that focuses on designing a discursive platform for individuated discrete participants who are the bearers of political will. It is easier to model respect for an individual's will when one adopts a view of that will as independent, stable, and purely internally generated. It is harder to do so when one conceives of that individual will as already in some unspecified degree rooted in exchange with others about what an individual is to value and prefer.

Culture has, of course, been incorporated into political theory as a central part of the critique of liberalism. The politics of culture have been a staple of critical theory since Marx first wrote that "Religion . . . is the opium of the people" and that "to call on them to give up their illusions about their condition is to call on them to give up a condition that requires illusions."[/1] The twentieth century saw a wide array of critique, from cultural Marxism to poststructuralism and postmodernism. However, much of mainstream liberal political theory has chosen to ignore, rather than respond and adapt to, these critiques. In *Political Liberalism*, for example, Rawls acknowledges "the fact" of reasonable pluralism - of groups that persistently and reasonably hold competing comprehensive doctrines - and aims for political pluralism as a mode of managing the irreconcilable differences. This leaves the formation of the comprehensive doctrine and the systems of belief within which it is rendered "reasonable" a black box to liberal theory. This may be an adequate strategy for analyzing the structure of formal political institutions at the broadest level of abstraction. However, it disables liberal political theory from dealing with more fine-grained questions of policy that act within the black box.

As a practical matter, treating culture as a black box disables a political theory as a mechanism for diagnosing the actual conditions of life in a society in terms of its own political values. It does so in precisely the same way that a formal conception of autonomy disables those who hold it from diagnosing the conditions of autonomy in practical life. Imagine for a moment that we had received a revelation that a crude version of Antonio Gramsci's hegemony theory was perfectly correct as a matter of descriptive sociology. Ruling classes do, in fact, consciously and successfully manipulate the culture in order to make the oppressed classes compliant. It would be difficult, then, to continue to justify holding a position about political institutions, or autonomy, that treated the question of how culture, generally, or even the narrow subset of reasonably held comprehensive doctrines like religion, are made, as a black box. It would be difficult to defend respect for autonomous choices as respect for an individual's will, if an objective observer could point to a social process, external to the individual and acting upon him or her, as the cause of the individual holding that will. It would be difficult to focus one's political design imperatives on public processes that allow people to express their beliefs and preferences, argue about them, and ultimately vote on them, if it is descriptively correct that those beliefs and preferences are themselves the product of manipulation of some groups by others.

The point is not, of course, that Gramsci was descriptively right or that any of the broad range of critical theories of culture is correct as a descriptive matter. It is that liberal theories that ignore culture are rendered incapable of answering some questions that arise in the real world and have real implications for individuals and polities. There is a range of sociological, psychological, or linguistic descriptions that could characterize the culture of a society as more or less in accord with the concern of liberalism with individual and collective self-determination. Some such descriptive theory of culture can provide us with enough purchase on the role of culture to diagnose the attractiveness of a cultural production system from a political-theory perspective. It does not require that liberal theory abandon individuals as the bearers of the claims of political morality. It

does not require that liberal political theory refocus on culture as opposed to formal political institutions. It does require, however, that liberal theory at least be able to diagnose different conditions in the practical cultural life of a society as more or less attractive from the perspective of liberal political theory.

The efforts of deliberative liberal theories to account for culture offer the most obvious source of such an insight. These political theories have worked to develop a conception of culture and its relationship to liberalism precisely because at a minimum, they require mutual intelligibility across individuals, which cannot adequately be explained without some conception of culture. In Jurgen Habermas's work, culture plays the role of a basis for mutual intelligibility. As the basis for "interpersonal intelligibility," we see culture playing such a role in the work of Bruce Ackerman, who speaks of acculturation as the necessary condition to liberal dialogue. "Cultural coherence" is something he sees children requiring as a precondition to becoming liberal citizens: it allows them to "Talk" and defend their claims in terms without which there can be no liberal conversation.[/2] Michael Walzer argues that, "in matters of morality, argument is simply the appeal to common meanings."[/3] Will Kymlicka claims that for individual autonomy, "freedom involves making choices amongst various options, and our societal culture not only provides these options, but makes them meaningful to us." A societal culture, in turn, is a "shared vocabulary of tradition and convention" that is "embodied in social life[,] institutionally embodied - in schools, media, economy, government, etc."[/4] Common meanings in all these frameworks must mean more than simple comprehension of the words of another. It provides a common baseline, which is not itself at that moment the subject of conversation or inquiry, but forms the background on which conversation and inquiry take place. Habermas's definition of lifeworld as "background knowledge," for example, is a crisp rendering of culture in this role:

the lifeworld embraces us as an unmediated certainty, out of whose immediate proximity we live and speak. This all-penetrating, yet latent and unnoticed presence of the background of communicative action can be described as a more intense, yet deficient, form of knowledge and ability. To begin with, we make use of this knowledge involuntarily, without reflectively knowing *that* we possess it at all. What enables background knowledge to acquire absolute certainty in this way, and even augments its epistemic quality from a subjective standpoint, is precisely the property that robs it of a constitutive feature of knowledge: we make use of such knowledge without the awareness that it *could* be false. Insofar as all knowledge is fallible and is known to be such, background knowledge does not represent knowledge at all, in a strict sense. As background knowledge, it lacks the possibility of being challenged, that is, of being raised to the level of criticizable validity claims. One can do this only by converting it from a resource into a topic of discussion, at which point - just when it is thematized - it no longer functions as a lifeworld background but rather *disintegrates* in its background modality.[/5]

In other words, our understanding of meaning - how we are, how others are, what ought to be - are in some significant portion unexamined assumptions that we share with others, and to which we appeal as we engage in communication with them. This does not mean that culture is a version of false consciousness. It does not mean that background knowledge cannot be examined rationally or otherwise undermines the very possibility or coherence of a liberal individual or polity. It does mean, however, that at any given time, in any given context, there will be some set of historically contingent beliefs, attitudes, and social and psychological conditions that will in the normal course remain unexamined, and form the unexamined foundation of conversation. Culture is revisable through critical examination, at which point it ceases to be "common knowledge" and becomes a contested assumption. Nevertheless, some body of unexamined common knowledge is necessary

for us to have an intelligible conversation that does not constantly go around in circles, challenging the assumptions on whichever conversational move is made.

Culture, in this framework, is not destiny. It does not predetermine who we are, or what we can become or do, nor is it a fixed artifact. It is the product of a dynamic process of engagement among those who make up a culture. It is a frame of meaning from within which we must inevitably function and speak to each other, and whose terms, constraints, and affordances we always negotiate. There is no point outside of culture from which to do otherwise. An old Yiddish folktale tells of a naïve rabbi who, for safekeeping, put a ten-ruble note inside his copy of the Torah, at the page of the commandment, "thou shalt not steal." That same night, a thief stole into the rabbi's home, took the ten-ruble note, and left a five-ruble note in its place, at the page of the commandment, "thou shalt love thy neighbor as thyself." The rabbi and the thief share a common cultural framework (as do we, across the cultural divide), through which their various actions can be understood; indeed, without which their actions would be unintelligible. The story offers a theory of culture, power, and freedom that is more congenial to liberal political theory than critical theories, and yet provides a conception of the role of culture in human relations that provides enough friction, or viscosity, to allow meaning making in culture to play a role in the core concerns of liberal political theory. Their actions are part strategic and part communicative - that is to say, to some extent they seek to force an outcome, and to some extent they seek to engage the other in a conversation in order to achieve a commonly accepted outcome. The rabbi places the ten-ruble note in the Bible in order to impress upon the putative thief that he should leave the money where it is. He cannot exert force on the thief by locking the money up in a safe because he does not own one. Instead, he calls upon a shared understanding and a claim of authority within the governed society to persuade the thief. The thief, to the contrary, could have physically taken the ten-ruble note without replacing it, but he does not. He engages the rabbi in the same conversation. In part, he justifies his claim to five rubles. In part, he resists the authority of the rabbi - not by rejecting the culture that renders the rabbi a privileged expert, but by playing the game of Talmudic disputation. There is a price, though, for participating in the conversation. The thief must leave the five-ruble note; he cannot take the whole amount.

In this story, culture is open to interpretation and manipulation, but not infinitely so. Some moves may be valid within a cultural framework and alter it; others simply will not. The practical force of culture, on the other hand, is not brute force. It cannot force an outcome, but it can exert a real pull on the range of behaviors that people will seriously consider undertaking, both as individuals and as polities. The storyteller relies on the listener's cultural understanding about the limits of argument, or communicative action. The story exploits the open texture of culture, and the listener's shared cultural belief that stealing is an act of force, not a claim of justice; that those who engage in it do not conceive of themselves as engaged in legitimate defensible acts. The rabbi was naïve to begin with, but the thief's disputation is inconsistent with our sense of the nature of the act of stealing in exactly the same way that the rabbi's was, but inversely. The thief, the rabbi, and the storyteller participate in making, and altering, the meaning of the commandments.

Culture changes through the actions of individuals in the cultural context. Beliefs, claims, communicative moves that have one meaning before an intervention may begin to shift in their meaning as a result of other moves, made by other participants in the same cultural milieu. One need not adopt any given fully fledged meme theory of culture - like Richard Dawkins's, or Balkin's political adaptation of it as a theory of ideology - to accept that culture is created through communication among human beings, that it exerts some force on what they can say to each other and how it will be received, and that the parameters of a culture as a platform for making meaning in interaction among human beings change over time with use. How cultural moves are made, by

whom, and with what degree of perfect replication or subtle (and not so subtle) change, become important elements in determining the rate and direction of cultural change. These changes, over time, alter the platform individuals must use to make sense of the world they occupy, and for participants in conversation to be able to make intelligible communications to each other about the world they share and where it can and ought to go. Culture so understood is a social fact about particular sets of human beings in historical context. As a social fact, it constrains and facilitates the development, expression, and questioning of beliefs and positions. Whether and how Darwinism should be taught in public schools, for example, is a live political question in vast regions of the United States, and is played out as a debate over whether evolution is "merely a theory." Whether racial segregation should be practiced in these schools is no longer a viable or even conceivable political agenda. The difference between Darwinism and the undesirability of racial segregation is not that one is scientifically true and the other is not. The difference is that the former is not part of the "common knowledge" of a large section of society, whereas the latter is, in a way that no longer requires proof by detailed sociological and psychological studies of the type cited by the Supreme Court in support of its holding, in *Brown v. Board of Education*, that segregation in education was inherently unequal.

If culture is indeed part of how we form a shared sense of unexamined common knowledge, it plays a significant role in framing the meaning of the state of the world, the availability and desirability of choices, and the organization of discourse. The question of how culture is framed (and through it, meaning and the baseline conversational moves) then becomes germane to a liberal political theory. Between the Scylla of a fixed culture (with hierarchical, concentrated power to control its development and interpretation) and the Charybdis of a perfectly open culture (where nothing is fixed and everything is up for grabs, offering no anchor for meaning and mutual intelligibility), there is a wide range of practical social and economic arrangements around the production and use of culture. In evaluating the attractiveness of various arrangements from the perspective of liberal theory, we come to an already familiar trade-off, and an already familiar answer. As in the case of autonomy and political discourse, a greater ability of individuals to participate in the creation of the cultural meaning of the world they occupy is attractive from the perspective of the liberal commitments to individual freedom and democratic participation. As in both areas that we have already considered, a Babel objection appears: Too much freedom to challenge and remake our own cultural environment will lead to a lack of shared meaning. As in those two cases, however, the fears of too active a community of meaning making are likely exaggerated. Loosening the dominant power of Hollywood and television over contemporary culture is likely to represent an incremental improvement, from the perspective of liberal political commitments. It will lead to a greater transparency of culture, and therefore a greater capacity for critical reflection, and it will provide more opportunities for participating in the creation of culture, for interpolating individual glosses on it, and for creating shared variations on common themes.

The Transparency of Internet Culture

If you run a search for "Barbie" on three separate search engines - Google, Overture, and Yahoo! - you will get quite different results. Table 8.1 lists these results in the order in which they appear on each search engine. Overture is a search engine that sells placement to the parties who are being searched. Hits on this search engine are therefore ranked based on whoever paid Overture the most in order to be placed highly in response to a query. On this list, none of the top ten results represent anything other than sales-related Barbie sites. Critical sites begin to appear only around the twenty-fifth result, presumably after all paying clients have been served. Google, as we already know, uses a radically decentralized mechanism for assigning relevance. It counts how many sites on the Web have linked to a particular site that has the search term in it, and ranks the search results by

placing a site with a high number of incoming links above a site with a low number of incoming links. In effect, each Web site publisher "votes" for a site's relevance by linking to it, and Google aggregates these votes and renders them on their results page as higher ranking.

Table 8.1: Results for "Barbie" - Google versus Overture and Yahoo!

Google	*Overture*	*Yahoo!*
Barbie.com (Mattel's site)	Barbie at Amazon.com	Barbie.com
Barbie Collector: Official Mattel Web site for hobbyists and collectors	Toys and Leisure at QVC-Barbie	*Barbie Bazaar* Magazine
AdiosBarbie.com: A Body Image for Every Body (site created by women critical of Barbie's projected body image)	Barbie on Sale at KBToys	Barbie Collector
Barbie Bazaar Magazine (Barbie collectible news and Information)	Target.com: Barbies	My Scene.com
If You Were a Barbie, Which Messed Up Version Would You Be?	Barbie: Best prices and selection (bizrate.com)	EverythingGirl.com
Visible Barbie Project (macabre images of Barbie sliced as though in a science project)	Barbies, New and Preowned at NetDoll	Barbie History (fan-type history, mostly when various dolls were released)
Barbie: The Image of Us All (1995 undergraduate paper about Barbie's cultural history)	Barbies - compare prices (nextag.com)	Mattel, Inc.
Andigraph.free.fr (Barbie and Ken sex animation)	Barbie Toys (complete line of Barbie electronics online)	Spatula Jackson's Barbies (pictures of Barbie as various countercultural images).
Suicide bomber Barbie (Barbie with explosives strapped to waist)	Barbie Party supplies	Barbie! (fan site)
Barbies (Barbie dressed and painted as countercultural images)	Barbie and her accessories online	The Distorted Barbie

The little girl who searches for Barbie on Google will encounter a culturally contested figure. The same girl, searching on Overture, will encounter a commodity toy. In each case, the underlying efforts of Mattel, the producer of Barbie, have not changed. What is different is that in an environment where relevance is measured in nonmarket action - placing a link to a Web site

because you deem it relevant to whatever you are doing with your Web site - as opposed to in dollars, Barbie has become a more transparent cultural object. It is easier for the little girl to see that the doll is not only a toy, not only a symbol of beauty and glamour, but also a symbol of how norms of female beauty in our society can be oppressive to women and girls. The transparency does not force the girl to choose one meaning of Barbie or another. It does, however, render transparent that Barbie can have multiple meanings and that choosing meanings is a matter of political concern for some set of people who coinhabit this culture. Yahoo! occupies something of a middle ground - its algorithm does link to two of the critical sites among the top ten, and within the top twenty, identifies most of the sites that appear on Google's top ten that are not related to sales or promotion.

A similar phenomenon repeats itself in the context of explicit efforts to define Barbie-encyclopedias. There are, as of this writing, six general-interest online encyclopedias that are reasonably accessible on the Internet - that is to say, can be found with reasonable ease by looking at major search engines, sites that focus on education and parenting, and similar techniques. Five are commercial, and one is a quintessential commons-based peer-production project - *Wikipedia*. Of the five commercial encyclopedias, only one is available at no charge, the *Columbia Encyclopedia*, which is packaged in two primary forms - as encyclopedia.com and as part of Bartleby.com./6 The other four - *Britannica*, Microsoft's *Encarta*, the *World Book*, and *Grolier's Online Encyclopedia* - charge various subscription rates that range around fifty to sixty dollars a year. The *Columbia Encyclopedia* includes no reference to Barbie, the doll. The *World Book* has no "Barbie" entry, but does include a reference to Barbie as part of a fairly substantial article on "Dolls." The only information that is given is that the doll was introduced in 1959, that she has a large wardrobe, and in a different place, that dark-skinned Barbies were introduced in the 1980s. The article concludes with a guide of about three hundred words to good doll-collecting practices. Microsoft's *Encarta* also includes Barbie in the article on "Doll," but provides a brief separate definition as well, which replicates the *World Book* information in slightly different form: 1959, large wardrobe, and introduction of dark-skinned Barbies. The online photograph available with the definition is of a brown-skinned, black-haired Barbie. *Grolier's Online*'s major general-purpose encyclopedia, *Americana*, also has no entry for Barbie, but makes reference to the doll as part of the article on dolls. Barbie is described as a revolutionary new doll, made to resemble a teenage fashion model as part of a trend to realism in dolls. *Grolier's Online* does, however, include a more specialized *American Studies* encyclopedia that has an article on Barbie. That article heavily emphasizes the number of dolls sold and their value, provides some description of the chronological history of the doll, and makes opaque references to Barbie's physique and her emphasis on consumption. While the encyclopedia includes bibliographic references to critical works about Barbie, the textual references to cultural critique or problems she raises are very slight and quite oblique.

Only two encyclopedias focus explicitly on Barbie's cultural meaning: *Britannica* and *Wikipedia*. The *Britannica* entry was written by M. G. Lord, a professional journalist who authored a book entitled *Forever Barbie: The Unauthorized Biography of a Real Doll*. It is a tightly written piece that underscores the critique of Barbie, both on body dimensions and its relationship to the body image of girls, and excessive consumerism. It also, however, makes clear the fact that Barbie was the first doll to give girls a play image that was not focused on nurturing and family roles, but was an independent, professional adult: playing roles such as airline pilot, astronaut, or presidential candidate. The article also provides brief references to the role of Barbie in a global market economy - its manufacture outside the United States, despite its marketing as an American cultural icon, and its manufacturer's early adoption of direct-to-children marketing. *Wikipedia* provides more or less all the information provided in the *Britannica* definition, including a reference to Lord's own book, and adds substantially more material from within Barbie lore itself and a detailed time line of the doll's

history. It has a strong emphasis on the body image controversy, and emphasizes both the critique that Barbie encourages girls to focus on shallow consumption of fashion accessories, and that she represents an unattainable lifestyle for most girls who play with her. The very first version of the definition, posted January 3, 2003, included only a brief reference to a change in Barbie's waistline as a result of efforts by parents and anorexia groups concerned with the doll's impact on girls' nutrition. This remained the only reference to the critique of Barbie until December 15, 2003, when a user who was not logged in introduced a fairly roughly written section that emphasized both the body image concerns and the consumerism concerns with Barbie. During the same day, a number of regular contributors (that is, users with log-in names and their own talk pages) edited the new section and improved its language and flow, but kept the basic concepts intact. Three weeks later, on January 5, 2004, another regular user rewrote the section, reorganized the paragraphs so that the critique of Barbie's emphasis on high consumption was separated from the emphasis on Barbie's body dimensions, and also separated and clarified the qualifying claims that Barbie's independence and professional outfits may have had positive effects on girls' perception of possible life plans. This contributor also introduced a reference to the fact that the term "Barbie" is often used to denote a shallow or silly girl or woman. After that, with a change three weeks later from describing Barbie as available for most of her life only as "white Anglo-Saxon (and probably protestant)" to "white woman of apparently European descent" this part of the definition stabilized. As this description aims to make clear, *Wikipedia* makes the history of the evolution of the article entirely transparent. The software platform allows any reader to look at prior versions of the definition, to compare specific versions, and to read the "talk" pages - the pages where the participants discuss their definition and their thoughts about it.

The relative emphasis of Google and *Wikipedia*, on the one hand, and Overture, Yahoo!, and the commercial encyclopedias other than *Britannica*, on the other hand, is emblematic of a basic difference between markets and social conversations with regard to culture. If we focus on the role of culture as "common knowledge" or background knowledge, its relationship to the market - at least for theoretical economists - is exogenous. It can be taken as given and treated as "taste." In more practical business environments, culture is indeed a source of taste and demand, but it is not taken as exogenous. Culture, symbolism, and meaning, as they are tied with market-based goods, become a major focus of advertising and of demand management. No one who has been exposed to the advertising campaigns of Coca-Cola, Nike, or Apple Computers, as well as practically to any one of a broad range of advertising campaigns over the past few decades, can fail to see that these are not primarily a communication about the material characteristics or qualities of the products or services sold by the advertisers. They are about meaning. These campaigns try to invest the act of buying their products or services with a cultural meaning that they cultivate, manipulate, and try to generalize in the practices of the society in which they are advertising, precisely in order to shape taste. They offer an opportunity to generate rents, because the consumer has to have this company's shoe rather than that one, because that particular shoe makes the customer this kind of person rather than that kind - cool rather than stuffy, sophisticated rather than common. Neither the theoretical economists nor the marketing executives have any interest in rendering culture transparent or writable. Whether one treats culture as exogenous or as a domain for limiting the elasticity of demand for one's particular product, there is no impetus to make it easier for consumers to see through the cultural symbols, debate their significance, or make them their own. If there is business reason to do anything about culture, it is to try to shape the cultural meaning of an object or practice, in order to shape the demand for it, while keeping the role of culture hidden and assuring control over the careful cultural choreography of the symbols attached to the company. Indeed, in 1995, the U.S. Congress enacted a new kind of trademark law, the Federal Antidilution Act, which for the first time disconnects trademark protection from protecting consumers from confusion by knockoffs. The Antidilution Act of 1995 gives the owner of any

famous mark - and only famous marks - protection from any use that dilutes the meaning that the brand owner has attached to its own mark. It can be entirely clear to consumers that a particular use does not come from the owner of the brand, and still, the owner has a right to prevent this use. While there is some constitutional free-speech protection for criticism, there is also a basic change in the understanding of trademark law - from a consumer protection law intended to assure that consumers can rely on the consistency of goods marked in a certain way, to a property right in controlling the meaning of symbols a company has successfully cultivated so that they are, in fact, famous. This legal change marks a major shift in the understanding of the role of law in assigning control for cultural meaning generated by market actors.

Unlike market production of culture, meaning making as a social, nonmarket practice has no similar systematic reason to accept meaning as it comes. Certainly, some social relations do. When girls play with dolls, collect them, or exhibit them, they are rarely engaged in reflection on the meaning of the dolls, just as fans of Scarlett O'Hara, of which a brief Internet search suggests there are many, are not usually engaged in critique of *Gone with the Wind* as much as in replication and adoption of its romantic themes. Plainly, however, some conversations we have with each other are about who we are, how we came to be who we are, and whether we view the answers we find to these questions as attractive or not. In other words, some social interactions do have room for examining culture as well as inhabiting it, for considering background knowledge for what it is, rather than taking it as a given input into the shape of demand or using it as a medium for managing meaning and demand. People often engage in conversations with each other precisely to understand themselves in the world, their relationship to others, and what makes them like and unlike those others. One major domain in which this formation of self- and group identity occurs is the adoption or rejection of, and inquiry into, cultural symbols and sources of meaning that will make a group cohere or splinter; that will make people like or unlike each other.

The distinction I draw here between market-based and nonmarket-based activities is purposefully overstated to clarify the basic structural differences between these two modes of organizing communications and the degree of transparency of culture they foster. As even the very simple story of how Barbie is defined in Internet communications demonstrates, practices are not usually as cleanly divided. Like the role of the elite newspapers in providing political coverage, discussed in chapter 6, some market-based efforts do provide transparency; indeed, their very market rationale pushes them to engage in a systematic effort to provide transparency. Google's strategy from the start has been to assume that what individuals are interested in is a reflection of what other individuals - who are interested in roughly the same area, but spend more time on it, that is, Web page authors - think is worthwhile. The company built its business model around rendering transparent what people and organizations that make their information available freely consider relevant. Occasionally, Google has had to deal with "search engine optimizers," who have advised companies on how to game its search engine to achieve a high ranking. Google has fought these optimizers; sometimes by outright blocking access to traffic that originates with them. In these cases, we see a technical competition between firms - the optimizers - whose interest is in capturing attention based on the interests of those who pay them, and a firm, Google, whose strategic choice is to render the distributed judgments of relevance on the Web more or less faithfully. There, the market incentive actually drives Google's investment affirmatively toward transparency. However, the market decision must be strategic, not tactical, for this to be the case. Fear of litigation has, for example, caused Google to bury links that threatened it with liability. The most prominent of these cases occurred when the Church of Scientology threatened to sue Google over presenting links to www.xenu.net, a site dedicated to criticizing scientology. Google initially removed the link. However, its strategic interest was brought to the fore by widespread criticism of its decision on the Internet, and the firm relented. A search for "Scientology" as of this writing reveals a wide range of

sites, many critical of scientology, and xenu.net is the second link. A search for "scientology Google" will reveal many stories, not quite flattering either to Google or to the Church of Scientology, as the top links. We see similar diversity among the encyclopedias. *Britannica* offered as clear a presentation of the controversy over Barbie as *Wikipedia*. *Britannica* has built its reputation and business model on delivery of the knowledge and opinions of those in positions to claim authority in the name of high culture professional competence, and delivering that perspective to those who buy the encyclopedia precisely to gain access to that kind of knowledge base, judgment, and formal credibility. In both cases, the long-term business model of the companies calls for reflecting the views and insights of agents who are not themselves thoroughly within the market - whether they are academics who write articles for *Britannica*, or the many and diverse Web page owners on the Internet. In both cases, these business models lead to a much more transparent cultural representation than what Hollywood or Madison Avenue produce. Just as not all market-based organizations render culture opaque, not all nonmarket or social-relations-based conversations aim to explore and expose cultural assumptions. Social conversations can indeed be among the most highly deferential to cultural assumptions, and can repress critique more effectively and completely than market-based conversations. Whether in communities of unquestioning religious devotion or those that enforce strict egalitarian political correctness, we commonly see, in societies both traditional and contemporary, significant social pressures against challenging background cultural assumptions within social conversations. We have, for example, always had more cultural experimentation and fermentation in cities, where social ties are looser and communities can exercise less social control over questioning minds and conversation. Ubiquitous Internet communications expand something of the freedom of city parks and streets, but also the freedom of cafés and bars - commercial platforms for social interaction - so that it is available everywhere.

The claim I make here, as elsewhere throughout this book, is not that nonmarket production will, in fact, generally displace market production, or that such displacement is necessary to achieve the improvement in the degree of participation in cultural production and legibility. My claim is that the emergence of a substantial nonmarket alternative path for cultural conversation increases the degrees of freedom available to individuals and groups to engage in cultural production and exchange, and that doing so increases the transparency of culture to its inhabitants. It is a claim tied to the particular technological moment and its particular locus of occurrence - our networked communications environment. It is based on the fact that it is displacing the particular industrial form of information and cultural production of the twentieth century, with its heavy emphasis on consumption in mass markets. In this context, the emergence of a substantial sector of nonmarket production, and of peer production, or the emergence of individuals acting cooperatively as a major new source of defining widely transmissible statements and conversations about the meaning of the culture we share, makes culture substantially more transparent and available for reflection, and therefore for revision.

Two other dimensions are made very clear by the *Wikipedia* example. The first is the degree of self-consciousness that is feasible with open, conversation-based definition of culture that is itself rendered more transparent. The second is the degree to which the culture is writable, the degree to which individuals can participate in mixing and matching and making their own emphases, for themselves and for others, on the existing set of symbols. Fisher, for example, has used the term "semiotic democracy" to describe the potential embodied in the emerging openness of Internet culture to participation by users. The term originates from Fiske's *Television Culture* as a counterpoint to the claim that television was actually a purely one-way medium that only enacted culture on viewers. Instead, Fiske claimed that viewers resist these meanings, put them in their own contexts, use them in various ways, and subvert them to make their own meaning. However, much of this

resistance is unstated, some of it unself-conscious. There are the acts of reception and interpretation, or of using images and sentences in different contexts of life than those depicted in the television program; but these acts are local, enacted within small-scale local cultures, and are not the result of a self-conscious conversation among users of the culture about its limits, its meanings, and its subversions. One of the phenomena we are beginning to observe on the Internet is an emerging culture of conversation about culture, which is both self-conscious and informed by linking or quoting from specific reference points. The *Wikipedia* development of the definition of Barbie, its history, and the availability of a talk page alongside it for discussion about the definition, are an extreme version of self-conscious discussion about culture. The basic tools enabled by the Internet-cutting, pasting, rendering, annotating, and commenting - make active utilization and conscious discussion of cultural symbols and artifacts easier to create, sustain, and read more generally.

The flexibility with which cultural artifacts - meaning-carrying objects - can be rendered, preserved, and surrounded by different context and discussion makes it easy for anyone, anywhere, to make a self-conscious statement about culture. They enable what Balkin has called "glomming on" - taking that which is common cultural representation and reworking it into your own move in a cultural conversation.[/7](#) The low cost of storage, and the ubiquitous possibility of connecting from any connection location to any storage space make any such statement persistent and available to others. The ease of commenting, linking, and writing to other locations of statements, in turn, increases the possibility of response and counterresponse. These conversations can then be found by others, and at least read if not contributed to. In other words, as with other, purposeful peer-produced projects like *Wikipedia*, the basic characteristics of the Internet in general and the World Wide Web in particular have made it possible for anyone, anywhere, for any reason to begin to contribute to an accretion of conversation about well-defined cultural objects or about cultural trends and characteristics generally. These conversations can persist across time and exist across distance, and are available for both active participation and passive reading by many people in many places. The result is, as we are already seeing it, the emergence of widely accessible, self-conscious conversation about the meaning of contemporary culture by those who inhabit it. This "writability" is also the second characteristic that the *Wikipedia* definition process makes very clear, and the second major change brought about by the networked information economy in the digital environment.

The Plasticity of Internet Culture: The Future of High-Production-Value Folk Culture

I have already described the phenomena of blogs, of individually created movies like *The Jedi Saga*, and of Second Life, the game platform where users have made all the story lines and all the objects, while the commercial provider created the tools and hosts the platform for their collective storytelling. We are seeing the broad emergence of business models that are aimed precisely at providing users with the tools to write, compose, film, and mix existing materials, and to publish, play, render, and distribute what we have made to others, everywhere. Blogger, for example, provides simple tools for online publication of written materials. Apple Computer offers a product called GarageBand, that lets users compose and play their own music. It includes a large library of prerecorded building blocks - different instruments, riffs, loops - and an interface that allows the user to mix, match, record and add their own, and produce their own musical composition and play it. Video-editing utilities, coupled with the easy malleability of digital video, enable people to make films - whether about their own lives or, as in the case of *The Jedi Saga*, of fantasies. The emerging phenomenon of Machinima - short movies that are made using game platforms - underscores how digital platforms can also become tools for creation in unintended ways. Creators use the 3-D rendering capabilities of an existing game, but use the game to stage a movie scene or video

presentation, which they record as it is played out. This recording is then distributed on the Internet as a stand-alone short film. While many of these are still crude, the basic possibilities they present as modes of making movies is significant. Needless to say, not everyone is Mozart. Not everyone is even a reasonably talented musician, author, or filmmaker. Much of what can be and is done is not wildly creative, and much of it takes the form of Balkin's "glomming on": That is, users take existing popular culture, or otherwise professionally created culture, and perform it, sometimes with an effort toward fidelity to the professionals, but often with their own twists, making it their own in an immediate and unmediated way. However, just as learning how to read music and play an instrument can make one a better-informed listener, so too a ubiquitous practice of making cultural artifacts of all forms enables individuals in society to be better readers, listeners, and viewers of professionally produced culture, as well as contributors of our own statements into this mix of collective culture.

People have always created their own culture. Popular music did not begin with Elvis. There has always been a folk culture - of music, storytelling, and theater. What happened over the course of the twentieth century in advanced economies, and to a lesser extent but still substantially around the globe, is the displacement of folk culture by commercially produced mass popular culture. The role of the individuals and communities vis-à-vis cultural artifacts changed, from coproducers and replicators to passive consumers. The time frame where elders might tell stories, children might put on a show for the adults, or those gathered might sing songs came to be occupied by background music, from the radio or phonograph, or by television. We came to assume a certain level of "production values" - quality of sound and image, quality of rendering and staging - that are unattainable with our crude means and our relatively untrained voices or use of instruments. Not only time for local popular creation was displaced, therefore, but also a sense of what counted as engaging, delightful articulation of culture. In a now-classic article from 1937, "The Work of Art in the Age of Mechanical Reproduction," Walter Benjamin authored one of the only instances of critical theory that took an optimistic view of the emergence of popular culture in the twentieth century as a potentially liberating turn. Benjamin's core claim was that with mechanical replication of art, the "aura" that used to attach to single works of art is dissipated. Benjamin saw this aura of unique works of art as reinforcing a distance between the masses and the representations of culture, reinforcing the perception of their weakness and distance from truly great things. He saw in mechanical reproducibility the possibility of bringing copies down to earth, to the hands of the masses, and reversing the sense of distance and relative weakness of the mass culture. What Benjamin did not yet see were the ways in which mechanical reproduction would insert a different kind of barrier between many dispersed individuals and the capacity to make culture. The barrier of production costs, production values, and the star system that came along with them, replaced the iconic role of the unique work of art with new, but equally high barriers to participation in making culture. It is precisely those barriers that the capabilities provided by digital media begin to erode. It is becoming feasible for users to cut and paste, "glom on," to existing cultural materials; to implement their intuitions, tastes, and expressions through media that render them with newly acceptable degrees of technical quality, and to distribute them among others, both near and far. As Hollywood begins to use more computer-generated special effects, but more important, whole films - 2004 alone saw major releases like *Shrek 2*, *The Incredibles*, and *Polar Express* - and as the quality of widely available image-generation software and hardware improves, the production value gap between individual users or collections of users and the commercial-professional studios will decrease. As this book is completed in early 2005, nothing makes clearer the value of retelling basic stories through the prism of contemporary witty criticism of prevailing culture than do *Shrek 2* and *The Incredibles*, and, equally, nothing exposes the limits of purely technical, movie-star-centered quality than the lifelessness of *Polar Express*. As online games like Second Life provide users with new tools and platforms to tell and retell their own stories, or their own versions of well-trodden

paths, as digital multimedia tools do the same for individuals outside of the collaborative storytelling platforms, we can begin to see a reemergence of folk stories and songs as widespread cultural practices. And as network connections become ubiquitous, and search engines and filters improve, we can begin to see this folk culture emerging to play a substantially greater role in the production of our cultural environment.

A Participatory Culture: Toward Policy

Culture is too broad a concept to suggest an all-encompassing theory centered around technology in general or the Internet in particular. My focus is therefore much narrower, along two dimensions. First, I am concerned with thinking about the role of culture to human interactions that can be understood in terms of basic liberal political commitments - that is to say, a concern for the degree of freedom individuals have to form and pursue a life plan, and the degree of participation they can exercise in debating and determining collective action. Second, my claim is focused on the relative attractiveness of the twentieth-century industrial model of cultural production and what appears to be emerging as the networked model in the early twenty-first century, rather than on the relationship of the latter to some theoretically defined ideal culture.

A liberal political theory cannot wish away the role of culture in structuring human events. We engage in wide ranges of social practices of making and exchanging symbols that are concerned with how our life is and how it might be, with which paths are valuable for us as individuals to pursue and which are not, and with what objectives we as collective communities - from the local to the global - ought to pursue. This unstructured, ubiquitous conversation is centrally concerned with things that a liberal political system speaks to, but it is not amenable to anything like an institutionalized process that could render its results "legitimate." Culture operates as a set of background assumptions and common knowledge that structure our understanding of the state of the world and the range of possible actions and outcomes open to us individually and collectively. It constrains the range of conversational moves open to us to consider what we are doing and how we might act differently. In these regards, it is a source of power in the critical-theory sense - a source that exerts real limits on what we can do and how we can be. As a source of power, it is not a natural force that stands apart from human endeavor and is therefore a fact that is not itself amenable to political evaluation. As we see well in the efforts of parents and teachers, advertising agencies and propaganda departments, culture is manipulable, manageable, and a direct locus of intentional action aimed precisely at harnessing its force as a way of controlling the lives of those who inhabit it. At the same time, however, culture is not the barrel of a gun or the chains of a dungeon. There are limits on the degree to which culture can actually control those who inhabit it. Those degrees depend to a great extent on the relative difficulty or ease of seeing through culture, of talking about it with others, and of seeing other alternatives or other ways of symbolizing the possible and the desirable.

Understanding that culture is a matter of political concern even within a liberal framework does not, however, translate into an agenda of intervention in the cultural sphere as an extension of legitimate political decision making. Cultural discourse is systematically not amenable to formal regulation, management, or direction from the political system. First, participation in cultural discourse is intimately tied to individual self-expression, and its regulation would therefore require levels of intrusion in individual autonomy that would render any benefits in terms of a participatory political system Pyrrhic indeed. Second, culture is much more intricately woven into the fabric of everyday life than political processes and debates. It is language - the basic framework within which we can comprehend anything, and through which we do so everywhere. To regulate culture is to regulate our very comprehension of the world we occupy. Third, therefore, culture infuses our thoughts at a

wide range of levels of consciousness. Regulating culture, or intervening in its creation and direction, would entail self-conscious action to affect citizens at a subconscious or weakly conscious level. Fourth, and finally, there is no Archimedean point outside of culture on which to stand and decide - let us pour a little bit more of this kind of image or that, so that we achieve a better consciousness, one that better fits even our most just and legitimately arrived - at political determinations.

A systematic commitment to avoid direct intervention in cultural exchange does not leave us with nothing to do or say about culture, and about law or policy as it relates to it. What we have is the capacity and need to observe a cultural production and exchange system and to assure that it is as unconstraining and free from manipulation as possible. We must diagnose what makes a culture more or less opaque to its inhabitants; what makes it more or less liable to be strictly constraining of the conversations that rely on it; and what makes the possibility of many and diverse sources and forms of cultural intervention more or less likely. On the background of this project, I suggest that the emergence of Internet culture is an attractive development from the perspective of liberal political theory. This is so both because of the technical characteristics of digital objects and computer network communications, and because of the emerging industrial structure of the networked information economy - typified by the increased salience of nonmarket production in general and of individual production, alone or in concert with others, in particular. The openness of digital networks allows for a much wider range of perspectives on any particular symbol or range of symbols to be visible for anyone, everywhere. The cross section of views that makes it easy to see that Barbie is a contested symbol makes it possible more generally to observe very different cultural forms and perspectives for any individual. This transparency of background unstated assumptions and common knowledge is the beginning of self-reflection and the capacity to break out of given molds. Greater transparency is also a necessary element in, and a consequence of, collaborative action, as various participants either explicitly, or through negotiating the divergence of their nonexplicit different perspectives, come to a clearer statement of their assumptions, so that these move from the background to the fore, and become more amenable to examination and revision. The plasticity of digital objects, in turn, improves the degree to which individuals can begin to produce a new folk culture, one that already builds on the twentieth-century culture that was highly unavailable for folk retelling and re-creation. This plasticity, and the practices of writing your own culture, then feed back into the transparency, both because the practice of making one's own music, movie, or essay makes one a more self-conscious user of the cultural artifacts of others, and because in retelling anew known stories, we again come to see what the originals were about and how they do, or do not, fit our own sense of how things are and how they ought to be. There is emerging a broad practice of learning by doing that makes the entire society more effective readers and writers of their own culture.

By comparison to the highly choreographed cultural production system of the industrial information economy, the emergence of a new folk culture and of a wider practice of active personal engagement in the telling and retelling of basic cultural themes and emerging concerns and attachments offers new avenues for freedom. It makes culture more participatory, and renders it more legible to all its inhabitants. The basic structuring force of culture is not eliminated, of course. The notion of floating monads disconnected from a culture is illusory. Indeed, it is undesirable. However, the framework that culture offers us, the language that makes it possible for us to make statements and incorporate the statements of others in the daily social conversation that pervades life, is one that is more amenable to our own remaking. We become more sophisticated users of this framework, more self-conscious about it, and have a greater capacity to recognize, challenge, and change that which we find oppressive, and to articulate, exchange, and adopt that which we find enabling. As chapter 11 makes clear, however, the tension between the industrial model of

cultural production and the networked information economy is nowhere more pronounced than in the question of the degree to which the new folk culture of the twenty-first century will be permitted to build upon the outputs of the twentieth-century industrial model. In this battle, the stakes are high. One cannot make new culture ex nihilo. We are as we are today, as cultural beings, occupying a set of common symbols and stories that are heavily based on the outputs of that industrial period. If we are to make this culture our own, render it legible, and make it into a new platform for our needs and conversations today, we must find a way to cut, paste, and remix present culture. And it is precisely this freedom that most directly challenges the laws written for the twentieth-century technology, economy, and cultural practice.

Notes

1. Karl Marx, "Introduction to a Contribution to the Critique of Hegel's Philosophy of Right," *Deutsch-Französicher Jahrbucher* (1844).

2. Bruce A. Ackerman, *Social Justice and the Liberal State* (New Haven, CT, and London: Yale University Press, 1980), 333-335, 141-146.

3. Michael Walzer, *Spheres of Justice: A Defense of Pluralism and Equality* (New York: Basic Books, 1983), 29.

4. Will Kymlicka, *Multicultural Citizenship: A Liberal Theory of Minority Rights* (Oxford: Clarendon Press, 1995), 76, 83.

5. Jurgen Habermas, *Between Facts and Norms, Contributions to a Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1998), 22-23.

6. Encyclopedia.com is a part of Highbeam Research, Inc., which combines free and pay research services. Bartleby provides searching and access to many reference and high-culture works at no charge, combining it with advertising, a book store, and many links to Amazon.com or to the publishers for purchasing the printed versions of the materials.

7. Jack Balkin, "Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society," *New York University Law Review* 79 (2004): 1.

Chapter 9
Justice and Development

How will the emergence of a substantial sector of nonmarket, commons-based production in the information economy affect questions of distribution and human well-being? The pessimistic answer is, very little. Hunger, disease, and deeply rooted racial, ethnic, or class stratification will not be solved by a more decentralized, nonproprietary information production system. Without clean water, basic literacy, moderately well-functioning governments, and universal practical adoption of the commitment to treat all human beings as fundamentally deserving of equal regard, the fancy Internet-based society will have little effect on the billions living in poverty or deprivation, either in the rich world, or, more urgently and deeply, in poor and middle-income economies. There is enough truth in this pessimistic answer to require us to tread lightly in embracing the belief that the shift to a networked information economy can indeed have meaningful effects in the domain of justice and human development.

Despite the caution required in overstating the role that the networked information economy can play in solving issues of justice, it is important to recognize that information, knowledge, and culture are core inputs into human welfare. Agricultural knowledge and biological innovation are central to food security. Medical innovation and access to its fruits are central to living a long and healthy life. Literacy and education are central to individual growth, to democratic self-governance, and to economic capabilities. Economic growth itself is critically dependent on innovation and information. For all these reasons, information policy has become a critical element of development policy and the question of how societies attain and distribute human welfare and well-being. Access to knowledge has become central to human development. The emergence of the networked information economy offers definable opportunities for improvement in the normative domain of justice, as it does for freedom, by comparison to what was achievable in the industrial information economy.

We can analyze the implications of the emergence of the networked information economy for justice or equality within two quite different frames. The first is liberal, and concerned primarily with some form of equality of opportunity. The second is social-democratic, or development oriented, and focused on universal provision of a substantial set of elements of human well-being. The availability of information from nonmarket sources and the range of opportunities to act within a nonproprietary production environment improve distribution in both these frameworks, but in different ways. Despite the differences, within both frameworks the effect crystallizes into one of access - access to opportunities for one's own action, and access to the outputs and inputs of the information economy. The industrial economy creates cost barriers and transactional-institutional barriers to both these domains. The networked information economy reduces both types of barriers, or creates alternative paths around them. It thereby equalizes, to some extent, both the opportunities to participate as an economic actor and the practical capacity to partake of the fruits of the increasingly information-based global economy.

The opportunities that the network information economy offers, however, often run counter to the central policy drive of both the United States and the European Union in the international trade and intellectual property systems. These two major powers have systematically pushed for ever-stronger proprietary protection and increasing reliance on strong patents, copyrights, and similar exclusive rights as the core information policy for growth and development. Chapter 2 explains why such a policy is suspect from a purely economic perspective concerned with optimizing innovation. A system that relies too heavily on proprietary approaches to information

production is not, however, merely inefficient. It is unjust. Proprietary rights are designed to elicit signals of people's willingness and ability to pay. In the presence of extreme distribution differences like those that characterize the global economy, the market is a poor measure of comparative welfare. A system that signals what innovations are most desirable and rations access to these innovations based on ability, as well as willingness, to pay, overrepresents welfare gains of the wealthy and underrepresents welfare gains of the poor. Twenty thousand American teenagers can simply afford, and will be willing to pay, much more for acne medication than the more than a million Africans who die of malaria every year can afford to pay for a vaccine. A system that relies too heavily on proprietary models for managing information production and exchange is unjust because it is geared toward serving small welfare increases for people who can pay a lot for incremental improvements in welfare, and against providing large welfare increases for people who cannot pay for what they need.

Liberal Theories of Justice and the Networked Information Economy

Liberal theories of justice can be categorized according to how they characterize the sources of inequality in terms of luck, responsibility, and structure. By luck, I mean reasons for the poverty of an individual that are beyond his or her control, and that are part of that individual's lot in life unaffected by his or her choices or actions. By responsibility, I mean causes for the poverty of an individual that can be traced back to his or her actions or choices. By structure, I mean causes for the inequality of an individual that are beyond his or her control, but are traceable to institutions, economic organizations, or social relations that form a society's transactional framework and constrain the behavior of the individual or undermine the efficacy of his or her efforts at self-help.

We can think of John Rawls's *Theory of Justice* as based on a notion that the poorest people are the poorest because of dumb luck. His proposal for a systematic way of defending and limiting redistribution is the "difference principle." A society should organize its redistribution efforts in order to make those who are least well-off as well-off as they can be. The theory of desert is that, because any of us could in principle be the victim of this dumb luck, we would all have agreed, if none of us had known where we would be on the distribution of bad luck, to minimize our exposure to really horrendous conditions. The practical implication is that while we might be bound to sacrifice some productivity to achieve redistribution, we cannot sacrifice too much. If we did that, we would most likely be hurting, rather than helping, the weakest and poorest. Libertarian theories of justice, most prominently represented by Robert Nozick's entitlement theory, on the other hand, tend to ignore bad luck or impoverishing structure. They focus solely on whether the particular holdings of a particular person at any given moment are unjustly obtained. If they are not, they may not justly be taken from the person who holds them. Explicitly, these theories ignore the poor. As a practical matter and by implication, they treat responsibility as the source of the success of the wealthy, and by negation, the plight of the poorest - leading them to be highly resistant to claims of redistribution.

The basic observation that an individual's economic condition is a function of his or her own actions does not necessarily resolve into a blanket rejection of redistribution, as we see in the work of other liberals. Ronald Dworkin's work on inequality offers a critique of Rawls's, in that it tries to include a component of responsibility alongside recognition of the role of luck. In his framework, if (1) resources were justly distributed and (2) bad luck in initial endowment were compensated through some insurance scheme, then poverty that resulted from bad choices, not bad luck, would not deserve help through redistribution. While Rawls's theory ignores personal responsibility, and in this regard, is less attractive from the perspective of a liberal theory that respects individual autonomy, it has the advantage of offering a much clearer metric for a just system. One can

measure the welfare of the poorest under different redistribution rules in market economies. One can then see how much redistribution is too much, in the sense that welfare is reduced to the point that the poorest are actually worse off than they would be under a less-egalitarian system. You could compare the Soviet Union, West Germany, and the United States of the late 1960s-early 1970s, and draw conclusions. Dworkin's insurance scheme would require too fine an ability to measure the expected incapacitating effect of various low endowments - from wealth to intelligence to health - in a market economy, and to calibrate wealth endowments to equalize them, to offer a measuring rod for policy. It does, however, have the merit of distinguishing - for purposes of judging desert to benefit from society's redistribution efforts - between a child of privilege who fell into poverty through bad investments coupled with sloth and a person born into a poor family with severe mental defects. Bruce Ackerman's *Social Justice and the Liberal State* also provides a mechanism of differentiating the deserving from the undeserving, but adds policy tractability by including the dimension of structure to luck and responsibility. In addition to the dumb luck of how wealthy your parents are when you are born and what genetic endowment you are born with, there are also questions of the education system you grow up with and the transactional framework through which you live your life - which opportunities it affords, and which it cuts off or burdens. His proposals therefore seek to provide basic remedies for those failures, to the extent that they can, in fact, be remedied. One such proposal is Anne Alstott and Ackerman's idea of a government-funded personal endowment at birth, coupled with the freedom to squander it and suffer the consequential reduction in welfare.[/1] He also emphasizes a more open and egalitarian transactional framework that would allow anyone access to opportunities to transact with others, rather than depending on, for example, unequal access to social links as a precondition to productive behavior.

The networked information economy improves justice from the perspective of every single one of these theories of justice. Imagine a good that improves the welfare of its users - it could be software, or an encyclopedia, or a product review. Now imagine a policy choice that could make production of that good on a nonmarket, peer-production basis too expensive to perform, or make it easy for an owner of an input to exclude competitors - both market-based and social-production based. For example, a government might decide to: recognize patents on software interfaces, so that it would be very expensive to buy the right to make your software work with someone else's; impose threshold formal education requirements on the authors of any encyclopedia available for school-age children to read, or impose very strict copyright requirements on using information contained in other sources (as opposed to only prohibiting copying their language) and impose high penalties for small omissions; or give the putative subjects of reviews very strong rights to charge for the privilege of reviewing a product - such as by expanding trademark rights to refer to the product, or prohibiting a reviewer to take apart a product without permission. The details do not matter. I offer them only to provide a sense of the commonplace kinds of choices that governments could make that would, as a practical matter, differentially burden nonmarket producers, whether nonprofit organizations or informal peer-production collaborations. Let us call a rule set that is looser from the perspective of access to existing information resources Rule Set A, and a rule set that imposes higher costs on access to information inputs Rule Set B. As explained in chapter 2, it is quite likely that adopting B would depress information production and innovation, even if it were intended to increase the production of information by, for example, strengthening copyright or patent. This is because the added incentives for some producers who produce with the aim of capturing the rents created by copyright or patents must be weighed against their costs. These include (a) the higher costs even for those producers and (b) the higher costs for all producers who do not rely on exclusive rights at all, but instead use either a nonproprietary market model-like service - or a nonmarket model, like nonprofits and individual authors, and that do not benefit in any way from the increased appropriation. However, let us make here a much weaker assumption - that an increase in the rules of exclusion will not affect overall production. Let us

assume that there will be exactly enough increased production by producers who rely on a proprietary model to offset the losses of production in the nonproprietary sectors.

It is easy to see why a policy shift from A to B would be regressive from the perspective of theories like Rawls's or Ackerman's. Under Rule A, let us say that in this state of affairs, State A, there are five online encyclopedias. One of them is peer produced and freely available for anyone to use. Rule B is passed. In the new State B, there are still five encyclopedias. It has become too expensive to maintain the free encyclopedia, however, and more profitable to run commercial online encyclopedias. A new commercial encyclopedia has entered the market in competition with the four commercial encyclopedias that existed in State A, and the free encyclopedia folded. From the perspective of the difference principle, we can assume that the change has resulted in a stable overall welfare in the Kaldor-Hicks sense. (That is, overall welfare has increased enough so that, even though some people may be worse off, those who have been made better off are sufficiently better off that they could, in principle, compensate everyone who is worse off enough to make everyone either better off or no worse off than they were before.) There are still five encyclopedias. However, now they all charge a subscription fee. The poorest members of society are worse off, even if we posit that total social welfare has remained unchanged. In State A, they had access for free to an encyclopedia. They could use the information (or the software utility, if the example were software) without having to give up any other sources of welfare. In State B, they must choose between the same amount of encyclopedia usage as they had before, and less of some other source of welfare, or the same welfare from other sources, and no encyclopedia. If we assume, contrary to theory and empirical evidence from the innovation economics literature, that the move to State B systematically and predictably improves the incentives and investments of the commercial producers, that would still by itself not justify the policy shift from the perspective of the difference principle. One would have to sustain a much stricter claim: that the marginal improvement in the quality of the encyclopedias, and a decline in price from the added market competition that was not felt by the commercial producers when they were competing with the free, peer-produced version, would still make the poorest better off, even though they now must pay for any level of encyclopedia access, than they were when they had four commercial competitors with their prior levels of investment operating in a competitive landscape of four commercial and one free encyclopedia.

From the perspective of Ackerman's theory of justice, the advantages of the networked information economy are clearer yet. Ackerman characterizes some of the basic prerequisites for participating in a market economy as access to a transactional framework, to basic information, and to an adequate educational endowment. To the extent that any of the basic utilities required to participate in an information economy at all are available without sensitivity to price - that is, free to anyone - they are made available in a form that is substantially insulated from the happenstance of initial wealth endowments. In this sense at least, the development of a networked information economy overcomes some of the structural components of continued poverty - lack of access to information about market opportunities for production and cheaper consumption, about the quality of goods, or lack of communications capacity to people or places where one can act productively. While Dworkin's theory does not provide a similarly clear locus for mapping the effect of the networked information economy on justice, there is some advantage, and no loss, from this perspective, in having more of the information economy function on a nonmarket basis. As long as one recognizes bad luck as a partial reason for poverty, then having information resources available for free use is one mechanism of moderating the effects of bad luck in endowment, and lowers the need to compensate for those effects insofar as they translate to lack of access to information resources. This added access results from voluntary communication by the producers and a respect for their willingness to communicate what they produced freely. While the benefits flow to individuals

irrespective of whether their present state is due to luck or irresponsibility, it does not involve a forced redistribution from responsible individuals to irresponsible individuals.

From the perspective of liberal theories of justice, then, the emergence of the networked information economy is an unqualified improvement. Except under restrictive assumptions inconsistent with what we know as a matter of both theory and empirics about the economics of innovation and information production, the emergence of a substantial sector of information production and exchange that is based on social transactional frameworks, rather than on a proprietary exclusion business model, improves distribution in society. Its outputs are available freely to anyone, as basic inputs into their own actions - whether market-based or nonmarket-based. The facilities it produces improve the prospects of all who are connected to the Internet - whether they are seeking to use it as consumers or as producers. It softens some of the effects of resource inequality. It offers platforms for greater equality of opportunity to participate in market- and nonmarket-based enterprises. This characteristic is explored in much greater detail in the next segment of this chapter, but it is important to emphasize here that equality of opportunity to act in the face of unequal endowment is central to all liberal theories of justice. As a practical matter, these characteristics of the networked information economy make the widespread availability of Internet access a more salient objective of redistribution policy. They make policy debates, which are mostly discussed in today's political sphere in terms of innovation and growth, and sometimes in terms of freedom, also a matter of liberal justice.

Commons-Based Strategies for Human Welfare and Development

There is a long social-democratic tradition of focusing not on theoretical conditions of equality in a liberal society, but on the actual well-being of human beings in a society. This conception of justice shares with liberal theories the acceptance of market economy as a fundamental component of free societies. However, its emphasis is not equality of opportunity or even some level of social insurance that still allows the slothful to fall, but on assuring a basic degree of well-being to everyone in society. Particularly in the European social democracies, the ambition has been to make that basic level quite high, but the basic framework of even American Social Security - unless it is fundamentally changed in the coming years - has this characteristic. The literature on global poverty and its alleviation was initially independent of this concern, but as global communications and awareness increased, and as the conditions of life in most advanced market economies for most people improved, the lines between the concerns with domestic conditions and global poverty blurred. We have seen an increasing merging of the concerns into a concern for basic human well-being everywhere. It is represented in no individual's work more clearly than in that of Amartya Sen, who has focused on the centrality of development everywhere to the definition not only of justice, but of freedom as well.

The emerging salience of global development as the core concern of distributive justice is largely based on the sheer magnitude of the problems faced by much of the world's population./2 In the world's largest democracy, 80 percent of the population - slightly more people than the entire population of the United States and the expanded European Union combined - lives on less than two dollars a day, 39 percent of adults are illiterate, and 47 percent of children under the age of five are underweight for their age. In Africa's wealthiest democracy, a child at birth has a 45 percent probability of dying before he or she reaches the age of forty. India and South Africa are far from being the worst-off countries. The scope of destitution around the globe exerts a moral pull on any acceptable discussion of justice. Intuitively, these problems seem too fundamental to be seriously affected by the networked information economy - what has *Wikipedia* got to do with the 49 percent of the population of Congo that lacks sustainable access to improved water sources? It is, indeed,

important not to be overexuberant about the importance of information and communications policy in the context of global human development. But it is also important not to ignore the centrality of information to most of our more-advanced strategies for producing core components of welfare and development. To see this, we can begin by looking at the components of the Human Development Index (HDI).

The Human Development Report was initiated in 1990 as an effort to measure a broad set of components of what makes a life livable, and, ultimately, attractive. It was developed in contradistinction to indicators centered on economic output, like gross domestic product (GDP) or economic growth alone, in order to provide a more refined sense of what aspects of a nation's economy and society make it more or less livable. It allows a more nuanced approach toward improving the conditions of life everywhere. As Sen pointed out, the people of China, Kerala in India, and Sri Lanka lead much longer and healthier lives than other countries, like Brazil or South Africa, which have a higher per capita income./3 The Human Development Report measures a wide range of outcomes and characteristics of life. The major composite index it tracks is the Human Development Index. The HDI tries to capture the capacity of people to live long and healthy lives, to be knowledgeable, and to have material resources sufficient to provide a decent standard of living. It does so by combining three major components: life expectancy at birth, adult literacy and school enrollment, and GDP per capita. As Figure 9.1 illustrates, in the global information economy, each and every one of these measures is significantly, though not solely, a function of access to information, knowledge, and information-embedded goods and services. Life expectancy is affected by adequate nutrition and access to life-saving medicines. Biotechnological innovation for agriculture, along with agronomic innovation in cultivation techniques and other, lower-tech modes of innovation, account for a high portion of improvements in the capacity of societies to feed themselves and in the availability of nutritious foods. Medicines depend on pharmaceutical research and access to its products, and health care depends on research and publication for the development and dissemination of information about best-care practices. Education is also heavily dependent, not surprisingly, on access to materials and facilities for teaching. This includes access to basic textbooks, libraries, computation and communications systems, and the presence of local academic centers. Finally, economic growth has been understood for more than half a century to be centrally driven by innovation. This is particularly true of latecomers, who can improve their own condition most rapidly by adopting best practices and advanced technology developed elsewhere, and then adapting to local conditions and adding their own from the new technological platform achieved in this way. All three of these components are, then, substantially affected by access to, and use of, information and knowledge. The basic premise of the claim that the emergence of the networked information economy can provide significant benefits to human development is that the manner in which we produce new information - and equally important, the institutional framework we use to manage the stock of existing information and knowledge around the world - can have significant impact on human development.
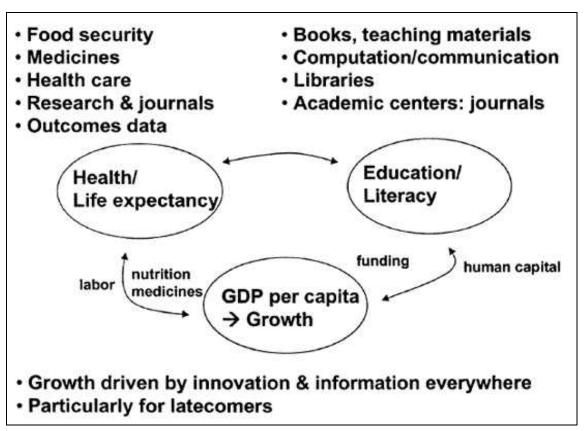
Figure 9.1: HDI and Information

Information-Embedded Goods and Tools, Information, and Knowledge

One can usefully idealize three types of information-based advantages that developed economies have, and that would need to be available to developing and less-developed economies if one's goal were the improvement in conditions in those economies and the opportunities for innovation in them. These include information-embedded material resources - consumption goods and production tools - information, and knowledge.

Information-Embedded Goods. These are goods that are not themselves information, but that are better, more plentiful, or cheaper because of some technological advance embedded in them or associated with their production. Pharmaceuticals and agricultural goods are the most obvious examples in the areas of health and food security, respectively. While there are other constraints on access to innovative products in these areas - regulatory and political in nature - a perennial barrier is cost. And a perennial barrier to competition that could reduce the cost is the presence of exclusive rights, mostly in the form of patents, but also in the form of internationally recognized breeders' rights and regulatory data exclusivity. In the areas of computation and communication, hardware and software are the primary domains of concern. With hardware, there have been some efforts toward developing cheaper equipment - like the simputer and the Jhai computer efforts to develop inexpensive computers. Because of the relatively commoditized state of most components of these systems, however, marginal cost, rather than exclusive rights, has been the primary barrier to access. The solution, if one has emerged, has been aggregation of demand - a networked computer for a village, rather than an individual. For software, the initial solution was piracy. More recently, we have seen an increased use of free software instead. The former cannot genuinely be described as a "solution," and is being eliminated gradually by trade policy efforts. The latter - adoption of free

software to obtain state-of-the-art software - forms the primary template for the class of commons-based solutions to development that I explore in this chapter.

Information-Embedded Tools. One level deeper than the actual useful material things one would need to enhance welfare are tools necessary for innovation itself. In the areas of agricultural biotechnology and medicines, these include enabling technologies for advanced research, as well as access to materials and existing compounds for experimentation. Access to these is perhaps the most widely understood to present problems in the patent system of the developed world, as much as it is for the developing world - an awareness that has mostly crystallized under Michael Heller's felicitous phrase "anti-commons," or Carl Shapiro's "patent thicket." The intuition, whose analytic basis is explained in chapter 2, is that innovation is encumbered more than it is encouraged when basic tools for innovation are proprietary, where the property system gives owners of these tools proprietary rights to control innovation that relies on their tools, and where any given new innovation requires the consent of, and payment to, many such owners. This problem is not unique to the developing world. Nonetheless, because of the relatively small dollar value of the market for medicines that treat diseases that affect only poorer countries or of crop varieties optimized for those countries, the cost hurdle weighs more heavily on the public or nonprofit efforts to achieve food security and health in poor and middle-income countries. These nonmarket-based research efforts into diseases and crops of concern purely to these areas are not constructed to appropriate gains from exclusive rights to research tools, but only bear their costs on downstream innovation.

Information. The distinction between information and knowledge is a tricky one. I use "information" here colloquially, to refer to raw data, scientific reports of the output of scientific discovery, news, and factual reports. I use "knowledge" to refer to the set of cultural practices and capacities necessary for processing the information into either new statements in the information exchange, or more important in our context, for practical use of the information in appropriate ways to produce more desirable actions or outcomes from action. Three types of information that are clearly important for purposes of development are scientific publications, scientific and economic data, and news and factual reports. Scientific publication has seen a tremendous cost escalation, widely perceived to have reached crisis proportions even by the terms of the best-endowed university libraries in the wealthiest countries. Over the course of the 1990s, some estimates saw a 260 percent increase in the prices of scientific publications, and libraries were reported choosing between journal subscription and monograph purchases.[4] In response to this crisis, and in reliance on what were perceived to be the publication cost-reduction opportunities for Internet publication, some scientists - led by Nobel laureate and then head of the National Institutes of Health Harold Varmus - began to agitate for a scientist-based publication system.[5] The debates were, and continue to be, heated in this area. However, currently we are beginning to see the emergence of scientist-run and -driven publication systems that distribute their papers for free online, either within a traditional peer-review system like the Public Library of Science (PLoS), or within tightly knit disciplines like theoretical physics, with only post-publication peer review and revision, as in the case of the Los Alamos Archive, or ArXiv.org. Together with free software and peer production on the Internet, the PLoS and ArXiv.org models offer insights into the basic shape of the class of commons-based, nonproprietary production solutions to problems of information production and exchange unhampered by intellectual property.

Scientific and economic data present a parallel conceptual problem, but in a different legal setting. In the case of both types of data, much of it is produced by government agencies. In the United States, however, raw data is in the public domain, and while initial access may require payment of the cost of distribution, reworking of the data as a tool in information production and innovation - and its redistribution by those who acquired access initially - is considered to be in the public

domain. In Europe, this has not been the case since the 1996 Database Directive, which created a property-like right in raw data in an effort to improve the standing of European database producers. Efforts to pass similar legislation in the United States have been mounted and stalled in practically every Congress since the mid-1990s. These laws continue to be introduced, driven by the lobby of the largest owners of nongovernment databases, and irrespective of the fact that for almost a decade, Europe's database industry has grown only slowly in the presence of a right, while the U.S. database industry has flourished without an exclusive rights regime.

News, market reports, and other factual reporting seem to have escaped the problems of barriers to access. Here it is most likely that the value-appropriation model simply does not depend on exclusive rights. Market data is generated as a by-product of the market function itself. Tiny time delays are sufficient to generate a paying subscriber base, while leaving the price trends necessary for, say, farmers to decide at what prices to sell their grain in the local market, freely available./6 As I suggested in chapter 2, the advertising-supported press has never been copyright dependent, but has instead depended on timely updating of news to capture attention, and then attach that attention to advertising. This has not changed, but the speed of the update cycle has increased and, more important, distribution has become global, so that obtaining most information is now trivial to anyone with access to an Internet connection. While this continues to raise issues with deployment of communications hardware and the knowledge of how to use it, these issues can be, and are being, approached through aggregation of demand in either public or private forms. These types of information do not themselves appear to exhibit significant barriers to access once network connectivity is provided.

Knowledge. In this context, I refer mostly to two types of concern. The first is the possibility of the transfer of implicit knowledge, which resists codification into what would here be treated as "information" - for example, training manuals. The primary mechanism for transfer of knowledge of this type is learning by doing, and knowledge transfer of this form cannot happen except through opportunities for local practice of the knowledge. The second type of knowledge transfer of concern here is formal instruction in an education context (as compared with dissemination of codified outputs for self-teaching). Here, there is a genuine limit on the capacity of the networked information economy to improve access to knowledge. Individual, face-to-face instruction does not scale across participants, time, and distance. However, some components of education, at all levels, are nonetheless susceptible to improvement with the increase in nonmarket and radically decentralized production processes. The MIT Open Courseware initiative is instructive as to how the universities of advanced economies can attempt to make at least their teaching materials and manuals freely available to teachers throughout the world, thereby leaving the pedagogy in local hands but providing more of the basic inputs into the teaching process on a global scale. More important perhaps is the possibility that teachers and educators can collaborate, both locally and globally, on an open platform model like *Wikipedia*, to coauthor learning objects, teaching modules, and, more ambitiously, textbooks that could then be widely accessed by local teachers

Industrial Organization of HDI-Related Information Industries

The production of information and knowledge is very different from the production of steel or automobiles. Chapter 2 explains in some detail that information production has always included substantial reliance on nonmarket actors and on nonmarket, nonproprietary settings as core modalities of production. In software, for example, we saw that Mickey and romantic maximizer-type producers, who rely on exclusive rights directly, have accounted for a stable 36-37 percent of market-based revenues for software developers, while the remainder was focused on both supply-side and demand-side improvements in the capacity to offer software services. This

number actually overstates the importance of software publishing, because it does not at all count free software development except when it is monetized by an IBM or a Red Hat, leaving tremendous value unaccounted for. A very large portion of the investments and research in any of the information production fields important to human development occur within the category that I have broadly described as "Joe Einstein." These include both those places formally designated for the pursuit of information and knowledge in themselves, like universities, and those that operate in the social sphere, but produce information and knowledge as a more or less central part of their existence - like churches or political parties. Moreover, individuals acting as social beings have played a central role in our information production and exchange system. In order to provide a more sector-specific analysis of how commons-based, as opposed to proprietary, strategies can contribute to development, I offer here a more detailed breakdown specifically of software, scientific publication, agriculture, and biomedical innovation than is provided in chapter 2. Table 9.1 presents a higher-resolution statement of the major actors in these fields, within both the market and the nonmarket sectors, from which we can then begin to analyze the path toward, and the sustainability of, more significant commons-based production of the necessities of human development.

Table 9.1: Map of Players and Roles in Major Relevant Sectors

Actor Sector	Government	Universities, Libraries, etc.	IP-Based Industry	Non-IP-Based Industry	NGOs/ Nonprofits	Individuals
Software	Research funding, defense, procurement	Basic research and design; components "incubate" much else	Software publishing (1/3 annual revenue)	Software services, customization (~2/3 annual revenue)	FSF; Apache; W3C; IETF	Free/ open-source software
Scientific publication	Research funding	University presses; salaries; promotion and tenure	Elsevier Science; professional associations	Biomed Central	PLoS; ArXiv	Working papers; Web-based self-publishing
Agricultural Biotech	Grants and government labs; NARS	Basic research; tech transfer (24% of patenting activity)	Monsanto, DuPont, Syngenta (~74% of patents)	No obvious equivalent	CAMBIA BIOS CGIAR	Farmers
Biomed/ Health	Grants and government labs	Basic research; tech transfer (~50%?)	Big Pharma; Biotech (~50%?)	Generics	OneWorld Health	None

Table 9.1 identifies the relative role of each of the types of main actors in information and knowledge production across the major sectors relevant to contemporary policy debates. It is most

important to extract from this table the diversity of business models and roles not only in each industry, but also among industries. This diversity means that different types of actors can have different relative roles: nonprofits as opposed to individuals, universities as opposed to government, or nonproprietary market actors - that is, market actors whose business model is service based or otherwise does not depend on exclusive appropriation of information - as compared to nonmarket actors. The following segments look at each of these sectors more specifically, and describe the ways in which commons-based strategies are already, or could be, used to improve the access to information, knowledge, and the information-embedded goods and tools for human development. However, even a cursory look at the table shows that the current production landscape of software is particularly well suited to having a greater role for commons-based production. For example, exclusive proprietary producers account for only one-third of software-related revenues, even within the market. The remainder is covered by various services and relationships that are compatible with nonproprietary treatment of the software itself. Individuals and nonprofit associations also have played a very large role, and continue to do so, not only in free software development, but in the development of standards as well. As we look at each sector, we see that they differ in their incumbent industrial landscape, and these differences mean that each sector may be more or less amenable to commons-based strategies, and, even if in principle amenable, may present harder or easier transition problems.

Toward Adopting Commons-Based Strategies for Development

The mainstream understanding of intellectual property by its dominant policy-making institutions - the Patent Office and U.S. trade representative in the United States, the Commission in the European Union, and the World Intellectual Property Organization (WIPO) and Trade-Related Aspects of Intellectual Property (TRIPS) systems internationally - is that strong protection is good, and stronger protection is better. In development and trade policy, this translates into a belief that the primary mechanism for knowledge transfer and development in a global information economy is for all nations, developing as well as developed, to ratchet up their intellectual property law standards to fit the most protective regimes adopted in the United States and Europe. As a practical political matter, the congruence between the United States and the European Union in this area means that this basic understanding is expressed in the international trade system, in the World Trade Organization (WTO) and its TRIPS agreement, and in international intellectual property treaties, through the WIPO. The next few segments present an alternative view. Intellectual property as an institution is substantially more ambiguous in its effects on information production than the steady drive toward expansive rights would suggest. The full argument is in chapter 2.

Intellectual property is particularly harmful to net information importers. In our present world trade system, these are the poor and middle-income nations. Like all users of information protected by exclusive rights, these nations are required by strong intellectual property rights to pay more than the marginal cost of the information at the time that they buy it. In the standard argument, this is intended to give producers incentives to create information that users want. Given the relative poverty of these countries, however, practically none of the intellectual-property-dependent producers develop products specifically with returns from poor or even middle-income markets in mind. The pharmaceutical industry receives about 5 percent of its global revenues from low- and middle-income countries. That is why we have so little investment in drugs for diseases that affect only those parts of the world. It is why most agricultural research that has focused on agriculture in poorer areas of the world has been public sector and nonprofit. Under these conditions, the above-marginal-cost prices paid in these poorer countries are purely regressive redistribution. The information, knowledge, and information-embedded goods paid for would have been developed in expectation of rich world rents alone. The prospects of rents from poorer countries do not affect

their development. They do not affect either the rate or the direction of research and development. They simply place some of the rents that pay for technology development in the rich countries on consumers in poor and middle-income countries. The morality of this redistribution from the world's poor to the world's rich has never been confronted or defended in the European or American public spheres. It simply goes unnoticed. When crises in access to information-embedded goods do appear - such as in the AIDS/HIV access to medicines crisis - these are seldom tied to our basic institutional choice. In our trade policies, Americans and Europeans push for ever-stronger protection. We thereby systematically benefit those who own much of the stock of usable human knowledge. We do so at the direct expense of those who need access to knowledge in order to feed themselves and heal their sick.

The practical politics of the international intellectual property and trade regime make it very difficult to reverse the trend toward ever-increasing exclusive property protections. The economic returns to exclusive proprietary rights in information are highly concentrated in the hands of those who own such rights. The costs are widely diffuse in the populations of both the developing and developed world. The basic inefficiency of excessive property protection is difficult to understand by comparison to the intuitive, but mistaken, Economics 101 belief that property is good, more property is better, and intellectual property must be the same. The result is that pressures on the governments that represent exporters of intellectual property rights permissions - in particular, the United States and the European Union - come in this area mostly from the owners, and they continuously push for ever-stronger rights. Monopoly is a good thing to have if you can get it. Its value for rent extraction is no less valuable for a database or patent-based company than it is for the dictator's nephew in a banana republic. However, its value to these supplicants does not make it any more efficient or desirable.

The political landscape is, however, gradually beginning to change. Since the turn of the twenty-first century, and particularly in the wake of the urgency with which the HIV/AIDS crisis in Africa has infused the debate over access to medicines, there has been a growing public interest advocacy movement focused on the intellectual property trade regime. This movement is, however, confronted with a highly playable system. A victory for developing world access in one round in the TRIPS context always leaves other places to construct mechanisms for exclusivity. Bilateral trade negotiations are one domain that is beginning to play an important role. In these, the United States or the European Union can force a rice- or cotton-exporting country to concede a commitment to strong intellectual property protection in exchange for favorable treatment for their core export. The intellectual property exporting nations can then go to WIPO, and push for new treaties based on the emerging international practice of bilateral agreements. This, in turn, would cycle back and be generalized and enforced through the trade regimes. Another approach is for the exporting nations to change their own laws, and then drive higher standards elsewhere in the name of "harmonization." Because the international trade and intellectual property system is highly "playable" and manipulable in these ways, systematic resistance to the expansion of intellectual property laws is difficult.

The promise of the commons-based strategies explored in the remainder of this chapter is that they can be implemented without changes in law - either national or international. They are paths that the emerging networked information economy has opened to individuals, nonprofits, and public-sector organizations that want to help in improving human development in the poorer regions of the world to take action on their own. As with decentralized speech for democratic discourse, and collaborative production by individuals of the information environment they occupy as autonomous agents, here too we begin to see that self-help and cooperative action outside the proprietary system offer an opportunity for those who wish to pursue it. In this case, it is an

opportunity to achieve a more just distribution of the world's resources and a set of meaningful improvements in human development. Some of these solutions are "commons-based," in the sense that they rely on free access to existing information that is in the commons, and they facilitate further use and development of that information and those information-embedded goods and tools by releasing their information outputs openly, and managing them as a commons, rather than as property. Some of the solutions are specifically peer-production solutions. We see this most clearly in software, and to some extent in the more radical proposals for scientific publication. I will also explore here the viability of peer-production efforts in agricultural and biomedical innovation, although in those fields, commons-based approaches grafted onto traditional public-sector and nonprofit organizations at present hold the more clearly articulated alternatives.

Software

The software industry offers a baseline case because of the proven large scope for peer production in free software. As in other information-intensive industries, government funding and research have played an enormously important role, and university research provides much of the basic science. However, the relative role of individuals, nonprofits, and nonproprietary market producers is larger in software than in the other sectors. First, two-thirds of revenues derived from software in the United States are from services and do not depend on proprietary exclusion. Like IBM's "Linux-related services" category, for which the company claimed more than two billion dollars of revenue for 2003, these services do not depend on exclusion from the software, but on charging for service relationships./7 Second, some of the most basic elements of the software environment - like standards and protocols - are developed in nonprofit associations, like the Internet Engineering Taskforce or the World Wide Web Consortium. Third, the role of individuals engaged in peer production - the free and open-source software development communities - is very large. Together, these make for an organizational ecology highly conducive to nonproprietary production, whose outputs can be freely usable around the globe. The other sectors have some degree of similar components, and commons-based strategies for development can focus on filling in the missing components and on leveraging nonproprietary components already in place.

In the context of development, free software has the potential to play two distinct and significant roles. The first is offering low-cost access to high-performing software for developing nations. The second is creating the potential for participation in software markets based on human ability, even without access to a stock of exclusive rights in existing software. At present, there is a movement in both developing and the most advanced economies to increase reliance on free software. In the United States, the Presidential Technology Advisory Commission advised the president in 2000 to increase use of free software in mission-critical applications, arguing the high quality and dependability of such systems. To the extent that quality, reliability, and ease of self-customization are consistently better with certain free software products, they are attractive to developing-country governments for the same reasons that they are to the governments of developed countries. In the context of developing nations, the primary additional arguments that have been made include cost, transparency, freedom from reliance on a single foreign source (read, Microsoft), and the potential of local software programmers to learn the program, acquire skills, and therefore easily enter the global market with services and applications for free software./8 The question of cost, despite the confusion that often arises from the word "free," is not obvious. It depends to some extent on the last hope - that local software developers will become skilled in the free software platforms. The cost of software to any enterprise includes the extent, cost, and efficacy with which the software can be maintained, upgraded, and fixed when errors occur. Free software may or may not involve an up-front charge. Even if it does not, that does not make it cost-free. However, free software enables an open market in free software servicing, which in turn improves and lowers the cost of

servicing the software over time. More important, because the software is open for all to see and because developer communities are often multinational, local developers can come, learn the software, and become relatively low-cost software service providers for their own government. This, in turn, helps realize the low-cost promise over and above the licensing fees avoided. Other arguments in favor of government procurement of free software focus on the value of transparency of software used for public purposes. The basic thrust of these arguments is that free software makes it possible for constituents to monitor the behavior of machines used in governments, to make sure that they are designed to do what they are publicly reported to do. The most significant manifestation of this sentiment in the United States is the hitherto-unsuccessful, but fairly persistent effort to require states to utilize voting machines that use free software, or at a minimum, to use software whose source code is open for public inspection. This is a consideration that, if valid, is equally suitable for developing nations. The concern with independence from a single foreign provider, in the case of operating systems, is again not purely a developing-nation concern. Just as the United States required American Marconi to transfer its assets to an American company, RCA, so that it would not be dependent for a critical infrastructure on a foreign provider, other countries may have similar concerns about Microsoft. Again, to the extent that this is a valid concern, it is so for rich nations as much as it is for poor, with the exceptions of the European Union and Japan, which likely do have bargaining power with Microsoft to a degree that smaller markets do not.

The last and quite distinct potential gain is the possibility of creating a context and an anchor for a free software development sector based on service. This was cited as the primary reason behind Brazil's significant push to use free software in government departments and in telecenters that the federal government is setting up to provide Internet service access to some of its poorer and more remote areas. Software services represent a very large industry. In the United States, software services are an industry roughly twice the size of the movie and video industry. Software developers from low- and middle-income countries can participate in the growing free software segment of this market by using their skills alone. Unlike with service for the proprietary domain, they need not buy licenses to learn and practice the services. Moreover, if Brazil, China, India, Indonesia, and other major developing countries were to rely heavily on free software, then the "internal market," within the developing world, for free software-related services would become very substantial. Building public-sector demand for these services would be one place to start. Moreover, because free software development is a global phenomenon, free software developers who learn their skills within the developing world would be able to export those skills elsewhere. Just as India's call centers leverage the country's colonial past with its resulting broad availability of English speakers, so too countries like Brazil can leverage their active free software development community to provide software services for free software platforms anywhere in the developed and developing worlds. With free software, the developing-world providers can compete as equals. They do not need access to permissions to operate. Their relationships need not replicate the "outsourcing" model so common in proprietary industries, where permission to work on a project is the point of control over the ability to do so. There will still be branding issues that undoubtedly will affect access to developed markets. However, there will be no baseline constraints of minimal capital necessary to enter the market and try to develop a reputation for reliability. As a development strategy, then, utilization of free software achieves transfer of information-embedded goods for free or at low cost. It also transfers information about the nature of the product and its operation - the source code. Finally, it enables transfer, at least potentially, of opportunities for learning by doing and of opportunities for participating in the global market. These would depend on knowledge of a free software platform that anyone is free to learn, rather than on access to financial capital or intellectual property inventories as preconditions to effective participation.

Scientific Publication

Scientific publication is a second sector where a nonproprietary strategy can be implemented readily and is already developing to supplant the proprietary model. Here, the existing market structure is quite odd in a way that likely makes it unstable. Authoring and peer review, the two core value-creating activities, are done by scientists who perform neither task in expectation of royalties or payment. The model of most publications, however, is highly proprietary. A small number of business organizations, like Elsevier Science, control most of the publications. Alongside them, professional associations of scientists also publish their major journals using a proprietary model. Universities, whose scientists need access to the papers, incur substantial cost burdens to pay for the publications as a basic input into their own new work. While the effects of this odd system are heavily felt in universities in rich countries, the burden of subscription rates that go into the thousands of dollars per title make access to up-to-date scientific research prohibitive for universities and scientists working in poorer economies. Nonproprietary solutions are already beginning to emerge in this space. They fall into two large clusters.

The first cluster is closer to the traditional peer-review publication model. It uses Internet communications to streamline the editorial and peer-review system, but still depends on a small, salaried editorial staff. Instead of relying on subscription payments, it relies on other forms of payments that do not require charging a price for the outputs. In the case of the purely nonprofit Public Library of Science (PLoS), the sources of revenue combine author's payments for publication, philanthropic support, and university memberships. In the case of the for-profit BioMed Central, based in the United Kingdom, it is a combination of author payments, university memberships, and a variety of customized derivative products like subscription-based literature reviews and customized electronic update services. Author payments - fees authors must pay to have their work published - are built into the cost of scientific research and included in grant applications. In other words, they are intended to be publicly funded. Indeed, in 2005, the National Institutes of Health (NIH), the major funding agency for biomedical science in the United States, announced a requirement that all NIH-funded research be made freely available on the Web within twelve months of publication. Both PLoS and BioMed Central have waiver processes for scientists who cannot pay the publication fees. The articles on both systems are available immediately for free on the Internet. The model exists. It works internally and is sustainable as such. What is left in determining the overall weight that these open-access journals will have in the landscape of scientific publication is the relatively conservative nature of universities themselves. The established journals, like *Science* or *Nature*, still carry substantially more prestige than the new journals. As long as this is the case, and as long as hiring and promotion decisions continue to be based on the prestige of the journal in which a scientist's work is published, the ability of the new journals to replace the traditional ones will be curtailed. Some of the established journals, however, are operated by professional associations of scientists. There is an internal tension between the interests of the associations in securing their revenue and the growing interest of scientists in open-access publication. Combined with the apparent economic sustainability of the open-access journals, it seems that some of these established journals will likely shift over to the open-access model. At a minimum, policy interventions like those proposed by the NIH will force traditional publications to adapt their business model by making access free after a few months. The point here, however, is not to predict the overall likely success of open-access journals. It is to combine them with what we have seen happening in software as another example of a reorganization of the components of the industrial structure of an information production system. Individual scientists, government funding agencies, nonprofits and foundations, and nonproprietary commercial business models can create the same good - scientific publication - but without the cost barrier that the old model imposed on

access to its fruits. Such a reorientation would significantly improve the access of universities and physicians in developing nations to the most advanced scientific publication.

The second approach to scientific publication parallels more closely free software development and peer production. This is typified by ArXiv and the emerging practices of self-archiving or self-publishing. ArXiv.org is an online repository of working papers in physics, mathematics, and computer science. It started out focusing on physics, and that is where it has become the sine qua non of publication in some subdisciplines. The archive does not perform review except for technical format compliance. Quality control is maintained by postpublication review and commentary, as well as by hosting updated versions of the papers with explanations (provided by authors) of the changes. It is likely that the reason ArXiv.org has become so successful in physics is the very small and highly specialized nature of the discipline. The universe of potential readers is small, and their capacity to distinguish good arguments from bad is high. Reputation effects of poor publications are likely immediate.

While ArXiv offers a single repository, a much broader approach has been the developing practice of self-archiving. Academics post their completed work on their own Web sites and make it available freely. The primary limitation of this mechanism is the absence of an easy, single location where one can search for papers on a topic of concern. And yet we are already seeing the emergence of tagging standards and protocols that allow anyone to search the universe of self-archived materials. Once completed, such a development process would in principle render archiving by single points of reference unnecessary. The University of Michigan Digital Library Production Service, for example, has developed a protocol called OAIster (pronounced like oyster, with the tagline "find the pearls"), which combines the acronym of Open Archives Initiative with the "ster" ending made popular in reference to peer-to-peer distribution technologies since Napster (AIMster, Grokster, Friendster, and the like). The basic impulse of the Open Archives Initiative is to develop a sufficiently refined set of meta-data tags that would allow anyone who archives their materials with OAI-compliant tagging to be searched easily, quickly, and accurately on the Web. In that case, a general Web search becomes a targeted academic search in a "database" of scientific publications. However, the database is actually a network of self-created, small personal databases that comply with a common tagging and search standard. Again, my point here is not to explore the details of one or another of these approaches. If scientists and other academics adopt this approach of self-archiving coupled with standardized interfaces for global, well-delimited searches, the problem of lack of access to academic publication because of their high-cost publication will be eliminated.

Other types of documents, for example, primary- and secondary-education textbooks, are in a much more rudimentary stage of the development of peer-production models. First, it should be recognized that responses to illiteracy and low educational completion in the poorer areas of the world are largely a result of lack of schoolteachers, physical infrastructure for classrooms, demand for children's schooling among parents who are themselves illiterate, and lack of effectively enforced compulsory education policy. The cost of textbooks contributes only a portion of the problem of cost. The opportunity cost of children's labor is probably the largest factor. Nonetheless, outdated materials and poor quality of teaching materials are often cited as one limit on the educational achievement of those who do attend school. The costs of books, school fees, uniforms, and stationery can amount to 20-30 percent of a family's income./9 The component of the problem contributed by the teaching materials may be alleviated by innovative approaches to textbook and education materials authoring. Chapter 4 already discussed some textbook initiatives. The most successful commons-based textbook authoring project, which is also the most relevant from the perspective of development, is the South African project, Free High School Science Texts

(FHSST). The FHSST initiative is more narrowly focused than the broader efforts of Wikibooks or the California initiative, more managed, and more successful. Nonetheless, in three years of substantial effort by a group of dedicated volunteers who administer the project, its product is one physics high school text, and advanced drafts of two other science texts. The main constraint on the efficacy of collaborative textbook authoring is that compliance requirements imposed by education ministries tend to require a great degree of coherence, which constrains the degree of modularity that these text-authoring projects adopt. The relatively large-grained contributions required limit the number of contributors, slowing the process. The future of these efforts is therefore likely to be determined by the extent to which their designers are able to find ways to make finer-grained modules without losing the coherence required for primary- and secondary-education texts. Texts at the post-secondary level likely present less of a problem, because of the greater freedom instructors have to select texts. This allows an initiative like MIT's Open Courseware Initiative to succeed. That initiative provides syllabi, lecture notes, problem sets, etc. from over 1,100 courses. The basic creators of the materials are paid academics who produce these materials for one of their core professional roles: teaching college- and graduate-level courses. The content is, by and large, a "side-effect" of teaching. What is left to be done is to integrate, create easy interfaces and search capabilities, and so forth. The university funds these functions through its own resources and dedicated grant funding. In the context of MIT, then, these functions are performed on a traditional model - a large, well-funded nonprofit provides an important public good through the application of full-time staff aimed at non-wealth-maximizing goals. The critical point here was the radical departure of MIT from the emerging culture of the 1980s and 1990s in American academia. When other universities were thinking of "distance education" in terms of selling access to taped lectures and materials so as to raise new revenue, MIT thought of what its basic mandate to advance knowledge and educate students in a networked environment entailed. The answer was to give anyone, anywhere, access to the teaching materials of some of the best minds in the world. As an intervention in the ecology of free knowledge and information and an act of leadership among universities, the MIT initiative was therefore a major event. As a model for organizational innovation in the domain of information production generally and the creation of educational resources in particular, it was less significant.

Software and academic publication, then, offer the two most advanced examples of commons-based strategies employed in a sector whose outputs are important to development, in ways that improve access to basic information, knowledge, and information-embedded tools. Building on these basic cases, we can begin to see how similar strategies can be employed to create a substantial set of commons-based solutions that could improve the distribution of information germane to human development.

Commons-Based Research for Food and Medicines

While computation and access to existing scientific research are important in the development of any nation, they still operate at a remove from the most basic needs of the world poor. On its face, it is far from obvious how the emergence of the networked information economy can grow rice to feed millions of malnourished children or deliver drugs to millions of HIV/AIDS patients. On closer observation, however, a tremendous proportion of the way modern societies grow food and develop medicines is based on scientific research and technical innovation. We have seen how the functions of mass media can be fulfilled by nonproprietary models of news and commentary. We have seen the potential of free and open source software and open-access publications to replace and redress some of the failures of proprietary software and scientific publication, respectively. These cases suggest that the basic choice between a system that depends on exclusive rights and business models that use exclusion to appropriate research outputs and a system that weaves

together various actors - public and private, organized and individual - in a nonproprietary social network of innovation, has important implications for the direction of innovation and for access to its products. Public attention has focused mostly on the HIV/AIDS crisis in Africa and the lack of access to existing drugs because of their high costs. However, that crisis is merely the tip of the iceberg. It is the most visible to many because of the presence of the disease in rich countries and its cultural and political salience in the United States and Europe. The exclusive rights system is a poor institutional mechanism for serving the needs of those who are worst off around the globe. Its weaknesses pervade the problems of food security and agricultural research aimed at increasing the supply of nourishing food throughout the developing world, and of access to medicines in general, and to medicines for developing-world diseases in particular. Each of these areas has seen a similar shift in national and international policy toward greater reliance on exclusive rights, most important of which are patents. Each area has also begun to see the emergence of commons-based models to alleviate the problems of patents. However, they differ from each other still. Agriculture offers more immediate opportunities for improvement because of the relatively larger role of public research - national, international, and academic - and of the long practices of farmer innovation in seed associations and local and regional frameworks. I explore it first in some detail, as it offers a template for what could be a path for development in medical research as well.

Food Security: Commons-Based Agricultural Innovation

Agricultural innovation over the past century has led to a vast increase in crop yields. Since the 1960s, innovation aimed at increasing yields and improving quality has been the centerpiece of efforts to secure the supply of food to the world's poor, to avoid famine and eliminate chronic malnutrition. These efforts have produced substantial increases in the production of food and decreases in its cost, but their benefits have varied widely in different regions of the world. Now, increases in productivity are not alone a sufficient condition to prevent famine. Sen's observations that democracies have no famines - that is, that good government and accountability will force public efforts to prevent famine - are widely accepted today. The contributions of the networked information economy to democratic participation and transparency are discussed in chapters 6-8, and to the extent that those chapters correctly characterize the changes in political discourse, should help alleviate human poverty through their effects on democracy. However, the cost and quality of food available to accountable governments of poor countries, or to international aid organizations or nongovernment organizations (NGOs) that step in to try to alleviate the misery caused by ineffective or malicious governments, affect how much can be done to avoid not only catastrophic famine, but also chronic malnutrition. Improvements in agriculture make it possible for anyone addressing food security to perform better than they could have if food production had lower yields, of less nutritious food, at higher prices. Despite its potential benefits, however, agricultural innovation has been subject to an unusual degree of sustained skepticism aimed at the very project of organized scientific and scientifically based innovation. Criticism combines biological-ecological concerns with social and economic concerns. Nowhere is this criticism more strident, or more successful at moving policy, than in current European resistance to genetically modified (GM) foods. The emergence of commons-based production strategies can go some way toward allaying the biological-ecological fears by locating much of the innovation at the local level. Its primary benefit, however, is likely to be in offering a path for agricultural and biological innovation that is sustainable and low cost, and that need not result in appropriation of the food production chain by a small number of multinational businesses, as many critics fear.

Scientific plant improvement in the United States dates back to the establishment of the U.S. Department of Agriculture, the land-grant universities, and later the state agricultural experiment stations during the Civil War and in the decades that followed. Public-sector investment dominated

agricultural research at the time, and with the rediscovery of Mendel's work in 1900, took a turn toward systematic selective breeding. Through crop improvement associations, seed certification programs, and open-release policies allowing anyone to breed and sell the certified new seeds, farmers were provided access to the fruits of public research in a reasonably efficient and open market. The development of hybrid corn through this system was the first major modern success that vastly increased agricultural yields. It reshaped our understanding not only of agriculture, but also more generally of the value of innovation, by comparison to efficiency, to growth. Yields in the United States doubled between the mid-1930s and the mid-1950s, and by the mid-1980s, cornfields had a yield six times greater than they had fifty years before. Beginning in the early 1960s, with funding from the Rockefeller and Ford foundations, and continuing over the following forty years, agricultural research designed to increase the supply of agricultural production and lower its cost became a central component of international and national policies aimed at securing the supply of food to the world's poor populations, avoiding famines and, ultimately, eliminating chronic malnutrition. The International Rice Research Institute (IRRI) in the Philippines was the first such institute, founded in the 1960s, followed by the International Center for Wheat and Maize Improvement (CIM-MYT) in Mexico (1966), and the two institutes for tropical agriculture in Colombia and Nigeria (1967). Together, these became the foundation for the Consultative Group for International Agricultural Research (CGIAR), which now includes sixteen centers. Over the same period, National Agricultural Research Systems (NARS) also were created around the world, focusing on research specific to local agroecological conditions. Research in these centers preceded the biotechnology revolution, and used various experimental breeding techniques to obtain high-yielding plants: for example, plants with shorter growing seasons, or more adapted to intensive fertilizer use. These efforts later introduced varieties that were resistant to local pests, diseases, and to various harsh environmental conditions.

The "Green Revolution," as the introduction of these new, scientific-research-based varieties has been called, indeed resulted in substantial increases in yields, initially in rice and wheat, in Asia and Latin America. The term "Green Revolution" is often limited to describing these changes in those regions in the 1960s and 1970s. A recent study shows, however, that the growth in yields has continued throughout the last forty years, and has, with varying degrees, occurred around the world.[10] More than eight thousand modern varieties of rice, wheat, maize, other major cereals, and root and protein crops have been released over the course of this period by more than four hundred public breeding programs. One of the most interesting finds of this study was that fewer than 1 percent of these modern varieties had any crosses with public or private breeding programs in the developed world, and that private-sector contributions in general were limited to hybrid maize, sorghum, and millet. The effort, in other words, was almost entirely public sector, and almost entirely based in the developing world, with complementary efforts of the international and national programs. Yields in Asia increased sevenfold from 1961 to 2000, and fivefold in Latin America, the Middle East/North Africa, and Sub-Saharan Africa. More than 60 percent of the growth in Asia and Latin America occurred in the 1960s-1980s, while the primary growth in Sub-Saharan Africa began in the 1980s. In Latin America, most of the early-stage increases in yields came from increasing cultivated areas (~40 percent), and from other changes in cultivation - increased use of fertilizer, mechanization, and irrigation. About 15 percent of the growth in the early period was attributable to the use of modern varieties. In the latter twenty years, however, more than 40 percent of the total increase in yields was attributable to the use of new varieties. In Asia in the early period, about 19 percent of the increase came from modern varieties, but almost the entire rest of the increase came from increased use of fertilizer, mechanization, and irrigation, not from increased cultivated areas. It is trivial to see why changes of this sort would elicit both environmental and a social-economic critique of the industrialization of farm work. Again, though, in the latter twenty years, 46 percent of the increase in yields is attributable to the use of modern

varieties. Modern varieties played a significantly less prominent role in the Green Revolution of the Middle East and Africa, contributing 5-6 percent of the growth in yields. In Sub-Saharan Africa, for example, early efforts to introduce varieties from Asia and Latin America failed, and local developments only began to be adopted in the 1980s. In the latter twenty-year period, however, the Middle East and North Africa did see a substantial role for modern varieties - accounting for close to 40 percent of a more than doubling of yields. In Sub-Saharan Africa, the overwhelming majority of the tripling of yields came from increasing area of cultivation, and about 16 percent came from modern varieties. Over the past forty years, then, research-based improvements in plants have come to play a larger role in increasing agricultural yields in the developing world. Their success was, however, more limited in the complex and very difficult environments of Sub-Saharan Africa. Much of the benefit has to do with local independence, as opposed to heavier dependence on food imports. Evenson and Gollin, for example, conservatively estimate that higher prices and a greater reliance on imports in the developing world in the absence of the Green Revolution would have resulted in 13-14 percent lower caloric intake in the developing world, and in a 6-8 percent higher proportion of malnourished children. While these numbers may not seem eye-popping, for populations already living on marginal nutrition, they represent significant differences in quality of life and in physical and mental development for millions of children and adults.

The agricultural research that went into much of the Green Revolution did not involve biotechnology - that is, manipulation of plant varieties at the genetic level through recombinant DNA techniques. Rather, it occurred at the level of experimental breeding. In the developed world, however, much of the research over the past twenty-five years has been focused on the use of biotechnology to achieve more targeted results than breeding can, has been more heavily based on private-sector investment, and has resulted in more private-sector ownership over the innovations. The promise of biotechnology, and particularly of genetically engineered or modified foods, has been that they could provide significant improvements in yields as well as in health effects, quality of the foods grown, and environmental effects. Plants engineered to be pest resistant could decrease the need to use pesticides, resulting in environmental benefits and health benefits to farmers. Plants engineered for ever-higher yields without increasing tilled acreage could limit the pressure for deforestation. Plants could be engineered to carry specific nutritional supplements, like golden rice with beta-carotene, so as to introduce necessarily nutritional requirements into subsistence diets. Beyond the hypothetically optimistic possibilities, there is little question that genetic engineering has already produced crops that lower the cost of production for farmers by increasing herbicide and pest tolerance. As of 2002, more than 50 percent of the world's soybean acreage was covered with genetically modified (GM) soybeans, and 20 percent with cotton. Twenty-seven percent of acreage covered with GM crops is in the developing world. This number will grow significantly now that Brazil has decided to permit the introduction of GM crops, given its growing agricultural role, and now that India, as the world's largest cotton producer, has approved the use of Bt cotton - a GM form of cotton that improves its resistance to a common pest. There are, then, substantial advantages to farmers, at least, and widespread adoption of GM crops both in the developed world outside of Europe and in the developing world.

This largely benign story of increasing yields, resistance, and quality has not been without critics, to put it mildly. The criticism predates biotechnology and the development of transgenic varieties. Its roots are in criticism of experimental breeding programs of the American agricultural sectors and the Green Revolution. However, the greatest public visibility and political success of these criticisms has been in the context of GM foods. The critique brings together odd intellectual and political bedfellows, because it includes five distinct components: social and economic critique of the industrialization of agriculture, environmental and health effects, consumer preference for "natural"

or artisan production of foodstuffs, and, perhaps to a more limited extent, protectionism of domestic farm sectors.

Perhaps the oldest component of the critique is the social-economic critique. One arm of the critique focuses on how mechanization, increased use of chemicals, and ultimately the use of nonreproducing proprietary seed led to incorporation of the agricultural sector into the capitalist form of production. In the United States, even with its large "family farm" sector, purchased inputs now greatly exceed nonpurchased inputs, production is highly capital intensive, and large-scale production accounts for the majority of land tilled and the majority of revenue captured from farming./11 In 2003, 56 percent of farms had sales of less than $10,000 a year. Roughly 85 percent of farms had less than $100,000 in sales./12 These farms account for only 42 percent of the farmland. By comparison, 3.4 percent of farms have sales of more than $500,000 a year, and account for more than 21 percent of land. In the aggregate, the 7.5 percent of farms with sales over $250,000 account for 37 percent of land cultivated. Of all principal owners of farms in the United States in 2002, 42.5 percent reported something other than farming as their principal occupation, and many reported spending two hundred or more days off-farm, or even no work days at all on the farm. The growth of large-scale "agribusiness," that is, mechanized, rationalized industrial-scale production of agricultural products, and more important, of agricultural inputs, is seen as replacing the family farm and the small-scale, self-sufficient farm, and bringing farm labor into the capitalist mode of production. As scientific development of seeds and chemical applications increases, the seed as input becomes separated from the grain as output, making farmers dependent on the purchase of industrially produced seed. This further removes farmwork from traditional modes of self-sufficiency and craftlike production to an industrial mode. This basic dynamic is repeated in the critique of the Green Revolution, with the added overlay that the industrial producers of seed are seen to be multinational corporations, and the industrialization of agriculture is seen as creating dependencies in the periphery on the industrial-scientific core of the global economy.

The social-economic critique has been enmeshed, as a political matter, with environmental, health, and consumer-oriented critiques as well. The environmental critiques focus on describing the products of science as monocultures, which, lacking the genetic diversity of locally used varieties, are more susceptible to catastrophic failure. Critics also fear contamination of existing varieties, unpredictable interactions with pests, and negative effects on indigenous species. The health effects concern focused initially on how breeding for yield may have decreased nutritional content, and in the more recent GM food debates, the concern that genetically altered foods will have some unanticipated negative health reactions that would only become apparent many years from now. The consumer concerns have to do with quality and an aesthetic attraction to artisan-mode agricultural products and aversion to eating industrial outputs. These social-economic and environmental-health-consumer concerns tend also to be aligned with protectionist lobbies, not only for economic purposes, but also reflecting a strong cultural attachment to the farming landscape and human ecology, particularly in Europe.

This combination of social-economic and postcolonial critique, environmentalism, public-health concerns, consumer advocacy, and farm-sector protectionism against the relatively industrialized American agricultural sector reached a height of success in the 1999 five-year ban imposed by the European Union on all GM food sales. A recent study of a governmental Science Review Board in the United Kingdom, however, found that there was no evidence for any of the environmental or health critiques of GM foods./13 Indeed, as Peter Pringle masterfully chronicled in *Food, Inc.*, both sides of the political debate could be described as having buffed their cases significantly. The successes and potential benefits have undoubtedly been overstated by enamored scientists and avaricious vendors. There is little doubt, too, that the near-hysterical pitch at which the failures and

risks of GM foods have been trumpeted has little science to back it, and the debate has degenerated to a state that makes reasoned, evidence-based consideration difficult. In Europe in general, however, there is wide acceptance of what is called a "precautionary principle." One way of putting it is that absence of evidence of harm is not evidence of absence of harm, and caution counsels against adoption of the new and at least theoretically dangerous. It was this precautionary principle rather than evidence of harm that was at the base of the European ban. This ban has recently been lifted, in the wake of a WTO trade dispute with the United States and other major producers who challenged the ban as a trade barrier. However, the European Union retained strict labeling requirements. This battle among wealthy countries, between the conservative "Fortress Europe" mentality and the growing reliance of American agriculture on biotechnological innovation, would have little moral valence if it did not affect funding for, and availability of, biotechnological research for the populations of the developing world. Partly as a consequence of the strong European resistance to GM foods, the international agricultural research centers that led the way in the development of the Green Revolution varieties, and that released their developments freely for anyone to sell and use without proprietary constraint, were slow to develop capacity in genetic engineering and biotechnological research more generally. Rather than the public national and international efforts leading the way, a study of GM use in developing nations concluded that practically all GM acreage is sown with seed obtained in the finished form from a developed-world supplier, for a price premium or technology licensing fee./14 The seed, and its improvements, is proprietary to the vendor in this model. It is not supplied in a form or with the rights to further improve locally and independently. Because of the critique of innovation in agriculture as part of the process of globalization and industrialization, of environmental degradation, and of consumer exploitation, the political forces that would have been most likely to support public-sector investment in agricultural innovation are in opposition to such investments. The result has not been retardation of biotechnological innovation in agriculture, but its increasing privatization: primarily in the United States and now increasingly in Latin America, whose role in global agricultural production is growing.

Private-sector investment, in turn, operates within a system of patents and other breeders' exclusive rights, whose general theoretical limitations are discussed in chapter 2. In agriculture, this has two distinct but mutually reinforcing implications. The first is that, while private-sector innovation has indeed accounted for most genetically engineered crops in the developing world, research aimed at improving agricultural production in the neediest places has not been significantly pursued by the major private-sector firms. A sector based on expectation of sales of products embedding its patents will not focus its research where human welfare will be most enhanced. It will focus where human welfare can best be expressed in monetary terms. The poor are systematically underserved by such a system. It is intended to elicit investments in research in directions that investors believe will result in outputs that serve the needs of those with the highest willingness and ability to pay for their outputs. The second is that even where the products of innovation can, as a matter of biological characteristics, be taken as inputs into local research and development - by farmers or by national agricultural research systems - the international system of patents and plant breeders' rights enforcement makes it illegal to do so without a license. This again retards the ability of poor countries and their farmers and research institutes to conduct research into local adaptations of improved crops.

The central question raised by the increasing privatization of agricultural biotechnology over the past twenty years is: What can be done to employ commons-based strategies to provide a foundation for research that will be focused on the food security of developing world populations? Is there a way of managing innovation in this sector so that it will not be heavily weighted in favor of populations with a higher ability to pay, and so that its outputs allow farmers and national

research efforts to improve and adapt to highly variable local agroecological environments? The continued presence of the public-sector research infrastructure - including the international and national research centers, universities, and NGOs dedicated to the problem of food security - and the potential of harnessing individual farmers and scientists to cooperative development of open biological innovation for agriculture suggest that commons-based paths for development in the area of food security and agricultural innovation are indeed feasible.

First, some of the largest and most rapidly developing nations that still have large poor populations - most prominently, China, India, and Brazil - can achieve significant advances through their own national agricultural research systems. Their research can, in turn, provide a platform for further innovation and adaptation by projects in poorer national systems, as well as in nongovernmental public and peer-production efforts. In this regard, China seems to be leading the way. The first rice genome to be sequenced was japonica, apparently sequenced in 2000 by scientists at Monsanto, but not published. The second, an independent and published sequence of japonica, was sequenced by scientists at Syngenta, and published as the first published rice genome sequence in *Science* in April 2002. To protect its proprietary interests, Syngenta entered a special agreement with *Science*, which permitted the authors not to deposit the genomic information into the public Genbank maintained by the National Institutes of Health in the United States.[/15](#) Depositing the information in GenBank makes it immediately available for other scientists to work with freely. All the major scientific publications require that such information be deposited and made publicly available as a standard condition of publication, but *Science* waved this requirement for the Syngenta japonica sequence. The same issue of *Science*, however, carried a similar publication, the sequence of Oryza sativa L.ssp. indica, the most widely cultivated subspecies in China. This was sequenced by a public Chinese effort, and its outputs were immediately deposited in GenBank. The simultaneous publication of the rice genome by a major private firm and a Chinese public effort was the first public exposure to the enormous advances that China's public sector has made in agricultural biotechnology, and its focus first and foremost on improving Chinese agriculture. While its investments are still an order of magnitude smaller than those of public and private sectors in the developed countries, China has been reported as the source of more than half of all expenditures in the developing world.[/16](#) China's longest experience with GM agriculture is with Bt cotton, which was introduced in 1997. By 2000, 20 percent of China's cotton acreage was sown to Bt cotton. One study showed that the average acreage of a farm was less than 0.5 hectare of cotton, and the trait that was most valuable to them was Bt cotton's reduced pesticide needs. Those who adopted Bt cotton used less pesticide, reducing labor for pest control and the pesticide cost per kilogram of cotton produced. This allowed an average cost savings of 28 percent. Another effect suggested by survey data - which, if confirmed over time, would be very important as a matter of public health, but also to the political economy of the agricultural biotechnology debate - is that farmers who do not use Bt cotton are four times as likely to report symptoms of a degree of toxic exposure following application of pesticides than farmers who did adopt Bt cotton.[/17](#) The point is not, of course, to sing the praises of GM cotton or the Chinese research system. China's efforts offer an example of how the larger national research systems can provide an anchor for agricultural research, providing solutions both for their own populations, and, by making the products of their research publicly and freely available, offer a foundation for the work of others.

Alongside the national efforts in developing nations, there are two major paths for commons-based research and development in agriculture that could serve the developing world more generally. The first is based on existing research institutes and programs cooperating to build a commons-based system, cleared of the barriers of patents and breeders' rights, outside and alongside the proprietary system. The second is based on the kind of loose affiliation of university scientists, nongovernmental organizations, and individuals that we saw play such a significant role in the

development of free and open-source software. The most promising current efforts in the former vein are the PIPRA (Public Intellectual Property for Agriculture) coalition of public-sector universities in the United States, and, if it delivers on its theoretical promises, the Generation Challenge Program led by CGIAR (the Consultative Group on International Agricultural Research). The most promising model of the latter, and probably the most ambitious commons-based project for biological innovation currently contemplated, is BIOS (Biological Innovation for an Open Society).

PIPRA is a collaboration effort among public-sector universities and agricultural research institutes in the United States, aimed at managing their rights portfolio in a way that will give their own and other researchers freedom to operate in an institutional ecology increasingly populated by patents and other rights that make work difficult. The basic thesis and underlying problem that led to PIPRA's founding were expressed in an article in *Science* coauthored by fourteen university presidents.[/18](#) They underscored the centrality of public-sector, land-grant university-based research to American agriculture, and the shift over the last twenty-five years toward increased use of intellectual property rules to cover basic discoveries and tools necessary for agricultural innovation. These strategies have been adopted by both commercial firms and, increasingly, by public-sector universities as the primary mechanism for technology transfer from the scientific institute to the commercializing firms. The problem they saw was that in agricultural research, innovation was incremental. It relies on access to existing germplasm and crop varieties that, with each generation of innovation, brought with them an ever-increasing set of intellectual property claims that had to be licensed in order to obtain permission to innovate further. The universities decided to use the power that ownership over roughly 24 percent of the patents in agricultural biotechnology innovations provides them as a lever with which to unravel the patent thickets and to reduce the barriers to research that they increasingly found themselves dealing with. The main story, one might say the "founding myth" of PIPRA, was the story of golden rice. Golden rice is a variety of rice that was engineered to provide dietary vitamin A. It was developed with the hope that it could introduce vitamin A supplement to populations in which vitamin A deficiency causes roughly 500,000 cases of blindness a year and contributes to more than 2 million deaths a year. However, when it came to translating the research into deliverable plants, the developers encountered more than seventy patents in a number of countries and six materials transfer agreements that restricted the work and delayed it substantially. PIPRA was launched as an effort of public-sector universities to cooperate in achieving two core goals that would respond to this type of barrier-preserving the right to pursue applications to subsistence crops and other developing-world-related crops, and preserving their own freedom to operate vis-à-vis each other's patent portfolios.

The basic insight of PIPRA, which can serve as a model for university alliances in the context of the development of medicines as well as agriculture, is that universities are not profit-seeking enterprises, and university scientists are not primarily driven by a profit motive. In a system that offers opportunities for academic and business tracks for people with similar basic skills, academia tends to attract those who are more driven by nonmonetary motivations. While universities have invested a good deal of time and money since the Bayh-Dole Act of 1980 permitted and indeed encouraged them to patent innovations developed with public funding, patent and other exclusive-rights-based revenues have not generally emerged as an important part of the revenue scheme of universities.

Table 9.2: Selected University Gross Revenues and Patent Licensing Revenues

	Total Revenues	Licensing and Royalties		Government Grants & Contracts	
	(millions $)	(millions $)	% of total	(millions $)	% of total
All universities	$227,000	$1,270	0.56%	$31,430	13.85%
University of California	$14,166	$81.3 $55 (net)/b	0.57% 0.39%	$2,372	16.74%
Stanford University	$3,475	$43.3 $36.8/c	1.25% 1.06%	$860	24.75%
Florida State	$2,646	$35.6	1.35%	$238	8.99%
Harvard	$2,473	$47.9	1.94%	$416 $548.7/d	16.82% 22.19%
Columbia University	$2,074	$178.4 $100-120/a	8.6% 4.9-5.9%	$532	25.65%
University of Wisconsin-Madison	$1,696	$32	1.89%	$417.4	24.61%
University of Minnesota	$1,237	$38.7	3.12%	$323.5	26.15%
Cal Tech	$531	$26.7/e $15.7/f	5.02% 2.95%	$268	50.47%

Sources: Aggregate revenues: U.S. Dept. of Education, National Center for Education Statistics, *Enrollment in Postsecondary Institutions, Fall 2001, and Financial Statistics, Fiscal Year 2001* (2003), Table F; Association of University Technology Management, *Annual Survey Summary FY 2002* (AUTM 2003), Table S-12. Individual institutions: publicly available annual reports of each university and/or its technology transfer office for FY 2003.

Notes:

1. Large ambiguity results because technology transfer office reports increased revenues for year-end 2003 as $178M without reporting expenses; University Annual Report reports licensing revenue with all "revenue from other educational and research activities," and reports a 10 percent decline in this category, "reflecting an anticipated decline in royalty and license income" from the $133M for the previous year-end, 2002. The table reflects an assumed net contribution to university revenues between $100-120M (the entire decline in the category due to royalty/royalties decreased proportionately with the category).

2.University of California Annual Report of the Office of Technology Transfer is more transparent than most in providing expenses - both net legal expenses and tech transfer direct operating expenses, which allows a clear separation of net revenues from technology transfer activities.

3.Minus direct expenses, not including expenses for unlicensed inventions.

4.Federal- and nonfederal-sponsored research.

5.Almost half of this amount is in income from a single Initial Public Offering, and therefore does not represent a recurring source of licensing revenue.

6.Technology transfer gross revenue minus the one-time event of an initial public offering of LiquidMetal Technologies.

As table 9.2 shows, except for one or two outliers, patent revenues have been all but negligible in university budgets.[/19] This fact makes it fiscally feasible for universities to use their patent portfolios to maximize the global social benefit of their research, rather than trying to maximize patent revenue. In particular, universities can aim to include provisions in their technology licensing agreements that are aimed at the dual goals of (a) delivering products embedding their innovations to developing nations at reasonable prices and (b) providing researchers and plant breeders the freedom to operate that would allow them to research, develop, and ultimately produce crops that would improve food security in the developing world.

While PIPRA shows an avenue for collaboration among universities in the public interest, it is an avenue that does not specifically rely on, or benefit in great measure from, the information networks or the networked information economy. It continues to rely on the traditional model of publicly funded research. More explicit in its effort to leverage the cost savings made possible by networked information systems is the Generation Challenge Program (GCP). The GCP is an effort to bring the CGIAR into the biotechnology sphere, carefully, given the political resistance to genetically modified foods, and quickly, given the already relatively late start that the international research centers have had in this area. Its stated emphasis is on building an architecture of innovation, or network of research relationships, that will provide low-cost techniques for the basic contemporary technologies of agricultural research. The program has five primary foci, but the basic thrust is to generate improvements both in basic genomics science and in breeding and farmer education, in both cases for developing world agriculture. One early focus would be on building a communications system that allows participating institutions and scientists to move information efficiently and utilize computational resources to pursue research. There are hundreds of thousands of samples of germplasm, from "landrace" (that is, locally agriculturally developed) and wild varieties to modern varieties, located in databases around the world in international, national, and academic institutions. There are tremendous high-capacity computation resources in some of the most advanced research institutes, but not in many of the national and international programs. One of the major goals articulated for the GCP is to develop Web-based interfaces to share these data and computational resources. Another is to provide a platform for sharing new questions and directions of research among participants. The work in this network will, in turn, rely on materials that have proprietary interests attached to them, and will produce outputs that could have proprietary interests attached to them as well. Just like the universities, the GCP institutes (national, international, and nonprofit) are looking for an approach aimed to secure open access to research

materials and tools and to provide humanitarian access to its products, particularly for subsistence crop development and use. As of this writing, however, the GCP is still in a formative stage, more an aspiration than a working model. Whether it will succeed in overcoming the political constraints placed on the CGIAR as well as the relative latecomer status of the international public efforts to this area of work remains to be seen. But the elements of the GCP certainly exhibit an understanding of the possibilities presented by commons-based networked collaboration, and an ambition to both build upon them and contribute to their development.

The most ambitious effort to create a commons-based framework for biological innovation in this field is BIOS. BIOS is an initiative of CAMBIA (Center for the Application of Molecular Biology to International Agriculture), a nonprofit agricultural research institute based in Australia, which was founded and is directed by Richard Jefferson, a pioneer in plant biotechnology. BIOS is based on the observation that much of contemporary agricultural research depends on access to tools and enabling technologies - such as mechanisms to identify genes or for transferring them into target plants. When these tools are appropriated by a small number of firms and available only as part of capital-intensive production techniques, they cannot serve as the basis for innovation at the local level or for research organized on nonproprietary models. One of the core insights driving the BIOS initiative is the recognition that when a subset of necessary tools is available in the public domain, but other critical tools are not, the owners of those tools appropriate the full benefits of public domain innovation without at the same time changing the basic structural barriers to use of the proprietary technology. To overcome these problems, the BIOS initiative includes both a strong informatics component and a fairly ambitious "copyleft"-like model (similar to the GPL described in chapter 3) of licensing CAMBIA's basic tools and those of other members of the BIOS initiative. The informatics component builds on a patent database that has been developed by CAMBIA for a number of years, and whose ambition is to provide as complete as possible a dataset of who owns what tools, what the contours of ownership are, and by implication, who needs to be negotiated with and where research paths might emerge that are not yet appropriated and therefore may be open to unrestricted innovation.

The licensing or pooling component is more proactive, and is likely the most significant of the project. BIOS is setting up a licensing and pooling arrangement, "primed" by CAMBIA's own significant innovations in tools, which are licensed to all of the initiative's participants on a free model, with grant-back provisions that perform an openness-binding function similar to copyleft.[/20] In coarse terms, this means that anyone who builds upon the contributions of others must contribute improvements back to the other participants. One aspect of this model is that it does not assume that all research comes from academic institutions or from traditional government-funded, nongovernmental, or intergovernmental research institutes. It tries to create a framework that, like the open-source development community, engages commercial and noncommercial, public and private, organized and individual participants into a cooperative research network. The platform for this collaboration is "BioForge," styled after Sourceforge, one of the major free and open-source software development platforms. The commitment to engage many different innovators is most clearly seen in the efforts of BIOS to include major international commercial providers and local potential commercial breeders alongside the more likely targets of a commons-based initiative. Central to this move is the belief that in agricultural science, the basic tools can, although this may be hard, be separated from specific applications or products. All actors, including the commercial ones, therefore have an interest in the open and efficient development of tools, leaving competition and profit making for the market in applications. At the other end of the spectrum, BIOS's focus on making tools freely available is built on the proposition that innovation for food security involves more than biotechnology alone. It involves environmental management, locale-specific adaptations, and social and economic adoption in forms that are locally and internally

sustainable, as opposed to dependent on a constant inflow of commoditized seed and other inputs. The range of participants is, then, much wider than envisioned by PIPRA or the GCP. It ranges from multinational corporations through academic scientists, to farmers and local associations, pooling their efforts in a communications platform and institutional model that is very similar to the way in which the GNU/Linux operating system has been developed. As of this writing, the BIOS project is still in its early infancy, and cannot be evaluated by its outputs. However, its structure offers the crispest example of the extent to which the peer-production model in particular, and commons-based production more generally, can be transposed into other areas of innovation at the very heart of what makes for human development - the ability to feed oneself adequately.

PIPRA and the BIOS initiative are the most salient examples of, and the most significant first steps in the development of commons-based strategies to achieve food security. Their vitality and necessity challenge the conventional wisdom that ever-increasing intellectual property rights are necessary to secure greater investment in research, or that the adoption of proprietary rights is benign. Increasing appropriation of basic tools and enabling technologies creates barriers to entry for innovators - public-sector, nonprofit organizations, and the local farmers themselves - concerned with feeding those who cannot signal with their dollars that they are in need. The emergence of commons-based techniques - particularly, of an open innovation platform that can incorporate farmers and local agronomists from around the world into the development and feedback process through networked collaboration platforms - promises the most likely avenue to achieve research oriented toward increased food security in the developing world. It promises a mechanism of development that will not increase the relative weight and control of a small number of commercial firms that specialize in agricultural production. It will instead release the products of innovation into a self-binding commons - one that is institutionally designed to defend itself against appropriation. It promises an iterative collaboration platform that would be able to collect environmental and local feedback in the way that a free software development project collects bug reports - through a continuous process of networked conversation among the user-innovators themselves. In combination with public investments from national governments in the developing world, from the developed world, and from more traditional international research centers, agricultural research for food security may be on a path of development toward constructing a sustainable commons-based innovation ecology alongside the proprietary system. Whether it follows this path will be partly a function of the engagement of the actors themselves, but partly a function of the extent to which the international intellectual property/trade system will refrain from raising obstacles to the emergence of these commons-based efforts.

Access to Medicines: Commons-Based Strategies for Biomedical Research

Nothing has played a more important role in exposing the systematic problems that the international trade and patent system presents for human development than access to medicines for HIV/AIDS. This is so for a number of reasons. First, HIV/AIDS has reached pandemic proportions. One quarter of all deaths from infectious and parasitic diseases in 2002 were caused by AIDS, accounting for almost 5 percent of all deaths in the world that year.[/21] Second, it is a new condition, unknown to medicine a mere twenty-five years ago, is communicable, and in principle is of a type - infectious diseases - that we have come to see modern medicine as capable of solving. This makes it different from much bigger killers - like the many cancers and forms of heart disease - which account for about nine times as many deaths globally. Third, it has a significant presence in the advanced economies. Because it was perceived there as a disease primarily affecting the gay community, it had a strong and well-defined political lobby and high cultural salience. Fourth, and finally, there have indeed been enormous advances in the development of medicines for

HIV/AIDS. Mortality for patients who are treated is therefore much lower than for those who are not. These treatments are new, under patent, and enormously expensive. As a result, death - as opposed to chronic illness - has become overwhelmingly a consequence of poverty. More than 75 percent of deaths caused by AIDS in 2002 were in Africa. HIV/AIDS drugs offer a vivid example of an instance where drugs exist for a disease but cannot be afforded in the poorest countries. They represent, however, only a part, and perhaps the smaller part, of the limitations that a patent-based drug development system presents for providing medicines to the poor. No less important is the absence of a market pull for drugs aimed at diseases that are solely or primarily developing-world diseases - like drugs for tropical diseases, or the still-elusive malaria vaccine.

To the extent that the United States and Europe are creating a global innovation system that relies on patents and market incentives as its primary driver of research and innovation, these wealthy democracies are, of necessity, choosing to neglect diseases that disproportionately affect the poor. There is nothing evil about a pharmaceutical company that is responsible to its shareholders deciding to invest where it expects to reap profit. It is not immoral for a firm to invest its research funds in finding a drug to treat acne, which might affect 20 million teenagers in the United States, rather than a drug that will cure African sleeping sickness, which affects 66 million Africans and kills about fifty thousand every year. If there is immorality to be found, it is in the legal and policy system that relies heavily on the patent system to induce drug discovery and development, and does not adequately fund and organize biomedical research to solve the problems that cannot be solved by relying solely on market pull. However, the politics of public response to patents for drugs are similar in structure to those that have to do with agricultural biotechnology exclusive rights. There is a very strong patent-based industry - much stronger than in any other patent-sensitive area. The rents from strong patents are enormous, and a rational monopolist will pay up to the value of its rents to maintain and improve its monopoly. The primary potential political push-back in the pharmaceutical area, which does not exist in the agricultural innovation area, is that the exorbitant costs of drugs developed under this system is hurting even the well-endowed purses of developed-world populations. The policy battles in the United States and throughout the developed world around drug cost containment may yet result in a sufficient loosening of the patent constraints to deliver positive side effects for the developing world. However, they may also work in the opposite direction. The unwillingness of the wealthy populations in the developed world to pay high rents for drugs retards the most immediate path to lower-cost drugs in the developing world - simple subsidy of below-cost sales in poor countries cross-subsidized by above-cost rents in wealthy countries.

The industrial structure of biomedical research and pharmaceutical development is different from that of agricultural science in ways that still leave a substantial potential role for commons-based strategies. However, these would be differently organized and aligned than in agriculture. First, while governments play an enormous role in funding basic biomedical science, there are no real equivalents of the national and international agricultural research institutes. In other words, there are few public-sector laboratories that actually produce finished drugs for delivery in the developing world, on the model of the International Rice Research Institute or one of the national agricultural research systems. On the other hand, there is a thriving generics industry, based in both advanced and developing economies, that stands ready to produce drugs once these are researched. The primary constraint on harnessing its capacity for low-cost drug production and delivery for poorer nations is the international intellectual property system. The other major difference is that, unlike with software, scientific publication, or farmers in agriculture, there is no existing framework for individuals to participate in research and development on drugs and treatments. The primary potential source of nongovernmental investment of effort and thought into biomedical research

and development are universities as institutions and scientists, if they choose to organize themselves into effective peer-production communities.

Universities and scientists have two complementary paths open to them to pursue commons-based strategies to provide improved research on the relatively neglected diseases of the poor and improved access to existing drugs that are available in the developed world but unaffordable in the developing. The first involves leveraging existing university patent portfolios - much as the universities allied in PIPRA are exploring and as CAMBIA is doing more aggressively. The second involves work in an entirely new model - constructing collaboration platforms to allow scientists to engage in peer production, cross-cutting the traditional grant-funded lab, and aiming toward research into diseases that do not exercise a market pull on the biomedical research system in the advanced economies.

Leveraging University Patents. In February 2001, the humanitarian organization Doctors Without Borders (also known as Médecins Sans Frontières, or MSF) asked Yale University, which held the key South African patent on stavudine - one of the drugs then most commonly used in combination therapies - for permission to use generic versions in a pilot AIDS treatment program. At the time, the licensed version of the drug, sold by Bristol-Myers-Squibb (BMS), cost $1,600 per patient per year. A generic version, manufactured in India, was available for $47 per patient per year. At that point in history, thirty-nine drug manufacturers were suing the South African government to strike down a law permitting importation of generics in a health crisis, and no drug company had yet made concessions on pricing in developing nations. Within weeks of receiving MSF's request, Yale negotiated with BMS to secure the sale of stavudine for fifty-five dollars a year in South Africa. Yale, the University of California at Berkeley, and other universities have, in the years since, entered into similar ad hoc agreements with regard to developing-world applications or distribution of drugs that depend on their patented technologies. These successes provide a template for a much broader realignment of how universities use their patent portfolios to alleviate the problems of access to medicines in developing nations.

We have already seen in table 9.2 that while universities own a substantial and increasing number of patents, they do not fiscally depend in any significant way on patent revenue. These play a very small part in the overall scheme of revenues. This makes it practical for universities to reconsider how they use their patents and to reorient toward using them to maximize their beneficial effects on equitable access to pharmaceuticals developed in the advanced economies. Two distinct moves are necessary to harness publicly funded university research toward building an information commons that is easily accessible for global redistribution. The first is internal to the university process itself. The second has to do with the interface between the university and patent-dependent and similar exclusive-rights-dependent market actors.

Universities are internally conflicted about their public and market goals. Dating back to the passage of the Bayh-Dole Act, universities have increased their patenting practices for the products of publicly funded research. Technology transfer offices that have been set up to facilitate this practice are, in many cases, measured by the number of patent applications, grants, and dollars they bring in to the university. These metrics for measuring the success of these offices tend to make them function, and understand their role, in a way that is parallel to exclusive-rights-dependent market actors, instead of as public-sector, publicly funded, and publicly minded institutions. A technology transfer officer who has successfully provided a royalty-free license to a nonprofit concerned with developing nations has no obvious metric in which to record and report the magnitude of her success (saving X millions of lives or displacing Y misery), unlike her colleague who can readily report X millions of dollars from a market-oriented license, or even merely Y dozens of patents

filed. Universities must consider more explicitly their special role in the global information and knowledge production system. If they recommit to a role focused on serving the improvement of the lot of humanity, rather than maximization of their revenue stream, they should adapt their patenting and licensing practices appropriately. In particular, it will be important following such a rededication to redefine the role of technology transfer offices in terms of lives saved, quality-of-life measures improved, or similar substantive measures that reflect the mission of university research, rather than the present metrics borrowed from the very different world of patent-dependent market production. While the internal process is culturally and politically difficult, it is not, in fact, analytically or technically complex. Universities have, for a very long time, seen themselves primarily as dedicated to the advancement of knowledge and human welfare through basic research, reasoned inquiry, and education. The long-standing social traditions of science have always stood apart from market incentives and orientations. The problem is therefore one of reawakening slightly dormant cultural norms and understandings, rather than creating new ones in the teeth of long-standing contrary traditions. The problem should be substantially simpler than, say, persuading companies that traditionally thought of their innovation in terms of patents granted or royalties claimed, as some technology industry participants have, to adopt free software strategies.

If universities do make the change, then the more complex problem will remain: designing an institutional interface between universities and the pharmaceutical industry that will provide sustainable significant benefits for developing-world distribution of drugs and for research opportunities into developing-world diseases. As we already saw in the context of agriculture, patents create two discrete kinds of barriers: The first is on distribution, because of the monopoly pricing power they purposefully confer on their owners. The second is on research that requires access to tools, enabling technologies, data, and materials generated by the developed-world research process, and that could be useful to research on developing-world diseases. Universities working alone will not provide access to drugs. While universities perform more than half of the basic scientific research in the United States, this effort means that more than 93 percent of university research expenditures go to basic and applied science, leaving less than 7 percent for development - the final research necessary to convert a scientific project into a usable product.[/22] Universities therefore cannot simply release their own patents and expect treatments based on their technologies to become accessible. Instead, a change is necessary in licensing practices that takes an approach similar to a synthesis of the general public license (GPL), of BIOS's licensing approach, and PIPRA.

Universities working together can cooperate to include in their licenses provisions that would secure freedom to operate for anyone conducting research into developing-world diseases or production for distribution in poorer nations. The institutional details of such a licensing regime are relatively complex and arcane, but efforts are, in fact, under way to develop such licenses and to have them adopted by universities.[/23] What is important here, for understanding the potential, is the basic idea and framework. In exchange for access to the university's patents, the pharmaceutical licensees will agree not to assert any of their own rights in drugs that require a university license against generics manufacturers who make generic versions of those drugs purely for distribution in low- and middle-income countries. An Indian or American generics manufacturer could produce patented drugs that relied on university patents and were licensed under this kind of an equitable-access license, as long as it distributed its products solely in poor countries. A government or nonprofit research institute operating in South Africa could work with patented research tools without concern that doing so would violate the patents. However, neither could then import the products of their production or research into the developed world without violating the patents of both the university and the drug company. The licenses would create a mechanism for redistribution of drug products and research tools from the developed economies to the developing. It would do so

without requiring the kind of regulatory changes advocated by others, such as Jean Lanjouw, who have advocated policy changes aimed similarly to achieve differential pricing in the developing and developed worlds. Because this redistribution could be achieved by universities acting through licensing, instead of through changes in law, it offers a more feasible political path for achieving the desired result. Such action by universities would, of course, not solve all the problems of access to medicines. First, not all health-related products are based on university research. Second, patents do not account for all, or perhaps even most, of the reason that patients in poor nations are not treated. A lack of delivery infrastructure, public-health monitoring and care, and stable conditions to implement disease-control policy likely weigh more heavily. Nonetheless, there are successful and stable government and nonprofit programs that could treat hundreds of thousands or millions of patients more than they do now, if the cost of drugs were lower. Achieving improved access for those patients seems a goal worthy of pursuit, even if it is no magic bullet to solve all the illnesses of poverty.

Nonprofit Research. Even a successful campaign to change the licensing practices of universities in order to achieve inexpensive access to the products of pharmaceutical research would leave the problem of research into diseases that affect primarily the poor. This is because, unless universities themselves undertake the development process, the patent-based pharmaceuticals have no reason to. The "simple" answer to this problem is more funding from the public sector or foundations for both basic research and development. This avenue has made some progress, and some foundations - particularly, in recent years, the Gates Foundation - have invested enormous amounts of money in searching for cures and improving basic public-health conditions of disease in Africa and elsewhere in the developing world. It has received a particularly interesting boost since 2000, with the founding of the Institute for One World Health, a nonprofit pharmaceutical dedicated to research and development specifically into developing-world diseases. The basic model of One World Health begins by taking contributions of drug leads that are deemed unprofitable by the pharmaceutical industry - from both universities and pharmaceutical companies. The firms have no reason not to contribute their patents on leads purely for purposes they do not intend to pursue. The group then relies on foundation and public-sector funding to perform synthesis, preclinical and clinical trials, in collaboration with research centers in the United States, India, Bangladesh, and Thailand, and when the time comes around for manufacturing, the institute collaborates with manufacturers in developing nations to produce low-cost instances of the drugs, and with government and NGO public-health providers to organize distribution. This model is new, and has not yet had enough time to mature and provide measurable success. However, it is promising.

Peer Production of Drug Research and Development. Scientists, scientists-in-training, and to some extent, nonscientists can complement university licensing practices and formally organized nonprofit efforts as a third component of the ecology of commons-based producers. The initial response to the notion that peer production can be used for drug development is that the process is too complex, expensive, and time consuming to succumb to commons-based strategies. This may, at the end of the day, prove true. However, this was also thought of complex software projects or of supercomputing, until free software and distributed computing projects like SETI@Home and Folding@Home came along and proved them wrong. The basic point is to see how distributed nonmarket efforts are organized, and to see how the scientific production process can be broken up to fit a peer-production model.

First, anything that can be done through computer modeling or data analysis can, in principle, be done on a peer-production basis. Increasing portions of biomedical research are done today through modeling, computer simulation, and data analysis of the large and growing databases, including a wide range of genetic, chemical, and biological information. As more of the process of

drug discovery of potential leads can be done by modeling and computational analysis, more can be organized for peer production. The relevant model here is open bioinformatics. Bioinformatics generally is the practice of pursuing solutions to biological questions using mathematics and information technology. Open bioinformatics is a movement within bioinformatics aimed at developing the tools in an open-source model, and in providing access to the tools and the outputs on a free and open basis. Projects like these include the Ensmbl Genome Browser, operated by the European Bioinformatics Institute and the Sanger Centre, or the National Center for Biotechnology Information (NCBI), both of which use computer databases to provide access to data and to run various searches on combinations, patterns, and so forth, in the data. In both cases, access to the data and the value-adding functionalities are free. The software too is developed on a free software model. These, in turn, are complemented by database policies like those of the International HapMap Project, an effort to map common variations in the human genome, whose participants have committed to releasing all the data they collect freely into the public domain. The economics of this portion of research into drugs are very similar to the economics of software and computation. The models are just software. Some models will be able to run on the ever-more-powerful basic machines that the scientists themselves use. However, anything that requires serious computation could be modeled for distributed computing. This would allow projects to harness volunteer computation resources, like Folding@Home, Genome@Home, or FightAIDS@Home - sites that already harness the computing power of hundreds of thousands of users to attack biomedical science questions. This stage of the process is the one that most directly can be translated into a peer-production model, and, in fact, there have been proposals, such as the Tropical Disease Initiative proposed by Maurer, Sali, and Rai./25

Second, and more complex, is the problem of building wet-lab science on a peer-production basis. Some efforts would have to focus on the basic science. Some might be at the phase of optimization and chemical synthesis. Some, even more ambitiously, would be at the stage of preclinical animal trials and even clinical trials. The wet lab seems to present an insurmountable obstacle for a serious role for peer production in biomedical science. Nevertheless, it is not clear that it is actually any more so than it might have seemed for the development of an operating system, or a supercomputer, before these were achieved. Laboratories have two immensely valuable resources that may be capable of being harnessed to peer production. Most important by far are postdoctoral fellows. These are the same characters who populate so many free software projects, only geeks of a different feather. They are at a similar life stage. They have the same hectic, overworked lives, and yet the same capacity to work one more hour on something else, something interesting, exciting, or career enhancing, like a special grant announced by the government. The other resources that have overcapacity might be thought of as petri dishes, or if that sounds too quaint and old-fashioned, polymerase chain reaction (PCR) machines or electrophoresis equipment. The point is simple. Laboratory funding currently is silo-based. Each lab is usually funded to have all the equipment it needs for run-of-the-mill work, except for very large machines operated on time-share principles. Those machines that are redundantly provisioned in laboratories have downtime. That downtime coupled with a postdoctoral fellow in the lab is an experiment waiting to happen. If a group that is seeking to start a project defines discrete modules of a common experiment, and provides a communications platform to allow people to download project modules, perform them, and upload results, it would be possible to harness the overcapacity that exists in laboratories. In principle, although this is a harder empirical question, the same could be done for other widely available laboratory materials and even animals for preclinical trials on the model of, "brother, can you spare a mouse?" One fascinating proposal and early experiment at the University of Indiana-Purdue University Indianapolis was suggested by William Scott, a chemistry professor. Scott proposed developing simple, low-cost kits for training undergraduate students in chemical synthesis, but which would use targets and molecules identified by computational biology as potential treatments

for developing-world diseases as their output. With enough redundancy across different classrooms and institutions around the world, the results could be verified while screening and synthesizing a significant number of potential drugs. The undergraduate educational experience could actually contribute to new experiments, as opposed simply to synthesizing outputs that are not really needed by anyone. Clinical trials provide yet another level of complexity, because the problem of delivering consistent drug formulations for testing to physicians and patients stretches the imagination. One option would be that research centers in countries affected by the diseases in question could pick up the work at this point, and create and conduct clinical trials. These too could be coordinated across regions and countries among the clinicians administering the tests, so that accruing patients and obtaining sufficient information could be achieved more rapidly and at lower cost. As in the case of One World Health, production and regulatory approval, from this stage on, could be taken up by the generics manufacturers. In order to prevent the outputs from being appropriated at this stage, every stage in the process would require a public-domain-binding license that would prevent a manufacturer from taking the outputs and, by making small changes, patenting the ultimate drug.

This proposal about medicine is, at this stage, the most imaginary among the commons-based strategies for development suggested here. However, it is analytically consistent with them, and, in principle, should be attainable. In combination with the more traditional commons-based approaches, university research, and the nonprofit world, peer production could contribute to an innovation ecology that could overcome the systematic inability of a purely patent-based system to register and respond to the health needs of the world's poor.

Commons-Based Strategies for Development: Conclusion

Welfare, development, and growth outside of the core economies heavily depend on the transfer of information-embedded goods and tools, information, and knowledge from the technologically advanced economies to the developing and less-developed economies and societies around the globe. These are important partly as finished usable components of welfare. Perhaps more important, however, they are necessary as tools and platforms on which innovation, research, and development can be pursued by local actors in the developing world itself - from the free software developers of Brazil to the agricultural scientists and farmers of Southeast Asia. The primary obstacles to diffusion of these desiderata in the required direction are the institutional framework of intellectual property and trade and the political power of the patent-dependent business models in the information-exporting economies. This is not because the proprietors of information goods and tools are evil. It is because their fiduciary duty is to maximize shareholder value, and the less-developed and developing economies have little money. As rational maximizers with a legal monopoly, the patent holders restrict output and sell at higher rates. This is not a bug in the institutional system we call "intellectual property." It is a known feature that has known undesirable side effects of inefficiently restricting access to the products of innovation. In the context of vast disparities in wealth across the globe, however, this known feature does not merely lead to less than theoretically optimal use of the information. It leads to predictable increase of morbidity and mortality and to higher barriers to development.

The rise of the networked information economy provides a new framework for thinking about how to work around the barriers that the international intellectual property regime places on development. Public-sector and other nonprofit institutions that have traditionally played an important role in development can do so with a greater degree of efficacy. Moreover, the emergence of peer production provides a model for new solutions to some of the problems of access to information and knowledge. In software and communications, these are directly available.

In scientific information and some educational materials, we are beginning to see adaptations of these models to support core elements of development and learning. In food security and health, the translation process may be more difficult. In agriculture, we are seeing more immediate progress in the development of a woven fabric of public-sector, academic, nonprofit, and individual innovation and learning to pursue biological innovation outside of the markets based on patents and breeders' rights. In medicine, we are still at a very early stage of organizational experiments and institutional proposals. The barriers to implementation are significant. However, there is growing awareness of the human cost of relying solely on the patent-based production system, and of the potential of commons-based strategies to alleviate these failures.

Ideally, perhaps, the most direct way to arrive at a better system for harnessing innovation to development would pass through a new international politics of development, which would result in a better-designed international system of trade and innovation policy. There is in fact a global movement of NGOs and developing nations pursuing this goal. It is possible, however, that the politics of international trade are sufficiently bent to the purposes of incumbent industrial information economy proprietors and the governments that support them as a matter of industrial policy that the political path of formal institutional reform will fail. Certainly, the history of the TRIPS agreement and, more recently, efforts to pass new expansive treaties through the WIPO suggest this. However, one of the lessons we learn as we look at the networked information economy is that the work of governments through international treaties is not the final word on innovation and its diffusion across boundaries of wealth. The emergence of social sharing as a substantial mode of production in the networked environment offers an alternative route for individuals and nonprofit entities to take a much more substantial role in delivering actual desired outcomes independent of the formal system. Commons-based and peer production efforts may not be a cure-all. However, as we have seen in the software world, these strategies can make a big contribution to quite fundamental aspects of human welfare and development. And this is where freedom and justice coincide.

The practical freedom of individuals to act and associate freely - free from the constraints of proprietary endowment, free from the constraints of formal relations of contract or stable organizations - allows individual action in ad hoc, informal association to emerge as a new global mover. It frees the ability of people to act in response to all their motivations. In doing so, it offers a new path, alongside those of the market and formal governmental investment in public welfare, for achieving definable and significant improvements in human development throughout the world.

Notes

1. Anne Alstott and Bruce Ackerman, *The Stakeholder Society* (New Haven, CT: Yale University Press, 1999).

2. Numbers are all taken from the *2004 Human Development Report* (New York: UN Development Programme, 2004).

3. Amartya Sen, *Development as Freedom* (New York: Knopf, 1999), 46-47.

4. Carol Tenopir and Donald W. King, *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers* (Washington, DC: Special Libraries Association, 2000), 273.

5. Harold Varmus, *E-Biomed: A Proposal for Electronic Publications in the Biomedical Sciences* (Bethesda, MD: National Institutes of Health, 1999).

6. C. K. Prahald, *The Fortune at the Bottom of the Pyramid: Eradicating Poverty Through Profits* (Upper Saddle River, NJ: Wharton School of Publishing, 2005), 319-357, Section 4, "The ITC e-Choupal Story."

7. For the sources of numbers for the software industry, see chapter 2 in this volume. IBM numbers, in particular, are identified in figure 2.1.

8. These arguments were set out most clearly and early in a public exchange of letters between Representative Villanueva Nunez in Peru and Microsoft's representatives in that country. The exchange can be found on the Web site of the Open Source Initiative, http://www.opensource.org/docs/peru_and_ms.php .

9. A good regional study of the extent and details of educational deprivation is Mahbub ul Haq and Khadija ul Haq, *Human Development in South Asia 1998: The Education Challenge* (Islamabad, Pakistan: Human Development Center).

10. Robert Evenson and D. Gollin, eds., *Crop Variety Improvement and Its Effect on Productivity: The Impact of International Agricultural Research* (New York: CABI Pub., 2002); results summarized in Robert Evenson and D. Gollin, "Assessing the Impact of the Green Revolution, 1960-2000," *Science* 300 (May 2003): 758-762.

11. Jack R. Kloppenburg, Jr., *First the Seed: The Political Economy of Plant Biotechnology 1492-2000* (Cambridge and New York: Cambridge University Press, 1988), table 2.2.

12. USDA National Agriculture Statistics Survey (2004), http://www.usda.gov/nass/aggraphs/fncht3.htm .

13. First Report of the GM Science Review Panel, *An Open Review of the Science Relevant to GM Crops and Food Based on the Interests and Concerns of the Public*, United Kingdom, July 2003.

14. Robert E. Evenson, "GMOs: Prospects for Productivity Increases in Developing Countries," *Journal of Agricultural and Food Industrial Organization* 2 (2004): article 2.

15. Elliot Marshall, "A Deal for the Rice Genome," *Science* 296 (April 2002): 34.

16. Jikun Huang et al., "Plant Biotechnology in China," *Science* 295 (2002): 674.

17. Huang et al., "Plant Biotechnology."

18. Richard Atkinson et al., "Public Sector Collaboration for Agricultural IP Management," *Science* 301 (2003): 174.

19. This table is a slightly expanded version of one originally published in Yochai Benkler, "Commons Based Strategies and the Problems of Patents," *Science* 305 (2004): 1110.

20. Wim Broothaertz et al., "Gene Transfer to Plants by Diverse Species of Bacteria," *Nature* 433 (2005): 629.

21. These numbers and others in this paragraph are taken from the 2004 WHO *World Health Report*, Annex Table 2.

22. National Science Foundation, Division of Science Resource Statistics, *Special Report: National Patterns of Research and Development Resources: 2003 NSF 05-308* (Arlington, VA: NSF, 2005), table 1.

23. The detailed analysis can be found in Amy Kapzcynzki et al., "Addressing Global Health Inequities: An Open Licensing Paradigm for Public Sector Inventions," *Berkeley Journal of Law and Technology* (Spring 2005).

24. See Jean Lanjouw, "A New Global Patent Regime for Diseases: U.S. and International Legal Issues," *Harvard Journal of Law & Technology* 16 (2002).

25. S. Maurer, A. Sali, and A. Rai, "Finding Cures for Tropical Disease: Is Open Source the Answer?" *Public Library of Science: Medicine* 1, no. 3 (December 2004): e56.

Chapter 10
Social Ties: Networking Together

Increased practical individual autonomy has been central to my claims throughout this book. It underlies the efficiency and sustainability of nonproprietary production in the networked information economy. It underlies the improvements I describe in both freedom and justice. Many have raised concerns that this new freedom will fray social ties and fragment social relations. On this view, the new freedom is one of detached monads, a freedom to live arid, lonely lives free of the many constraining attachments that make us grounded, well-adjusted human beings. Bolstered by early sociological studies, this perspective was one of two diametrically opposed views that typified the way the Internet's effect on community, or close social relations, was portrayed in the 1990s. The other view, popular among the digerati, was that "virtual communities" would come to represent a new form of human communal existence, providing new scope for building a shared experience of human interaction. Within a few short years, however, empirical research suggests that while neither view had it completely right, it was the dystopian view that got it especially wrong. The effects of the Internet on social relations are obviously complex. It is likely too soon to tell which social practices this new mode of communication will ultimately settle on. The most recent research, however, suggests that the Internet has some fairly well-defined effects on human community and intimate social relations. These effects mark neither breakdown nor transcendence, but they do represent an improvement over the world of television and telephone along most dimensions of normative concern with social relations.

We are seeing two effects: first, and most robustly, we see a thickening of preexisting relations with friends, family, and neighbors, particularly with those who were not easily reachable in the pre-Internet-mediated environment. Parents, for example, use instant messages to communicate with their children who are in college. Friends who have moved away from each other are keeping in touch more than they did before they had e-mail, because e-mail does not require them to coordinate a time to talk or to pay long-distance rates. However, this thickening of contacts seems to occur alongside a loosening of the hierarchical aspects of these relationships, as individuals weave their own web of supporting peer relations into the fabric of what might otherwise be stifling familial relationships. Second, we are beginning to see the emergence of greater scope for limited-purpose, loose relationships. These may not fit the ideal model of "virtual communities." They certainly do not fit a deep conception of "community" as a person's primary source of emotional context and support. They are nonetheless effective and meaningful to their participants. It appears that, as the digitally networked environment begins to displace mass media and telephones, its salient communications characteristics provide new dimensions to thicken existing social relations, while also providing new capabilities for looser and more fluid, but still meaningful social networks. A central aspect of this positive improvement in loose ties has been the technical-organizational shift from an information environment dominated by commercial mass media on a one-to-many model, which does not foster group interaction among viewers, to an information environment that both technically and as a matter of social practice enables user-centric, group-based active cooperation platforms of the kind that typify the networked information economy. This is not to say that the Internet necessarily effects all people, all social groups, and networks identically. The effects on different people in different settings and networks will likely vary, certainly in their magnitude. My purpose here, however, is to respond to the concern that enhanced individual capabilities entail social fragmentation and alienation. The available data do not support that claim as a description of a broad social effect.

From "Virtual Communities" to Fear of Disintegration

Angst about the fragmentation of organic deep social ties, the gemeinschaft community, the family, is hardly a creature of the Internet. In some form or another, the fear that cities, industrialization, rapid transportation, mass communications, and other accoutrements of modern industrial society are leading to alienation, breakdown of the family, and the disruption of community has been a fixed element of sociology since at least the mid-nineteenth century. Its mirror image - the search for real or imagined, more or less idealized community, "grounded" in preindustrial pastoral memory or postindustrial utopia - was often not far behind. Unsurprisingly, this patterned opposition of fear and yearning was replayed in the context of the Internet, as the transformative effect of this new medium made it a new focal point for both strands of thought.

In the case of the Internet, the optimists preceded the pessimists. In his now-classic *The Virtual Community*, Howard Rheingold put it most succinctly in 1993:

> My direct observations of online behavior around the world over the past ten years have led me to conclude that whenever CMC [computer mediated communications] technology becomes available to people anywhere, they inevitably build virtual communities with it, just as microorganisms inevitably create colonies. I suspect that one of the explanations for this phenomenon is the hunger for community that grows in the breasts of people around the world as more and more informal public spaces disappear from our real lives. I also suspect that these new media attract colonies of enthusiasts because CMC enables people to do things with each other in new ways, and to do altogether new kinds of things - just as telegraphs, telephones, and televisions did.

The Virtual Community was grounded on Rheingold's own experience in the WELL (Whole Earth 'Lectronic Link). The WELL was one the earliest well-developed instances of large-scale social interaction among people who started out as strangers but came to see themselves as a community. Its members eventually began to organize meetings in real space to strengthen the bonds, while mostly continuing their interaction through computer-mediated communications. Note the structure of Rheingold's claim in this early passage. There is a hunger for community, no longer satisfied by the declining availability of physical spaces for human connection. There is a newly available medium that allows people to connect despite their physical distance. This new opportunity inevitably and automatically brings people to use its affordances - the behaviors it makes possible - to fulfill their need for human connection. Over and above this, the new medium offers new ways of communicating and new ways of doing things together, thereby enhancing what was previously possible. Others followed Rheingold over the course of the 1990s in many and various ways. The basic structure of the claim about the potential of cyberspace to forge a new domain for human connection, one that overcomes the limitations that industrial mass-mediated society places on community, was oft repeated. The basic observation that the Internet permits the emergence of new relationships that play a significant role in their participants' lives and are anchored in online communications continues to be made. As discussed below, however, much of the research suggests that the new online relationships develop in addition to, rather than instead of, physical face-to-face human interaction in community and family - which turns out to be alive and well.

It was not long before a very different set of claims emerged about the Internet. Rather than a solution to the problems that industrial society creates for family and society, the Internet was seen as increasing alienation by absorbing its users. It made them unavailable to spend time with their families. It immersed them in diversions from the real world with its real relationships. In a

social-relations version of the Babel objection, it was seen as narrowing the set of shared cultural experiences to such an extent that people, for lack of a common sitcom or news show to talk about, become increasingly alienated from each other. One strand of this type of criticism questioned the value of online relationships themselves as plausible replacements for real-world human connection. Sherry Turkle, the most important early explorer of virtual identity, characterized this concern as: "is it really sensible to suggest that the way to revitalize community is to sit alone in our rooms, typing at our networked computers and filling our lives with virtual friends?"[/1] Instead of investing themselves with real relationships, risking real exposure and connection, people engage in limited-purpose, low-intensity relationships. If it doesn't work out, they can always sign off, and no harm done.

Another strand of criticism focused less on the thinness, not to say vacuity, of online relations, and more on sheer time. According to this argument, the time and effort spent on the Net came at the expense of time spent with family and friends. Prominent and oft cited in this vein were two early studies. The first, entitled *Internet Paradox*, was led by Robert Kraut.[/2] It was the first longitudinal study of a substantial number of users - 169 users in the first year or two of their Internet use. Kraut and his collaborators found a slight, but statistically significant, correlation between increases in Internet use and (a) decreases in family communication, (b) decreases in the size of social circle, both near and far, and (c) an increase in depression and loneliness. The researchers hypothesized that use of the Internet replaces strong ties with weak ties. They ideal-typed these communications as exchanging knitting tips with participants in a knitting Listserv, or jokes with someone you would meet on a tourist information site. These trivialities, they thought, came to fill time that, in the absence of the Internet, would be spent with people with whom one has stronger ties. From a communications theory perspective, this causal explanation was more sophisticated than the more widely claimed assimilation of the Internet and television - that a computer monitor is simply one more screen to take away from the time one has to talk to real human beings.[/3] It recognized that using the Internet is fundamentally different from watching TV. It allows users to communicate with each other, rather than, like television, encouraging passive reception in a kind of "parallel play." Using a distinction between strong ties and weak ties, introduced by Mark Granovetter in what later became the social capital literature, these researchers suggested that the kind of human contact that was built around online interactions was thinner and less meaningful, so that the time spent on these relationships, on balance, weakened one's stock of social relations.

A second, more sensationalist release of a study followed two years later. In 2000, the Stanford Institute for the Quantitative Study of Society's "preliminary report" on Internet and society, more of a press release than a report, emphasized the finding that "the more hours people use the Internet, the less time they spend with real human beings."[/4] The actual results were somewhat less stark than the widely reported press release. As among all Internet users, only slightly more than 8 percent reported spending less time with family; 6 percent reported spending more time with family, and 86 percent spent about the same amount of time. Similarly, 9 percent reported spending less time with friends, 4 percent spent more time, and 87 percent spent the same amount of time.[/5] The press release probably should not have read, "social isolation increases," but instead, "Internet seems to have indeterminate, but in any event small, effects on our interaction with family and friends" - hardly the stuff of front-page news coverage.[/6] The strongest result supporting the "isolation" thesis in that study was that 27 percent of respondents who were heavy Internet users reported spending less time on the phone with friends and family. The study did not ask whether they used e-mail instead of the phone to keep in touch with these family and friends, and whether they thought they had more or less of a connection with these friends and family as a result. Instead, as the author reported in his press release, "E-mail is a way to stay in touch, but you can't share coffee or beer with somebody on e-mail, or give them a hug" (as opposed, one supposes, to

the common practice of phone hugs).[7] As Amitai Etzioni noted in his biting critique of that study, the truly significant findings were that Internet users spent less time watching television and shopping. Forty-seven percent of those surveyed said that they watched less television than they used to, and that number reached 65 percent for heavy users and 27 percent for light users. Only 3 percent of those surveyed said they watched more TV. Nineteen percent of all respondents and 25 percent of those who used the Internet more than five hours a week said they shopped less in stores, while only 3 percent said they shopped more in stores. The study did not explore how people were using the time they freed by watching less television and shopping less in physical stores. It did not ask whether they used any of this newfound time to increase and strengthen their social and kin ties.[8]

A More Positive Picture Emerges Over Time

The concerns represented by these early studies of the effects of Internet use on community and family seem to fall into two basic bins. The first is that sustained, more or less intimate human relations are critical to well-functioning human beings as a matter of psychological need. The claims that Internet use is associated with greater loneliness and depression map well onto the fears that human connection ground into a thin gruel of electronic bits simply will not give people the kind of human connectedness they need as social beings. The second bin of concerns falls largely within the "social capital" literature, and, like that literature itself, can be divided largely into two main subcategories. The first, following James Coleman and Mark Granovetter, focuses on the economic function of social ties and the ways in which people who have social capital can be materially better off than people who lack it. The second, exemplified by Robert Putnam's work, focuses on the political aspects of engaged societies, and on the ways in which communities with high social capital-defined as social relations with people in local, stable, face-to-face interactions - will lead to better results in terms of political participation and the provisioning of local public goods, like education and community policing. For this literature, the shape of social ties, their relative strength, and who is connected to whom become more prominent features.

There are, roughly speaking, two types of responses to these concerns. The first is empirical. In order for these concerns to be valid as applied to increasing use of Internet communications, it must be the case that Internet communications, with all of their inadequacies, come to supplant real-world human interactions, rather than simply to supplement them. Unless Internet connections actually displace direct, unmediated, human contact, there is no basis to think that using the Internet will lead to a decline in those nourishing connections we need psychologically, or in the useful connections we make socially, that are based on direct human contact with friends, family, and neighbors. The second response is theoretical. It challenges the notion that the socially embedded individual is a fixed entity with unchanging needs that are, or are not, fulfilled by changing social conditions and relations. Instead, it suggests that the "nature" of individuals changes over time, based on actual social practices and expectations. In this case, we are seeing a shift from individuals who depend on social relations that are dominated by locally embedded, thick, unmediated, given, and stable relations, into networked individuals - who are more dependent on their own combination of strong and weak ties, who switch networks, cross boundaries, and weave their own web of more or less instrumental, relatively fluid relationships. Manuel Castells calls this the "networked society,"[9] Barry Wellman, "networked individualism."[10] To simplify vastly, it is not that people cease to depend on others and their context for both psychological and social well-being and efficacy. It is that the kinds of connections that we come to rely on for these basic human needs change over time. Comparisons of current practices to the old ways of achieving the desiderata of community, and fears regarding the loss of community, are more a form of nostalgia than a diagnosis of present social malaise.

Users Increase Their Connections with Preexisting Relations

The most basic response to the concerns over the decline of community and its implications for both the psychological and the social capital strands is the empirical one. Relations with one's local geographic community and with one's intimate friends and family do not seem to be substantially affected by Internet use. To the extent that these relationships are affected, the effect is positive. Kraut and his collaborators continued their study, for example, and followed up with their study subjects for an additional three years. They found that the negative effects they had reported in the first year or two dissipated over the total period of observation.[/11] Their basic hypothesis that the Internet probably strengthened weak ties, however, is consistent with other research and theoretical work. One of the earliest systematic studies of high-speed Internet access and its effects on communities in this vein was by Keith Hampton and Barry Wellman.[/12] They studied the aptly named Toronto suburb Netville, where homes had high-speed wiring years before broadband access began to be adopted widely in North America. One of their most powerful findings was that people who were connected recognized three times as many of their neighbors by name and regularly talked with twice as many as those who were not wired. On the other hand, however, stronger ties - indicated by actually visiting neighbors, as opposed to just knowing their name or stopping to say good morning - were associated with how long a person had lived in the neighborhood, not with whether or not they were wired. In other words, weak ties of the sort of knowing another's name or stopping to chat with them were significantly strengthened by Internet connection, even within a geographic neighborhood. Stronger ties were not. Using applications like a local e-mail list and personal e-mails, wired residents communicated with others in their neighborhood much more often than did nonwired residents. Moreover, wired residents recognized the names of people in a wider radius from their homes, while nonwired residents tended to know only people within their block, or even a few homes on each side. However, again, stronger social ties, like visiting and talking face-to-face, tended to be concentrated among physically proximate neighbors. Other studies also observed this increase of weak ties in a neighborhood with individuals who are more geographically distant than one's own immediate street or block.[/13] Perhaps the most visible aspect of the social capital implications of a well-wired geographic community was the finding that wired neighbors began to sit on their front porches, instead of in their backyard, thereby providing live social reinforcement of community through daily brief greetings, as well as creating a socially enforced community policing mechanism.

We now have quite a bit of social science research on the side of a number of factual propositions.[/14] Human beings, whether connected to the Internet or not, continue to communicate preferentially with people who are geographically proximate than with those who are distant.[/15] Nevertheless, people who are connected to the Internet communicate more with people who are geographically distant without decreasing the number of local connections. While the total number of connections continues to be greatest with proximate family members, friends, coworkers, and neighbors, the Internet's greatest effect is in improving the ability of individuals to add to these proximate relationships new and better-connected relationships with people who are geographically distant. This includes keeping more in touch with friends and relatives who live far away, and creating new weak-tie relationships around communities of interest and practice. To the extent that survey data are reliable, the most comprehensive and updated surveys support these observations. It now seems clear that Internet users "buy" their time to use the Internet by watching less television, and that the more Internet experience they have, the less they watch TV. People who use the Internet claim to have increased the number of people they stay in touch with, while mostly reporting no effect on time they spend with their family.[/16]

Connections with family and friends seemed to be thickened by the new channels of communication, rather than supplanted by them. Emblematic of this were recent results of a survey conducted by the Pew project on "Internet and American Life" on *Holidays Online*. Almost half of respondents surveyed reported using e-mail to organize holiday activities with family (48 percent) and friends (46 percent), 27 percent reported sending or receiving holiday greetings, and while a third described themselves as shopping online in order to save money, 51 percent said they went online to find an unusual or hard-to-find gift. In other words, half of those who used the Internet for holiday shopping did so in order to personalize their gift further, rather than simply to take advantage of the most obvious use of e-commerce-price comparison and time savings. Further support for this position is offered in another Pew study, entitled "Internet and Daily Life." In that survey, the two most common uses - both of which respondents claimed they did more of because of the Net than they otherwise would have - were connecting with family and friends and looking up information./17 Further evidence that the Internet is used to strengthen and service preexisting relations, rather than create new ones, is the fact that 79 percent of those who use the Internet at all do so to communicate with friends and family, while only 26 percent use the Internet to meet new people or to arrange dates. Another point of evidence is the use of instant messaging (IM). IM is a synchronous communications medium that requires its users to set time aside to respond and provides information to those who wish to communicate with an individual about whether that person is or is not available at any given moment. Because it is so demanding, IM is preferentially useful for communicating with individuals with whom one already has a preexisting relationship. This preferential use for strengthening preexisting relations is also indicated by the fact that two-thirds of IM users report using IM with no more than five others, while only one in ten users reports instant messaging with more than ten people. A recent Pew study of instant messaging shows that 53 million adults - 42 percent of Internet users in the United States - trade IM messages. Forty percent use IM to contact coworkers, one-third family, and 21 percent use it to communicate equally with both. Men and women IM in equal proportions, but women IM more than men do, averaging 433 minutes per month as compared to 366 minutes, respectively, and households with children IM more than households without children.

These studies are surveys and local case studies. They cannot offer a knockdown argument about how "we" - everyone, everywhere - are using the Internet. The same technology likely has different effects when it is introduced into cultures that differ from each other in their pre-Internet baseline./18 Despite these cautions, these studies do offer the best evidence we have about Internet use patterns. As best we can tell from contemporary social science, Internet use increases the contact that people have with others who traditionally have been seen as forming a person's "community": family, friends, and neighbors. Moreover, the Internet is also used as a platform for forging new relationships, in addition to those that are preexisting. These relationships are more limited in nature than ties to friends and family. They are detached from spatial constraints, and even time synchronicity; they are usually interest or practice based, and therefore play a more limited role in people's lives than the more demanding and encompassing relationships with family or intimate friends. Each discrete connection or cluster of connections that forms a social network, or a network of social relations, plays some role, but not a definitive one, in each participant's life. There is little disagreement among researchers that these kinds of weak ties or limited-liability social relationships are easier to create on the Internet, and that we see some increase in their prevalence among Internet users. The primary disagreement is interpretive - in other words, is it, on balance, a good thing that we have multiple, overlapping, limited emotional liability relationships, or does it, in fact, undermine our socially embedded being?

Networked Individuals

The interpretive argument about the normative value of the increase in weak ties is colored by the empirical finding that the time spent on the Internet in these limited relationships does not come at the expense of the number of communications with preexisting, real-world relationships. Given our current state of sociological knowledge, the normative question cannot be whether online relations are a reasonable replacement for real-world friendship. Instead, it must be how we understand the effect of the interaction between an increasingly thickened network of communications with preexisting relations and the casting of a broader net that captures many more, and more varied, relations. What is emerging in the work of sociologists is a framework that sees the networked society or the networked individual as entailing an abundance of social connections and more effectively deployed attention. The concern with the decline of community conceives of a scarcity of forms of stable, nurturing, embedding relations, which are mostly fixed over the life of an individual and depend on long-standing and interdependent relations in stable groups, often with hierarchical relations. What we now see emerging is a diversity of forms of attachment and an abundance of connections that enable individuals to attain discrete components of the package of desiderata that "community" has come to stand for in sociology. As Wellman puts it: "Communities and societies have been changing towards networked societies where boundaries are more permeable, interactions are with diverse others, linkages switch between multiple networks, and hierarchies are flatter and more recursive. . . . Their work and community networks are diffuse, sparsely knit, with vague, overlapping, social and spatial boundaries."[/19] In this context, the range and diversity of network connections beyond the traditional family, friends, stable coworkers, or village becomes a source of dynamic stability, rather than tension and disconnect.

The emergence of networked individuals is not, however, a mere overlay, "floating" on top of thickened preexisting social relations without touching them except to add more relations. The interpolation of new networked connections, and the individual's role in weaving those for him- or herself, allows individuals to reorganize their social relations in ways that fit them better. They can use their network connections to loosen social bonds that are too hierarchical and stifling, while filling in the gaps where their real-world relations seem lacking. Nowhere is this interpolation clearer than in Mizuko Ito's work on the use of mobile phones, primarily for text messaging and e-mail, among Japanese teenagers.[/20] Japanese urban teenagers generally live in tighter physical quarters than their American or European counterparts, and within quite strict social structures of hierarchy and respect. Ito and others have documented how these teenagers use mobile phones - primarily as platforms for text messages - that is, as a mobile cross between e-mail and instant messaging and more recently images, to loosen the constraints under which they live. They text at home and in the classroom, making connections to meet in the city and be together, and otherwise succeed in constructing a network of time- and space-bending emotional connections with their friends, without - and this is the critical observation - breaking the social molds they otherwise occupy. They continue to spend time in their home, with their family. They continue to show respect and play the role of child at home and at school. However, they interpolate that role and those relations with a sub-rosa network of connections that fulfill otherwise suppressed emotional needs and ties.

The phenomenon is not limited to youths, but is applicable more generally to the capacity of users to rely on their networked connections to escape or moderate some of the more constraining effects of their stable social connections. In the United States, a now iconic case - mostly described in terms of privacy - was that of U.S. Navy sailor Timothy McVeigh (not the Oklahoma bomber). McVeigh was discharged from the navy when his superiors found out that he was gay by accessing his AOL (America Online) account. The case was primarily considered in terms of McVeigh's e-mail account privacy. It settled for an undisclosed sum, and McVeigh retired from the navy with

benefits. However, what is important for us here is not the "individual rights" category under which the case was fought, but the practice that it revealed. Here was an eighteen-year veteran of the navy who used the space-time breaking possibilities of networked communications to loosen one of the most constraining attributes imaginable of the hierarchical framework that he nonetheless chose to be part of - the U.S. Navy. It would be odd to think that the navy did not provide McVeigh with a sense of identity and camaraderie that closely knit communities provide their members. Yet at the same time, it also stifled his ability to live one of the most basic of all human ties - his sexual identity. He used the network and its potential for anonymous and pseudonymous existence to coexist between these two social structures.

At the other end of the spectrum of social ties, we see new platforms emerging to generate the kinds of bridging relations that were so central to the identification of "weak ties" in social capital literature. Weak ties are described in the social capital literature as allowing people to transmit information across social networks about available opportunities and resources, as well as provide at least a limited form of vouching for others - as one introduces a friend to a friend of a friend. What we are seeing on the Net is an increase in the platforms developed to allow people to create these kinds of weak ties based on an interest or practice. Perhaps clearest of these is Meetup.com. Meetup is a Web site that allows users to search for others who share an interest and who are locally available to meet face-to-face. The search results show users what meetings are occurring within their requested area and interest. The groups then meet periodically, and those who sign up for them also are able to provide a profile and photo of themselves, to facilitate and sustain the real-world group meetings. The power of this platform is that it is not intended as a replacement for real-space meetings. It is intended as a replacement for the happenstance of social networks as they transmit information about opportunities for interest- and practice-based social relations. The vouching function, on the other hand, seems to have more mixed efficacy, as Dana Boyd's ethnography of Friendster suggests.[/21] Friendster was started as a dating Web site. It was built on the assumption that dating a friend of a friend of a friend is safer and more likely to be successful than dating someone based on a similar profile, located on a general dating site like match.com - in other words, that vouching as friends provides valuable information. As Boyd shows, however, the attempt of Friendster to articulate and render transparent the social networks of its users met with less than perfect success. The platform only permits users to designate friend/not friend, without the finer granularity enabled by a face-to-face conversation about someone, where one can answer or anticipate the question, "just how well do you know this person?" with a variety of means, from tone to express reservations. On Friendster, it seems that people cast broader networks, and for fear of offending or alienating others, include many more "friendsters" than they actually have "friends." The result is a weak platform for mapping general connections, rather than a genuine articulation of vouching through social networks. Nonetheless, it does provide a visible rendering of at least the thinnest of weak ties, and strengthens their effect in this regard. It enables very weak ties to perform some of the roles of real-world weak social ties.

The Internet as a Platform for Human Connection

Communication is constitutive of social relations. We cannot have relationships except by communicating with others. Different communications media differ from each other - in who gets to speak to whom and in what can be said. These differences structure the social relations that rely on these various modes of communication so that they differ from each other in significant ways. Technological determinism is not required to accept this. Some aspects of the difference are purely technical. Script allows text and more or less crude images to be transmitted at a distance, but not voice, touch, smell, or taste. To the extent that there are human emotions, modes of submission and exertion of authority, irony, love or affection, or information that is easily encoded and

conveyed in face-to-face communications but not in script, script-based communications are a poor substitute for presence. A long and romantic tradition of love letters and poems notwithstanding, there is a certain thinness to that mode in the hands of all but the most gifted writers relative to the fleshiness of unmediated love. Some aspects of the difference among media of communication are not necessarily technical, but are rather culturally or organizationally embedded. Television can transmit text. However, text distribution is not television's relative advantage in a sociocultural environment that already has mass-circulation print media, and in a technical context where the resolution of television images is relatively low. As a matter of cultural and business practice, therefore, from its inception, television emphasized moving images and sound, not text transmission. Radio could have been deployed as short-range, point-to-point personal communications systems, giving us a nation of walkie-talkies. However, as chapter 6 described, doing so would have required a very different set of regulatory and business decisions between 1919 and 1927. Communications media take on certain social roles, structures of control, and emphases of style that combine their technical capacities and limits with the sociocultural business context into which they were introduced, and through which they developed. The result is a cluster of use characteristics that define how a given medium is used within a given society, in a given historical context. They make media differ from each other, providing platforms with very different capacities and emphases for their users.

As a technical and organizational matter, the Internet allows for a radically more diverse suite of communications models than any of the twentieth-century systems permitted. It allows for textual, aural, and visual communications. It permits spatial and temporal asynchronicity, as in the case of e-mail or Web pages, but also enables temporal synchronicity - as in the case of IM, online game environments, or Voice over Internet Protocol (VoIP). It can even be used for subchannel communications within a spatially synchronous context, such as in a meeting where people pass electronic notes to each other by e-mail or IM. Because it is still highly textual, it requires more direct attention than radio, but like print, it is highly multiplexable - both between uses of the Internet and other media, and among Internet uses themselves. Similar to print media, you can pick your head up from the paper, make a comment, and get back to reading. Much more richly, one can be on a voice over IP conversation and e-mail at the same time, or read news interlaced with receiving and responding to e-mail. It offers one-to-one, one-to-few, few-to-few, one-to-many, and many-to-many communications capabilities, more diverse in this regard than any medium for social communication that preceded it, including - on the dimensions of distance, asynchronicity, and many-to-many capabilities - even that richest of media: face-to-face communications.

Because of its technical flexibility and the "business model" of Internet service providers as primarily carriers, the Internet lends itself to being used for a wide range of social relations. Nothing in "the nature of the technology" requires that it be the basis of rich social relations, rather than becoming, as some predicted in the early 1990s, a "celestial jukebox" for the mass distribution of prepackaged content to passive end points. In contradistinction to the dominant remote communications technologies of the twentieth century, however, the Internet offers some new easy ways to communicate that foster both of the types of social communication that the social science literature seems to be observing. Namely, it makes it easy to increase the number of communications with preexisting friends and family, and increases communication with geographically distant or more loosely affiliated others. Print, radio, television, film, and sound recording all operated largely on a one-to-many model. They did not, given the economics of production and transmission, provide a usable means of remote communication for individuals at the edges of these communication media. Television, film, sound recording, and print industries were simply too expensive, and their business organization was too focused on selling broadcast-model communications, to support significant individual communication. When cassette

tapes were introduced, we might have seen people recording a tape instead of writing a letter to friends or family. However, this was relatively cumbersome, low quality, and time consuming. Telephones were the primary means of communications used by individuals, and they indeed became the primary form of mediated personal social communications. However, telephone conversations require synchronicity, which means that they can only be used for socializing purposes when both parties have time. They were also only usable throughout this period for serial, one-to-one conversations. Moreover, for most of the twentieth century, a long-distance call was a very expensive proposition for most nonbusiness users, and outside of the United States, local calls too carried nontrivial time-sensitive prices in most places. Telephones were therefore a reasonable medium for social relations with preexisting friends and family. However, their utility dropped off radically with the cost of communication, which was at a minimum associated with geographic distance. In all these dimensions, the Internet makes it easier and cheaper to communicate with family and friends, at close proximity or over great distances, through the barriers of busy schedules and differing time zones. Moreover, because of the relatively low-impact nature of these communications, the Internet allows people to experiment with looser relations more readily. In other words, the Internet does not make us more social beings. It simply offers more degrees of freedom for each of us to design our own communications space than were available in the past. It could have been that we would have used that design flexibility to re-create the mass-media model. But to predict that it would be used in this fashion requires a cramped view of human desire and connectedness. It was much more likely that, given the freedom to design our own communications environment flexibly and to tailor it to our own individual needs dynamically over time, we would create a system that lets us strengthen the ties that are most important to us. It was perhaps less predictable, but unsurprising after the fact, that this freedom would also be used to explore a wider range of relations than simply consuming finished media goods.

There is an appropriate wariness in contemporary academic commentary about falling into the trap of "the mythos of the electrical sublime" by adopting a form of Internet utopianism.[/22](#) It is important, however, not to let this caution blind us to the facts about Internet use, and the technical, business, and cultural capabilities that the Internet makes feasible. The cluster of technologies of computation and communications that characterize the Internet today are, in fact, used in functionally different ways, and make for several different media of communication than we had in the twentieth century. The single technical platform might best be understood to enable several different "media" - in the sense of clusters of technical-social-economic practices of communication - and the number of these enabled media is growing. Instant messaging came many years after e-mail, and a few years after Web pages. Blogging one's daily journal on LiveJournal so that a group of intimates can check in on one's life as it unfolds was not a medium that was available to users until even more recently. The Internet is still providing its users with new ways to communicate with each other, and these represent a genuinely wide range of new capabilities. It is therefore unsurprising that connected social beings, such as we are, will take advantage of these new capabilities to form connections that were practically infeasible in the past. This is not media determinism. This is not millenarian utopianism. It is a simple observation. People do what they can, not what they cannot. In the daily humdrum of their lives, individuals do more of what is easier to do than what requires great exertion. When a new medium makes it easy for people to do new things, they may well, in fact, do them. And when these new things are systematically more user-centric, dialogic, flexible in terms of the temporal and spatial synchronicity they require or enable, and multiplexable, people will communicate with each other in ways and amounts that they could not before.

The Emergence of Social Software

The design of the Internet itself is agnostic as among the social structures and relations it enables. At its technical core is a commitment to push all the detailed instantiations of human communications to the edges of the network - to the applications that run on the computers of users. This technical agnosticism leads to a social agnosticism. The possibility of large-scale sharing and cooperation practices, of medium-scale platforms for collaboration and discussion, and of small-scale, one-to-one communications has led to the development of a wide range of software designs and applications to facilitate different types of communications. The World Wide Web was used initially as a global broadcast medium available to anyone and everyone, everywhere. In e-mail, we see a medium available for one-to-one, few-to-few, one-to-many and, to a lesser extent, many-to-many use. One of the more interesting phenomena of the past few years is the emergence of what is beginning to be called "social software." As a new design space, it is concerned with groups that are, as defined by Clay Shirky, who first articulated the concept, "Larger than a dozen, smaller than a few hundred, where people can actually have these conversational forms that can't be supported when you're talking about tens of thousands or millions of users, at least in a single group." The definition of the term is somewhat amorphous, but the basic concept is software whose design characteristic is that it treats genuine social phenomena as different from one-to-one or one-to-many communications. It seeks to build one's expectations about the social interactions that the software will facilitate into the design of the platform. The design imperative was most clearly articulated by Shirky when he wrote that from the perspective of the software designer, the user of social software is the group, not the individual./23

A simple example will help to illustrate. Take any given site that uses a collaborative authorship tool, like the Wiki that is the basis of *Wikipedia* and many other cooperative authorship exercises. From the perspective of an individual user, the ease of posting a comment on the Wiki, and the ease of erasing one's own comments from it, would be important characteristics: The fewer registration and sign-in procedures, the better. Not so from the perspective of the group. The group requires some "stickiness" to make the group as a group, and the project as a project, avoid the rending forces of individualism and self-reference. So, for example, design components that require registration for posting, or give users different rights to post and erase comments over time, depending on whether they are logged in or not, or depending on a record of their past cooperative or uncooperative behavior, are a burden for the individual user. However, that is precisely their point. They are intended to give those users with a greater stake in the common enterprise a slight, or sometimes large, edge in maintaining the group's cohesion. Similarly, erasing past comments may be useful for the individual, for example, if they were silly or untempered. Keeping the comments there is, however, useful to the group - as a source of experience about the individual or part of the group's collective memory about mistakes made in the past that should not be repeated by someone else. Again, the needs of the group as a group often differ from those of the individual participant. Thinking of the platform as social software entails designing it with characteristics that have a certain social-science or psychological model of the interactions of a group, and building the platform's affordances in order to enhance the survivability and efficacy of the group, even if it sometimes comes at the expense of the individual user's ease of use or comfort.

This emergence of social software-like blogs with opportunities to comment, Wikis, as well as social-norm-mediated Listservs or uses of the "cc" line in e-mail - underscores the nondeterministic nature of the claim about the relationship between the Internet and social relations. The Internet makes possible all sorts of human communications that were not technically feasible before its widespread adoption. Within this wide range of newly feasible communications patterns, we are beginning to see the emergence of different types of relationships - some positive, some, like spam

(unsolicited commercial e-mail), decidedly negative. In seeking to predict and diagnose the relationship between the increasing use of Internet communications and the shape of social relations, we see that the newly emerging constructive social possibilities are leading to new design challenges. These, in turn, are finding engineers and enthusiasts willing and able to design for them. The genuinely new capability - connecting among few and many at a distance in a dialogic, recursive form - is leading to the emergence of new design problems. These problems come from the fact that the new social settings come with their own social dynamics, but without long-standing structures of mediation and constructive ordering. Hence the early infamy of the tendency of Usenet and Listservs discussions to deteriorate into destructive flame wars. As social habits of using these kinds of media mature, so that users already know that letting loose on a list will likely result in a flame war and will kill the conversation, and as designers understand that social dynamics - including both those that allow people to form and sustain groups and those that rend them apart with equal if not greater force - we are seeing the coevolution of social norms and platform designs that are intended to give play to the former, and mediate or moderate the latter. These platforms are less likely to matter for sustaining the group in preexisting relations - as among friends or family. The structuring of those relationships is dominated by social norms. However, they do offer a new form and a stabilizing context for the newly emerging diverse set of social relations - at a distance, across interests and contexts - that typify both peer production and many forms of social interaction aimed purely at social reproduction.

The peer-production processes that are described in primarily economic terms in chapter 3 - like free software development, *Wikipedia*, or the Open Directory Project - represent one cluster of important instances of this new form of social relations. They offer a type of relationship that is nonhierarchical and organized in a radically decentralized pattern. Their social valence is given by some combination of the shared experience of joint creativity they enable, as well as their efficacy - their ability to give their users a sense of common purpose and mutual support in achieving it. Individuals adopt projects and purposes they consider worth pursuing. Through these projects they find others, with whom they initially share only a general sense of human connectedness and common practical interest, but with whom they then interact in ways that allow the relationship to thicken over time. Nowhere is this process clearer than on the community pages of *Wikipedia*. Because of the limited degree to which that platform uses technical means to constrain destructive behavior, the common enterprise has developed practices of user-to-user communication, multiuser mediation, and user-appointed mediation to resolve disputes and disagreements. Through their involvement in these, users increase their participation, their familiarity with other participants - at least in this limited role as coauthors - and their practices of mutual engagement with these others. In this way, peer production offers a new platform for human connection, bringing together otherwise unconnected individuals and replacing common background or geographic proximity with a sense of well-defined purpose and the successful common pursuit of this purpose as the condensation point for human connection. Individuals who are connected to each other in a peer-production community may or may not be bowling alone when they are off-line, but they are certainly playing together online.

The Internet and Human Community

This chapter began with a basic question. While the networked information economy may enhance the autonomy of individuals, does it not also facilitate the breakdown of community? The answer offered here has been partly empirical and partly conceptual.

Empirically, it seems that the Internet is allowing us to eat our cake and have it too, apparently keeping our (social) figure by cutting down on the social equivalent of deep-fried dough-television.

That is, we communicate more, rather than less, with the core constituents of our organic communities - our family and our friends - and we seem, in some places, also to be communicating more with our neighbors. We also communicate more with loosely affiliated others, who are geographically remote, and who may share only relatively small slivers of overlapping interests, or for only short periods of life. The proliferation of potential connections creates the social parallel to the Babel objection in the context of autonomy - with all these possible links, will any of them be meaningful? The answer is largely that we do, in fact, employ very strong filtering on our Internet-based social connections in one obvious dimension: We continue to use the newly feasible lines of communication primarily to thicken and strengthen connections with preexisting relationships - family and friends. The clearest indication of this is the parsimony with which most people use instant messaging. The other mechanism we seem to be using to avoid drowning in the noise of potential chitchat with ever-changing strangers is that we tend to find networks of connections that have some stickiness from our perspective. This stickiness could be the efficacy of a cluster of connections in pursuit of a goal one cares about, as in the case of the newly emerging peer-production enterprises. It could be the ways in which the internal social interaction has combined social norms with platform design to offer relatively stable relations with others who share common interests. Users do not amble around in a social equivalent of Brownian motion. They tend to cluster in new social relations, albeit looser and for more limited purposes than the traditional pillars of community.

The conceptual answer has been that the image of "community" that seeks a facsimile of a distant pastoral village is simply the wrong image of how we interact as social beings. We are a networked society now-networked individuals connected with each other in a mesh of loosely knit, overlapping, flat connections. This does not leave us in a state of anomie. We are well-adjusted, networked individuals; well-adjusted socially in ways that those who seek community would value, but in new and different ways. In a substantial departure from the range of feasible communications channels available in the twentieth century, the Internet has begun to offer us new ways of connecting to each other in groups small and large. As we have come to take advantage of these new capabilities, we see social norms and software coevolving to offer new, more stable, and richer contexts for forging new relationships beyond those that in the past have been the focus of our social lives. These do not displace the older relations. They do not mark a fundamental shift in human nature into selfless, community-conscious characters. We continue to be complex beings, radically individual and self-interested at the same time that we are entwined with others who form the context out of which we take meaning, and in which we live our lives. However, we now have new scope for interaction with others. We have new opportunities for building sustained limited-purpose relations, weak and intermediate-strength ties that have significant roles in providing us with context, with a source of defining part of our identity, with potential sources for support, and with human companionship. That does not mean that these new relationships will come to displace the centrality of our more immediate relationships. They will, however, offer increasingly attractive supplements as we seek new and diverse ways to embed ourselves in relation to others, to gain efficacy in weaker ties, and to interpolate different social networks in combinations that provide us both stability of context and a greater degree of freedom from the hierarchical and constraining aspects of some of our social relations.

Notes

1. Sherry Turkle, "Virtuality and Its Discontents, Searching for Community in Cyberspace," *The American Prospect* 7, no. 24 (1996); Sherry Turkle, *Life on the Screen: Identity in the Age of the Internet* (New York: Simon & Schuster, 1995).

2. Robert Kraut et al., "Internet Paradox, A Social Technology that Reduces Social Involvement and Psychological Well Being," *American Psychologist* 53 (1998): 1017-1031.

3. A fairly typical statement of this view, quoted in a study commissioned by the Kellogg Foundation, was: "TV or other media, such as computers, are no longer a kind of 'electronic hearth,' where a family will gather around and make decisions or have discussions. My position, based on our most recent studies, is that most media in the home are working against bringing families together." Christopher Lee et al., "Evaluating Information and Communications Technology: Perspective for a Balanced Approach," Report to the Kellogg Foundation (December 17, 2001), http://www.si.umich.edu/pne/kellogg/013.html .

4. Norman H. Nie and Lutz Ebring, "Internet and Society, A Preliminary Report," Stanford Institute for the Quantitative Study of Society, February 17, 2000, 15 (Press Release), http://www.pkp.ubc.ca/bctf/Stanford_Report.pdf .

5. Ibid., 42-43, tables CH-WFAM, CH-WFRN.

6. See John Markoff and A. Newer, "Lonelier Crowd Emerges in Internet Study," *New York Times*, February 16, 2000, section A, page 1, column 1.

7. Nie and Ebring, "Internet and Society," 19.

8. Amitai Etzioni, "Debating the Societal Effects of the Internet: Connecting with the World," *Public Perspective* 11 (May/June 2000): 42, also available at http://www.gwu.edu/~ccps/etzioni/A273.html .

9. Manuel Castells, *The Rise of Networked Society* 2d ed. (Malden, MA: Blackwell Publishers, Inc., 2000).

10. Barry Wellman et al., "The Social Affordances of the Internet for Networked Individualism," *Journal of Computer Mediated Communication* 8, no. 3 (April 2003).

11. Robert Kraut et al., "Internet Paradox Revisited," *Journal of Social Issues* 58, no. 1 (2002): 49.

12. Keith Hampton and Barry Wellman, "Neighboring in Netville: How the Internet Supports Community and Social Capital in a Wired Suburb," *City & Community* 2, no. 4 (December 2003): 277.

13. Gustavo S. Mesch and Yael Levanon, "Community Networking and Locally-Based Social Ties in Two Suburban Localities," *City & Community* 2, no. 4 (December 2003): 335.

14. Useful surveys include: Paul DiMaggio et al., "Social Implications of the Internet," *Annual Review of Sociology* 27 (2001): 307-336; Robyn B. Driskell and Larry Lyon, "Are Virtual Communities True Communities? Examining the Environments and Elements of Community," *City & Community* 1, no. 4 (December 2002): 349; James E. Katz and Ronald E. Rice, *Social Consequences of Internet Use: Access, Involvement, Interaction* (Cambridge, MA: MIT Press, 2002).

15. Barry Wellman, "Computer Networks as Social Networks," *Science* 293, issue 5537 (September 2001): 2031.

16. Jeffery I. Cole et al., "The UCLA Internet Report: Surveying the Digital Future, Year Three" (UCLA Center for Communication Policy, January 2003), 33, 55, 62, http://www.ccp.ucla.edu/pdf/UCLA-Internet-Report-Year-Three.pdf .

17. Pew Internet and Daily Life Project (August 11, 2004), report available at http://www.pewinternet.org/PPF/r/131/report_display.asp .

18. See Barry Wellman, "The Social Affordances of the Internet for Networked Individualism," *Journal of Computer Mediated Communication* 8, no. 3 (April 2003); Gustavo S. Mesch and Yael Levanon, "Community Networking and Locally-Based Social Ties in Two Suburban Localities, *City & Community* 2, no. 4 (December 2003): 335.

19. Barry Wellman, "The Social Affordances of the Internet."

20. A review of Ito's own work and that of other scholars of Japanese techno-youth culture is Mizuko Ito, "Mobile Phones, Japanese Youth, and the Re-Placement of Social Contact," forthcoming in *Mobile Communications: Re-negotiation of the Social Sphere*, ed., Rich Ling and P. Pedersen (New York: Springer, 2005).

21. Dana M. Boyd, "Friendster and Publicly Articulated Social Networking," *Conference on Human Factors and Computing Systems (CHI 2004)* (Vienna: ACM, April 24-29, 2004).

22. James W. Carrey, *Communication as Culture: Essays on Media and Society* (Boston: Unwin Hyman, 1989).

23. Clay Shirky, "A Group Is Its Own Worst Enemy," published first in *Networks, Economics and Culture* mailing list July 1, 2003.

The Wealth of Networks:
How Social Production Transforms Markets and Freedom
by Yochai Benkler, Yale University Press

Part III
Policies of Freedom at a Moment of Transformation

Epigraph

"Human nature is not a machine to be built after a model, and set to do exactly the work prescribed for it, but a tree, which requires to grow and develop itself on all sides, according to the tendency of the inward forces which make it a living thing."

"Such are the differences among human beings in their sources of pleasure, their susceptibilities of pain, and the operation on them of different physical and moral agencies, that unless there is a corresponding diversity in their modes of life, they neither obtain their fair share of happiness, nor grow up to the mental, moral, and aesthetic stature of which their nature is capable."

John Stuart Mill, On Liberty *(1859)*

Part III
Policies of Freedom at a Moment of Transformation

Part I of this book offers a descriptive, progressive account of emerging patterns of nonmarket individual and cooperative social behavior, and an analysis of why these patterns are internally sustainable and increase information economy productivity. Part II combines descriptive and normative analysis to claim that these emerging practices offer defined improvements in autonomy, democratic discourse, cultural creation, and justice. I have noted periodically, however, that the descriptions of emerging social practices and the analysis of their potential by no means imply that these changes will necessarily become stable or provide the benefits I ascribe them. They are not a deterministic consequence of the adoption of networked computers as core tools of information production and exchange. There is no inevitable historical force that drives the technological-economic moment toward an open, diverse, liberal equilibrium. If the transformation I describe actually generalizes and stabilizes, it could lead to substantial redistribution of power and money. The twentieth-century industrial producers of information, culture, and communications - like Hollywood, the recording industry, and some of the telecommunications giants - stand to lose much. The winners would be a combination of the widely diffuse population of individuals around the globe and the firms or other toolmakers and platform providers who supply these newly capable individuals with the context for participating in the networked information economy. None of the industrial giants of yore are taking this threat lying down. Technology will not overcome their resistance through an insurmountable progressive impulse of history. The reorganization of production and the advances it can bring in freedom and justice will emerge only as a result of social practices and political actions that successfully resist efforts to regulate the emergence of the networked information economy in order to minimize its impact on the incumbents.

Since the middle of the 1990s, we have seen intensifying battles over the institutional ecology within which the industrial mode of information production and the newly emerging networked modes compete. Partly, this has been a battle over telecommunications infrastructure regulation. Most important, however, this has meant a battle over "intellectual property" protection, very broadly defined. Building upon and extending a twenty-five-year trend of expansion of copyrights, patents, and similar exclusive rights, the last half-decade of the twentieth century saw expansion of institutional mechanisms for exerting exclusive control in multiple dimensions. The term of copyright was lengthened. Patent rights were extended to cover software and business methods. Trademarks were extended by the Antidilution Act of 1995 to cover entirely new values, which became the basis for liability in the early domain-name trademark disputes. Most important, we saw a move to create new legal tools with which information vendors could hermetically seal access to their materials to an extent never before possible. The Digital Millennium Copyright Act (DMCA) prohibited the creation and use of technologies that would allow users to get at materials whose owners control through encryption. It prohibited even technologies that users can employ to use the materials in ways that the owners have no right to prevent. Today we are seeing efforts to further extend similar technological regulations - down to the level of regulating hardware to make sure that it complies with design specifications created by the copyright industries. At other layers of the communications environment, we see efforts to expand software patents, to control the architecture of personal computing devices, and to create ever-stronger property rights in physical infrastructure - be it the telephone lines, cable plant, or wireless frequencies. Together, these legislative and judicial acts have formed what many have been calling a second enclosure movement: A concerted effort to shape the institutional ecology in order to help proprietary models of information production at the expense of burdening nonmarket, nonproprietary production.[/1] The

new enclosure movement is not driven purely by avarice and rent seeking - though it has much of that too. Some of its components are based in well-meaning judicial and regulatory choices that represent a particular conception of innovation and its relationship to exclusive rights. That conception, focused on mass-media-type content, movies, and music, and on pharmaceutical-style innovation systems, is highly solicitous of the exclusive rights that are the bread and butter of those culturally salient formats. It is also suspicious of, and detrimental to, the forms of nonmarket, commons-based production emerging in the networked information economy.

This new enclosure movement has been the subject of sustained and diverse academic critique since the mid-1980s./2 The core of this rich critique has been that the cases and statutes of the past decade or so have upset the traditional balance, in copyrights in particular, between seeking to create incentives through the grant of exclusive rights and assuring access to information through the judicious limitation of these rights and the privileging of various uses. I do not seek to replicate that work here, or to offer a comprehensive listing of all the regulatory moves that have increased the scope of proprietary rights in digital communications networks. Instead, I offer a way of framing these various changes as moves in a large-scale battle over the institutional ecology of the digital environment. By "institutional ecology," I mean to say that institutions matter to behavior, but in ways that are more complex than usually considered in economic models. They interact with the technological state, the cultural conceptions of behaviors, and with incumbent and emerging social practices that may be motivated not only by self-maximizing behavior, but also by a range of other social and psychological motivations. In this complex ecology, institutions - most prominently, law - affect these other parameters, and are, in turn, affected by them. Institutions coevolve with technology and with social and market behavior. This coevolution leads to periods of relative stability, punctuated by periods of disequilibrium, which may be caused by external shocks or internally generated phase shifts. During these moments, the various parameters will be out of step, and will pull and tug at the pattern of behavior, at the technology, and at the institutional forms of the behavior. After the tugging and pulling has shaped the various parameters in ways that are more consistent with each other, we should expect to see periods of relative stability and coherence.

Chapter 11 is devoted to an overview of the range of discrete policy areas that are shaping the institutional ecology of digital networks, in which proprietary, market-based models of information production compete with those that are individual, social, and peer produced. In almost all contexts, when presented with a policy choice, advanced economies have chosen to regulate information production and exchange in ways that make it easier to pursue a proprietary, exclusion-based model of production of entertainment goods at the expense of commons- and service-based models of information production and exchange. This has been true irrespective of the political party in power in the United States, or the cultural differences in the salience of market orientation between Europe and the United States. However, the technological trajectory, the social practices, and the cultural understanding are often working at cross-purposes with the regulatory impulse. The equilibrium on which these conflicting forces settle will shape, to a large extent, the way in which information, knowledge, and culture are produced and used over the coming few decades. Chapter 12 concludes the book with an overview of what we have seen about the political economy of information and what we might therefore understand to be at stake in the policy choices that liberal democracies and advanced economies will be making in the coming years.

Chapter 11 The Battle Over the Institutional Ecology of the Digital Environment

Institutional Ecology and Path Dependence

A Framework for Mapping the Institutional Ecology

The Physical Layer

The Logical Layer

The Content Layer

The Problem of Security

Chapter 12 Conclusion: The Stakes of Information Law and Policy

Notes

1. For a review of the literature and a substantial contribution to it, see James Boyle, "The Second Enclosure Movement and the Construction of the Public Domain," *Law and Contemporary Problems* 66 (Winter-Spring 2003): 33-74.

2. Early versions in the legal literature of the skepticism regarding the growth of exclusive rights were Ralph Brown's work on trademarks, Benjamin Kaplan's caution over the gathering storm that would become the Copyright Act of 1976, and Stephen Breyer's work questioning the economic necessity of copyright in many industries. Until, and including the 1980s, these remained, for the most part, rare voices - joined in the 1980s by David Lange's poetic exhortation for the public domain; Pamela Samuelson's systematic critique of the application of copyright to computer programs, long before anyone was paying attention; Jessica Litman's early work on the political economy of copyright legislation and the systematic refusal to recognize the public domain as such; and William Fisher's theoretical exploration of fair use. The 1990s saw a significant growth of academic questioning of enclosure: Samuelson continued to press the question of copyright in software and digital materials; Litman added a steady stream of prescient observations as to where the digital copyright was going and how it was going wrong; Peter Jaszi attacked the notion of the romantic author; Ray Patterson developed a user-centric view of copyright; Diane Zimmerman revitalized the debate over the conflict between copyright and the first amendment; James Boyle introduced erudite criticism of the theoretical coherence of the relentless drive to propertization; Niva Elkin Koren explored copyright and democracy; Keith Aoki questioned trademark, patents, and global trade systems; Julie Cohen early explored technical protection systems and privacy; and Eben Moglen began mercilessly to apply the insights of free software to hack at the foundations of intellectual property apologia. Rebecca Eisenberg, and more recently, Arti Rai, questioned the wisdom of patents on research tools to biomedical innovation. In this decade, William Fisher, Larry Lessig, Litman, and Siva Vaidhyanathan have each described the various forms that the enclosure movement has taken and exposed its many limitations. Lessig and Vaidhyanathan, in particular, have begun to explore the relations between the institutional battles and the freedom in the networked environment.

Chapter 11
The Battle Over the Institutional Ecology of the Digital Environment

The decade straddling the turn of the twenty-first century has seen high levels of legislative and policy activity in the domains of information and communications. Between 1995 and 1998, the United States completely overhauled its telecommunications law for the first time in sixty years, departed drastically from decades of practice on wireless regulation, revolutionized the scope and focus of trademark law, lengthened the term of copyright, criminalized individual user infringement, and created new paracopyright powers for rights holders that were so complex that the 1998 Digital Millennium Copyright Act (DMCA) that enacted them was longer than the entire Copyright Act. Europe covered similar ground on telecommunications, and added a new exclusive right in raw facts in databases. Both the United States and the European Union drove for internationalization of the norms they adopted, through the new World Intellectual Property Organization (WIPO) treaties and, more important, though the inclusion of intellectual property concerns in the international trade regime. In the seven years since then, legal battles have raged over the meaning of these changes, as well as over efforts to extend them in other directions. From telecommunications law to copyrights, from domain name assignment to trespass to server, we have seen a broad range of distinct regulatory moves surrounding the question of control over the basic resources needed to create, encode, transmit, and receive information, knowledge, and culture in the digital environment. As we telescope up from the details of sundry regulatory skirmishes, we begin to see a broad pattern of conflict over the way that access to these core resources will be controlled.

Much of the formal regulatory drive has been to increase the degree to which private, commercial parties can gain and assert exclusivity in core resources necessary for information production and exchange. At the physical layer, the shift to broadband Internet has been accompanied by less competitive pressure and greater legal freedom for providers to exclude competitors from, and shape the use of, their networks. That freedom from both legal and market constraints on exercising control has been complemented by increasing pressures from copyright industries to require that providers exercise greater control over the information flows in their networks in order to enforce copyrights. At the logical layer, anticircumvention provisions and the efforts to squelch peer-to-peer sharing have created institutional pressures on software and protocols to offer a more controlled and controllable environment. At the content layer, we have seen a steady series of institutional changes aimed at tightening exclusivity.

At each of these layers, however, we have also seen countervailing forces. At the physical layer, the Federal Communications Commission's (FCC's) move to permit the development of wireless devices capable of self-configuring as user-owned networks offers an important avenue for a commons-based last mile. The open standards used for personal computer design have provided an open platform. The concerted resistance against efforts to require computers to be designed so they can more reliably enforce copyrights against their users has, to this point, prevented extension of the DMCA approach to hardware design. At the logical layer, the continued centrality of open standard-setting processes and the emergence of free software as a primary modality of producing mission-critical software provide significant resistance to efforts to enclose the logical layer. At the content layer, where law has been perhaps most systematically one-sided in its efforts to enclose, the cultural movements and the technical affordances that form the foundation of the transformation described throughout this book stand as the most significant barrier to enclosure.

It is difficult to tell how much is really at stake, from the long-term perspective, in all these legal battles. From one point of view, law would have to achieve a great deal in order to replicate the twentieth-century model of industrial information economy in the new technical-social context. It would have to curtail some of the most fundamental technical characteristics of computer networks and extinguish some of our most fundamental human motivations and practices of sharing and cooperation. It would have to shift the market away from developing ever-cheaper general-purpose computers whose value to users is precisely their on-the-fly configurability over time, toward more controllable and predictable devices. It would have to squelch the emerging technologies in wireless, storage, and computation that are permitting users to share their excess resources ever more efficiently. It would have to dampen the influence of free software, and prevent people, young and old, from doing the age-old human thing: saying to each other, "here, why don't you take this, you'll like it," with things they can trivially part with and share socially. It is far from obvious that law can, in fact, achieve such basic changes. From another viewpoint, there may be no need to completely squelch all these things. Lessig called this the principle of bovinity: a small number of rules, consistently applied, suffice to control a herd of large animals. There is no need to assure that all people in all contexts continue to behave as couch potatoes for the true scope of the networked information economy to be constrained. It is enough that the core enabling technologies and the core cultural practices are confined to small groups - some teenagers, some countercultural activists. There have been places like the East Village or the Left Bank throughout the period of the industrial information economy. For the gains in autonomy, democracy, justice, and a critical culture that are described in part II to materialize, the practices of nonmarket information production, individually free creation, and cooperative peer production must become more than fringe practices. They must become a part of life for substantial portions of the networked population. The battle over the institutional ecology of the digitally networked environment is waged precisely over how many individual users will continue to participate in making the networked information environment, and how much of the population of consumers will continue to sit on the couch and passively receive the finished goods of industrial information producers.

Institutional Ecology and Path Dependence

The century-old pragmatist turn in American legal thought has led to the development of a large and rich literature about the relationship of law to society and economy. It has both Right and Left versions, and has disciplinary roots in history, economics, sociology, psychology, and critical theory. Explanations are many: some simple, some complex; some analytically tractable, many not. I do not make a substantive contribution to that debate here, but rather build on some of its strains to suggest that the process is complex, and particularly, that the relationship of law to social relations is one of punctuated equilibrium - there are periods of stability followed by periods of upheaval, and then adaptation and stabilization anew, until the next cycle. Hopefully, the preceding ten chapters have provided sufficient reason to think that we are going through a moment of social-economic transformation today, rooted in a technological shock to our basic modes of information, knowledge, and cultural production. Most of this chapter offers a sufficient description of the legislative and judicial battles of the past few years to make the case that we are in the midst of a significant perturbation of some sort. I suggest that the heightened activity is, in fact, a battle, in the domain of law and policy, over the shape of the social settlement that will emerge around the digital computation and communications revolution.

The basic claim is made up of fairly simple components. First, law affects human behavior on a micromotivational level and on a macro-social-organizational level. This is in contradistinction to, on the one hand, the classical Marxist claim that law is epiphenomenal, and, on the other hand, the increasingly rare simple economic models that ignore transaction costs and institutional barriers and

simply assume that people will act in order to maximize their welfare, irrespective of institutional arrangements. Second, the causal relationship between law and human behavior is complex. Simple deterministic models of the form "if law X, then behavior Y" have been used as assumptions, but these are widely understood as, and criticized for being, oversimplifications for methodological purposes. Laws do affect human behavior by changing the payoffs to regulated actions directly. However, they also shape social norms with regard to behaviors, psychological attitudes toward various behaviors, the cultural understanding of actions, and the politics of claims about behaviors and practices. These effects are not all linearly additive. Some push back and nullify the law, some amplify its effects; it is not always predictable which of these any legal change will be. Decreasing the length of a "Walk" signal to assure that pedestrians are not hit by cars may trigger wider adoption of jaywalking as a norm, affecting ultimate behavior in exactly the opposite direction of what was intended. This change may, in turn, affect enforcement regarding jaywalking, or the length of the signals set for cars, because the risks involved in different signal lengths change as actual expected behavior changes, which again may feed back on driving and walking practices. Third, and as part of the complexity of the causal relation, the effects of law differ in different material, social, and cultural contexts. The same law introduced in different societies or at different times will have different effects. It may enable and disable a different set of practices, and trigger a different cascade of feedback and countereffects. This is because human beings are diverse in their motivational structure and their cultural frames of meaning for behavior, for law, or for outcomes. Fourth, the process of lawmaking is not exogenous to the effects of law on social relations and human behavior. One can look at positive political theory or at the history of social movements to see that the shape of law itself is contested in society because it makes (through its complex causal mechanisms) some behaviors less attractive, valuable, or permissible, and others more so. The "winners" and the "losers" battle each other to tweak the institutional playing field to fit their needs. As a consequence of these, there is relatively widespread acceptance that there is path dependence in institutions and social organization. That is, the actual organization of human affairs and legal systems is not converging through a process of either Marxist determinism or its neoclassical economics mirror image, "the most efficient institutions win out in the end." Different societies will differ in initial conditions and their historically contingent first moves in response to similar perturbations, and variances will emerge in their actual practices and institutional arrangements that persist over time - irrespective of their relative inefficiency or injustice.

The term "institutional ecology" refers to this context-dependent, causally complex, feedback-ridden, path-dependent process. An example of this interaction in the area of communications practices is the description in chapter 6 of how the introduction of radio was received and embedded in different legal and economic systems early in the twentieth century. A series of organizational and institutional choices converged in all nations on a broadcast model, but the American broadcast model, the BBC model, and the state-run monopoly radio models created very different journalistic styles, consumption expectations and styles, and funding mechanisms in these various systems. These differences, rooted in a series of choices made during a short period in the 1920s, persisted for decades in each of the respective systems. Paul Starr has argued in *The Creation of the Media* that basic institutional choices - from postage pricing to freedom of the press - interacted with cultural practices and political culture to underwrite substantial differences in the print media of the United States, Britain, and much of the European continent in the late eighteenth and throughout much of the nineteenth centuries.[1] Again, the basic institutional and cultural practices were put in place around the time of the American Revolution, and were later overlaid with the introduction of mass-circulation presses and the telegraph in the mid-1800s. Ithiel de Sola Pool's *Technologies of Freedom* describes the battle between newspapers and telegraph operators in the United States and Britain over control of telegraphed news flows. In Britain, this resulted in the nationalization of telegraph and the continued dominance of London and *The Times*.

In the United States, it resolved into the pooling model of the Associated Press, based on private lines for news delivery and sharing - the prototype for newspaper chains and later network-television models of mass media.[/2](#) The possibility of multiple stable equilibria alongside each other evoked by the stories of radio and print media is a common characteristic to both ecological models and analytically tractable models of path dependency. Both methodological approaches depend on feedback effects and therefore suggest that for any given path divergence, there is a point in time where early actions that trigger feedbacks can cause large and sustained differences over time.

Systems that exhibit path dependencies are characterized by periods of relative pliability followed by periods of relative stability. Institutions and social practices coevolve through a series of adaptations - feedback effects from the institutional system to social, cultural, and psychological frameworks; responses into the institutional system; and success and failure of various behavioral patterns and belief systems - until a society reaches a stage of relative stability. It can then be shaken out of that stability by external shocks - like Admiral Perry's arrival in Japan - or internal buildup of pressure to a point of phase transition, as in the case of slavery in the United States. Of course, not all shocks can so neatly be categorized as external or internal - as in the case of the Depression and the New Deal. To say that there are periods of stability is not to say that in such periods, everything is just dandy for everyone. It is only to say that the political, social, economic settlement is too widely comfortable for, accepted or acquiesced in, by too many agents who in that society have the power to change practices for institutional change to have substantial effects on the range of lived human practices.

The first two parts of this book explained why the introduction of digital computer-communications networks presents a perturbation of transformative potential for the basic model of information production and exchange in modern complex societies. They focused on the technological, economic, and social patterns that are emerging, and how they differ from the industrial information economy that preceded them. This chapter offers a fairly detailed map of how law and policy are being tugged and pulled in response to these changes. Digital computers and networked communications as a broad category will not be rolled back by these laws. Instead, we are seeing a battle - often but not always self-conscious - over the precise shape of these technologies. More important, we are observing a series of efforts to shape the social and economic practices as they develop to take advantage of these new technologies.

A Framework for Mapping the Institutional Ecology

Two specific examples will illustrate the various levels at which law can operate to shape the use of information and its production and exchange. The first example builds on the story from chapter 7 of how embarrassing internal e-mails from Diebold, the electronic voting machine maker, were exposed by investigative journalism conducted on a nonmarket and peer-production model. After students at Swarthmore College posted the files, Diebold made a demand under the DMCA that the college remove the materials or face suit for contributory copyright infringement. The students were therefore forced to remove the materials. However, in order keep the materials available, the students asked students at other institutions to mirror the files, and injected them into the eDonkey, BitTorrent, and FreeNet file-sharing and publication networks. Ultimately, a court held that the unauthorized publication of files that were not intended for sale and carried such high public value was a fair use. This meant that the underlying publication of the files was not itself a violation, and therefore the Internet service provider was not liable for providing a conduit. However, the case was decided on September 30, 2004 - long after the information would have been relevant to the voting equipment certification process in California. What kept the

information available for public review was not the ultimate vindication of the students' publication. It was the fact that the materials were kept in the public sphere even under threat of litigation. Recall also that at least some of the earlier set of Diebold files that were uncovered by the activist who had started the whole process in early 2003 were zipped, or perhaps encrypted in some form. Scoop, the Web site that published the revelation of the initial files, published - along with its challenge to the Internet community to scour the files and find holes in the system - links to locations in which utilities necessary for reading the files could be found.

There are four primary potential points of failure in this story that could have conspired to prevent the revelation of the Diebold files, or at least to suppress the peer-produced journalistic mode that made them available. First, if the service provider-the college, in this case - had been a sole provider with no alternative physical transmission systems, its decision to block the materials under threat of suit would have prevented publication of the materials throughout the relevant period. Second, the existence of peer-to-peer networks that overlay the physical networks and were used to distribute the materials made expunging them from the Internet practically impossible. There was no single point of storage that could be locked down. This made the prospect of threatening other universities futile. Third, those of the original files that were not in plain text were readable with software utilities that were freely available on the Internet, and to which Scoop pointed its readers. This made the files readable to many more critical eyes than they otherwise would have been. Fourth, and finally, the fact that access to the raw materials - the e-mails - was ultimately found to be privileged under the fair-use doctrine in copyright law allowed all the acts that had been performed in the preceding period under a shadow of legal liability to proceed in the light of legality.

The second example does not involve litigation, but highlights more of the levers open to legal manipulation. In the weeks preceding the American-led invasion of Iraq, a Swedish video artist produced an audio version of Diana Ross and Lionel Richie's love ballad, "Endless Love," lip-synched to news footage of U.S. president George Bush and British prime minister Tony Blair. By carefully synchronizing the lip movements from the various news clips, the video produced the effect of Bush "singing" Richie's part, and Blair "singing" Ross's, serenading each other with an eternal love ballad. No legal action with regard to the release of this short video has been reported. However, the story adds two components not available in the context of the Diebold files context. First, it highlights that quotation from video and music requires actual copying of the digital file. Unlike text, you cannot simply transcribe the images or the sound. This means that access to the unencrypted bits is more important than in the case of text. Second, it is not at all clear that using the entire song, unmodified, is a "fair use." While it is true that the Swedish video is unlikely to cut into the market for the original song, there is nothing in the video that is a parody either of the song itself or of the news footage. The video uses "found materials," that is, materials produced by others, to mix them in a way that is surprising, creative, and creates a genuinely new statement. However, its use of the song is much more complete than the minimalist uses of digital sampling in recorded music, where using a mere two-second, three-note riff from another's song has been found to be a violation unless done with a negotiated license./3

Combined, the two stories suggest that we can map the resources necessary for a creative communication, whether produced on a market model or a nonmarket model, as including a number of discrete elements. First, there is the universe of "content" itself: existing information, cultural artifacts and communications, and knowledge structures. These include the song and video footage, or the e-mail files, in the two stories. Second, there is the cluster of machinery that goes into capturing, manipulating, fixing and communicating the new cultural utterances or communications made of these inputs, mixed with the creativity, knowledge, information, or

communications capacities of the creator of the new statement or communication. These include the physical devices - the computers used by the students and the video artist, as well as by their readers or viewers - and the physical transmission mechanisms used to send the information or communications from one place to another. In the Diebold case, the firm tried to use the Internet service provider liability regime of the DMCA to cut off the machine storage and mechanical communications capacity provided to the students by the university. However, the "machinery" also includes the logical components - the software necessary to capture, read or listen to, cut, paste, and remake the texts or music; the software and protocols necessary to store, retrieve, search, and communicate the information across the Internet.

As these stories suggest, freedom to create and communicate requires use of diverse things and relationships - mechanical devices and protocols, information, cultural materials, and so forth. Because of this diversity of components and relationships, the institutional ecology of information production and exchange is a complex one. It includes regulatory and policy elements that affect different industries, draw on various legal doctrines and traditions, and rely on diverse economic and political theories and practices. It includes social norms of sharing and consumption of things conceived of as quite different - bandwidth, computers, and entertainment materials. To make these cohere into a single problem, for several years I have been using a very simple, three-layered representation of the basic functions involved in mediated human communications. These are intended to map how different institutional components interact to affect the answer to the basic questions that define the normative characteristics of a communications system - who gets to say what, to whom, and who decides?/4

These are the physical, logical, and content layers. The physical layer refers to the material things used to connect human beings to each other. These include the computers, phones, handhelds, wires, wireless links, and the like. The content layer is the set of humanly meaningful statements that human beings utter to and with one another. It includes both the actual utterances and the mechanisms, to the extent that they are based on human communication rather than mechanical processing, for filtering, accreditation, and interpretation. The logical layer represents the algorithms, standards, ways of translating human meaning into something that machines can transmit, store, or compute, and something that machines process into communications meaningful to human beings. These include standards, protocols, and software - both general enabling platforms like operating systems, and more specific applications. A mediated human communication must use all three layers, and each layer therefore represents a resource or a pathway that the communication must use or traverse in order to reach its intended destination. In each and every one of these layers, we have seen the emergence of technical and practical capabilities for using that layer on a nonproprietary model that would make access cheaper, less susceptible to control by any single party or class of parties, or both. In each and every layer, we have seen significant policy battles over whether these nonproprietary or open-platform practices will be facilitated or even permitted. Looking at the aggregate effect, we see that at all these layers, a series of battles is being fought over the degree to which some minimal set of basic resources and capabilities necessary to use and participate in constructing the information environment will be available for use on a nonproprietary, nonmarket basis.

In each layer, the policy debate is almost always carried out in local, specific terms. We ask questions like, Will this policy optimize "spectrum management" in these frequencies, or, Will this decrease the number of CDs sold? However, the basic, overarching question that we must learn to ask in all these debates is: Are we leaving enough institutional space for the social-economic practices of networked information production to emerge? The networked information economy requires access to a core set of capabilities - existing information and culture, mechanical means to process,

store, and communicate new contributions and mixes, and the logical systems necessary to connect them to each other. What nonmarket forms of production need is a core common infrastructure that anyone can use, irrespective of whether their production model is market-based or not, proprietary or not. In almost all these dimensions, the current trajectory of technological-economic-social trends is indeed leading to the emergence of such a core common infrastructure, and the practices that make up the networked information economy are taking advantage of open resources. Wireless equipment manufacturers are producing devices that let users build their own networks, even if these are now at a primitive stage. The open-innovation ethos of the programmer and Internet engineering community produce both free software and proprietary software that rely on open standards for providing an open logical layer. The emerging practices of free sharing of information, knowledge, and culture that occupy most of the discussion in this book are producing an ever-growing stream of freely and openly accessible content resources. The core common infrastructure appears to be emerging without need for help from a guiding regulatory hand. This may or may not be a stable pattern. It is possible that by some happenstance one or two firms, using one or two critical technologies, will be able to capture and control a bottleneck. At that point, perhaps regulatory intervention will be required. However, from the beginning of legal responses to the Internet and up to this writing in the middle of 2005, the primary role of law has been reactive and reactionary. It has functioned as a point of resistance to the emergence of the networked information economy. It has been used by incumbents from the industrial information economies to contain the risks posed by the emerging capabilities of the networked information environment. What the emerging networked information economy therefore needs, in almost all cases, is not regulatory protection, but regulatory abstinence.

The remainder of this chapter provides a more or less detailed presentation of the decisions being made at each layer, and how they relate to the freedom to create, individually and with others, without having to go through proprietary, market-based transactional frameworks. Because so many components are involved, and so much has happened since the mid-1990s, the discussion is of necessity both long in the aggregate and truncated in each particular category. To overcome this expositional problem, I have collected the various institutional changes in table 11.1.

Table 11.1: Overview of the Institutional Ecology

	Enclosure	Openness
Physical Transport	• Broadband treated by FCC as information service • DMCA ISP liability • Municipal broadband barred by states	• Open wireless networks • Municipal broadband initiatives
Physical Devices	• CBDPTA: regulatory requirements to implement "trusted systems"; private efforts toward the same goal • Operator-controlled mobile phones	• Standardization • Fiercely competitive market in commodity components

Logical Transmission protocols	• Privatized DNS/ICANN	• TCP/IP • IETF • p2p networks
Logical Software	• DMCA anticircumvention; • Proprietary OS; • Web browser • Software patents	• Free software • W3C • P2p software widely used • social acceptability of widespread hacking of copy protection

Content	• Copyright expansion
	  ○ "Right to read"
	  ○ No de minimis digital sampling
	  ○ "Fair use" narrowed: effect on potential market "commercial" defined broadly
	  ○ Criminalization
	  ○ Term extension
	• Contractual enclosure: UCITA
	• Trademark dilution
	• Database protection
	• Linking and trespass to chattels
	• International "harmonization" and trade enforcement of maximal exclusive rights regimes
		• Increasing sharing practices and adoption of sharing licensing practices
		• Musicians distribute music freely
		• Creative Commons; other open publication models
		• Widespread social disdain for copyright
		• International jurisdictional arbitrage
		• Early signs of a global access to knowledge movement combining developing nations with free information ecology advocates, both market and nonmarket, raising a challenge to the enclosure movement

For readers interested only in the overarching claim of this chapter - that is, that there is, in fact, a battle over the institutional environment, and that many present choices interact to increase or decrease the availability of basic resources for information production and exchange - table 11.1 may provide sufficient detail. For those interested in a case study of the complex relationship between law, technology, social behavior, and market structure, the discussion of peer-to-peer networks may be particularly interesting to pursue.

A quick look at table 11.1 reveals that there is a diverse set of sources of openness. A few of these are legal. Mostly, they are based on technological and social practices, including resistance to legal

and regulatory drives toward enclosure. Examples of policy interventions that support an open core common infrastructure are the FCC's increased permission to deploy open wireless networks and the various municipal broadband initiatives. The former is a regulatory intervention, but its form is largely removal of past prohibitions on an entire engineering approach to building wireless systems. Municipal efforts to produce open broadband networks are being resisted at the state legislation level, with statutes that remove the power to provision broadband from the home rule powers of municipalities. For the most part, the drive for openness is based on individual and voluntary cooperative action, not law. The social practices of openness take on a quasi-normative face when practiced in standard-setting bodies like the Internet Engineering Task Force (IETF) or the World Wide Web Consortium (W3C). However, none of these have the force of law. Legal devices also support openness when used in voluntaristic models like free software licensing and Creative Commons-type licensing. However, most often when law has intervened in its regulatory force, as opposed to its contractual-enablement force, it has done so almost entirely on the side of proprietary enclosure.

Another characteristic of the social-economic-institutional struggle is an alliance between a large number of commercial actors and the social sharing culture. We see this in the way that wireless equipment manufacturers are selling into a market of users of WiFi and similar unlicensed wireless devices. We see this in the way that personal computer manufacturers are competing over decreasing margins by producing the most general-purpose machines that would be most flexible for their users, rather than machines that would most effectively implement the interests of Hollywood and the recording industry. We see this in the way that service and equipment-based firms, like IBM and Hewlett-Packard (HP), support open-source and free software. The alliance between the diffuse users and the companies that are adapting their business models to serve them as users, instead of as passive consumers, affects the political economy of this institutional battle in favor of openness. On the other hand, security consciousness in the United States has led to some efforts to tip the balance in favor of closed proprietary systems, apparently because these are currently perceived as more secure, or at least more amenable to government control. While orthogonal in its political origins to the battle between proprietary and commons-based strategies for information production, this drive does tilt the field in favor of enclosure, at least at the time of this writing in 2005.

Over the past few years, we have also seen that the global character of the Internet is a major limit on effective enclosure, when openness is a function of technical and social practices, and enclosure is a function of law./5 When Napster was shut down in the United States, for example, KaZaa emerged in the Netherlands, from where it later moved to Australia. This force is meeting the countervailing force of international harmonization - a series of bilateral and multilateral efforts to "harmonize" exclusive rights regimes internationally and efforts to coordinate international enforcement. It is difficult at this stage to predict which of these forces will ultimately have the upper hand. It is not too early to map in which direction each is pushing. And it is therefore not too early to characterize the normative implications of the success or failure of these institutional efforts.

The Physical Layer

The physical layer encompasses both transmission channels and devices for producing and communicating information. In the broadcast and telephone era, devices were starkly differentiated. Consumers owned dumb terminals. Providers owned sophisticated networks and equipment: transmitters and switches. Consumers could therefore consume whatever providers could produce most efficiently that the providers believed consumers would pay for. Central to the emergence of

the freedom of users in the networked environment is an erosion of the differentiation between consumer and provider equipment. Consumers came to use general-purpose computers that could do whatever their owners wanted, instead of special-purpose terminals that could only do what their vendors designed them to do. These devices were initially connected over a transmission network - the public phone system - that was regulated as a common carrier. Common carriage required the network owners to carry all communications without differentiating by type or content. The network was neutral as among communications. The transition to broadband networks, and to a lesser extent the emergence of Internet services on mobile phones, are threatening to undermine that neutrality and nudge the network away from its end-to-end, user-centric model to one designed more like a five-thousand-channel broadcast model. At the same time, Hollywood and the recording industry are pressuring the U.S. Congress to impose regulatory requirements on the design of personal computers so that they can be relied on not to copy music and movies without permission. In the process, the law seeks to nudge personal computers away from being purely general-purpose computation devices toward being devices with factory-defined behaviors vis-à-vis predicted-use patterns, like glorified televisions and CD players. The emergence of the networked information economy as described in this book depends on the continued existence of an open transport network connecting general-purpose computers. It therefore also depends on the failure of the efforts to restructure the network on the model of proprietary networks connecting terminals with sufficiently controlled capabilities to be predictable and well behaved from the perspective of incumbent production models.

Transport: Wires and Wireless

Recall the Cisco white paper quoted in chapter 5. In it, Cisco touted the value of its then new router, which would allow a broadband provider to differentiate streams of information going to and from the home at the packet level. If the packet came from a competitor, or someone the user wanted to see or hear but the owner preferred that the user did not, the packet could be slowed down or dropped. If it came from the owner or an affiliate, it could be speeded up. The purpose of the router was not to enable evil control over users. It was to provide better-functioning networks. America Online (AOL), for example, has been reported as blocking its users from reaching Web sites that have been advertised in spam e-mails. The theory is that if spammers know their Web site will be inaccessible to AOL customers, they will stop.[/6] The ability of service providers to block sites or packets from certain senders and promote packets from others may indeed be used to improve the network. However, whether this ability will in fact be used to improve service depends on the extent to which the interests of all users, and particularly those concerned with productive uses of the network, are aligned with the interests of the service providers. Clearly, when in 2005 Telus, Canada's second largest telecommunications company, blocked access to the Web site of the Telecommunications Workers Union for all of its own clients and those of internet service providers that relied on its backbone network, it was not seeking to improve service for those customers' benefit, but to control a conversation in which it had an intense interest. When there is a misalignment, the question is what, if anything, disciplines the service providers' use of the technological capabilities they possess? One source of discipline would be a genuinely competitive market. The transition to broadband has, however, severely constrained the degree of competition in Internet access services. Another would be regulation: requiring owners to treat all packets equally. This solution, while simple to describe, remains highly controversial in the policy world. It has strong supporters and strong opposition from the incumbent broadband providers, and has, as a practical matter, been rejected for the time being by the FCC. The third type of solution would be both more radical and less "interventionist" from the perspective of regulation. It would involve eliminating contemporary regulatory barriers to the emergence of a user-owned wireless infrastructure. It would allow users to deploy their own equipment, share their wireless capacity, and

create a "last mile" owned by all users in common, and controlled by none. This would, in effect, put equipment manufacturers in competition to construct the "last mile" of broadband networks, and thereby open up the market in "middle-mile" Internet connection services.

Since the early 1990s, when the Clinton administration announced its "Agenda for Action" for what was then called "the information superhighway," it was the policy of the United States to "let the private sector lead" in deployment of the Internet. To a greater or lesser degree, this commitment to private provisioning was adopted in most other advanced economies in the world. In the first few years, this meant that investment in the backbone of the Internet was private, and heavily funded by the stock bubble of the late 1990s. It also meant that the last distribution bottleneck - the "last mile" - was privately owned. Until the end of the 1990s, the last mile was made mostly of dial-up connections over the copper wires of the incumbent local exchange carriers. This meant that the physical layer was not only proprietary, but that it was, for all practical purposes, monopolistically owned. Why, then, did the early Internet nonetheless develop into a robust, end-to-end neutral network? As Lessig showed, this was because the telephone carriers were regulated as common carriers. They were required to carry all traffic without discrimination. Whether a bit stream came from Cable News Network (CNN) or from an individual blog, all streams - upstream from the user and downstream to the user - were treated neutrally.

Broadband Regulation

The end of the 1990s saw the emergence of broadband networks. In the United States, cable systems, using hybrid fiber-coaxial systems, moved first, and became the primary providers. The incumbent local telephone carriers have been playing catch-up ever since, using digital subscriber line (DSL) techniques to squeeze sufficient speed out of their copper infrastructure to remain competitive, while slowly rolling out fiber infrastructure closer to the home. As of 2003, the incumbent cable carriers and the incumbent local telephone companies accounted for roughly 96 percent of all broadband access to homes and small offices.[7] In 1999-2000, as cable was beginning to move into a more prominent position, academic critique began to emerge, stating that the cable broadband architecture could be manipulated to deviate from the neutral, end-to-end architecture of the Internet. One such paper was written by Jerome Saltzer, one of the authors of the paper that originally defined the "end-to-end" design principle of the Internet in 1980, and Lessig and Mark Lemley wrote another. These papers began to emphasize that cable broadband providers technically could, and had commercial incentive to, stop treating all communications neutrally. They could begin to move from a network where almost all functions are performed by user-owned computers at the ends of the network to one where more is done by provider equipment at the core. The introduction of the Cisco policy router was seen as a stark marker of how things could change.

The following two years saw significant regulatory battles over whether the cable providers would be required to behave as commons carriers. In particular, the question was whether they would be required to offer competitors nondiscriminatory access to their networks, so that these competitors could compete in Internet services. The theory was that competition would discipline the incumbents from skewing their networks too far away from what users valued as an open Internet. The first round of battles occurred at the municipal level. Local franchising authorities tried to use

their power over cable licenses to require cable operators to offer open access to their competitors if they chose to offer cable broadband. The cable providers challenged these regulations in courts. The most prominent decision came out of Portland, Oregon, where the Federal Court of Appeals for the Ninth Circuit held that broadband was part information service and part telecommunications service, but not a cable service. The FCC, not the cable franchising authority, had power to regulate it./8 At the same time, as part of the approval of the AOL-Time Warner merger, the Federal Trade Commission (FTC) required the new company to give at least three competitors open access to its broadband facilities, should AOL be offered cable broadband facilities over Time Warner.

The AOL-Time Warner merger requirements, along with the Ninth Circuit's finding that cable broadband included a telecommunications component, seemed to indicate that cable broadband transport would come to be treated as a common carrier. This was not to be. In late 2001 and the middle of 2002, the FCC issued a series of reports that would reach the exact opposite result. Cable broadband, the commission held, was an information service, not a telecommunications service. This created an imbalance with the telecommunications status of broadband over telephone infrastructure, which at the time was treated as a telecommunications service. The commission dealt with this imbalance by holding that broadband over telephone infrastructure, like broadband over cable, was now to be treated as an information service. Adopting this definition was perhaps admissible as a matter of legal reasoning, but it certainly was not required by either sound legal reasoning or policy. The FCC's reasoning effectively took the business model that cable operators had successfully used to capture two-thirds of the market in broadband - bundling two discrete functionalities, transport (carrying bits) and higher-level services (like e-mail and Web hosting) - and treated it as though it described the intrinsic nature of "broadband cable" as a service. Because that service included more than just carriage of bits, it could be called an information service. Of course, it would have been as legally admissible, and more technically accurate, to do as the Ninth Circuit had done. That is, to say that cable broadband bundles two distinct services: carriage and information-use tools. The former is a telecommunications service. In June of 2005, the Supreme Court in the *Brand X* case upheld the FCC's authority to make this legally admissible policy error, upholding as a matter of deference to the expert agency the Commission's position that cable broadband services should be treated as information services./9 As a matter of policy, the designation of broadband services as "information services" more or less locked the FCC into a "no regulation" approach. As information services, broadband providers obtained the legal power to "edit" their programming, just like any operator of an information service, like a Web site. Indeed, this new designation has placed a serious question mark over whether future efforts to regulate carriage decisions would be considered constitutional, or would instead be treated as violations of the carriers' "free speech" rights as a provider of information. Over the course of the 1990s, there were a number of instances where carriers - particularly cable, but also telephone companies - were required by law to carry some signals from competitors. In particular, cable providers were required to carry over-the-air broadcast television, telephone carriers, in FCC rules called "video dialtone," were required to offer video on a common carriage basis, and cable providers that chose to offer broadband were required to make their infrastructure available to competitors on a common carrier model. In each of these cases, the carriage requirements were subjected to First Amendment scrutiny by courts. In the case of cable carriage of broadcast television, the carriage requirements were only upheld after six years of litigation./10 In cases involving video common carriage requirements applied

to telephone companies and cable broadband, lower courts struck down the carriage requirements as violating the telephone and cable companies' free-speech rights. /11 To a large extent, then, the FCC's regulatory definition left the incumbent cable and telephone providers - who control 96 percent of broadband connections to home and small offices - unregulated, and potentially constitutionally immune to access regulation and carriage requirements.

Since 2003 the cable access debate - over whether competitors should get access to the transport networks of incumbent broadband carriers - has been replaced with an effort to seek behavioral regulation in the form of "network neutrality." This regulatory concept would require broadband providers to treat all packets equally, without forcing them to open their network up to competitors or impose any other of the commitments associated with common carriage. The concept has the backing of some very powerful actors, including Microsoft, and more recently MCI, which still owns much of the Internet backbone, though not the last mile. For this reason, if for no other, it remains as of this writing a viable path for institutional reform that would balance the basic structural shift of Internet infrastructure from a common-carriage to a privately controlled model. Even if successful, the drive to network neutrality would keep the physical infrastructure a technical bottleneck, owned by a small number of firms facing very limited competition, with wide legal latitude for using that control to affect the flow of information over their networks.

Open Wireless Networks

A more basic and structural opportunity to create an open broadband infrastructure is, however, emerging in the wireless domain. To see how, we must first recognize that opportunities to control the broadband infrastructure in general are not evenly distributed throughout the networked infrastructure. The long-haul portions of the network have multiple redundant paths with no clear choke points. The primary choke point over the physical transport of bits across the Internet is in the last mile of all but the most highly connected districts. That is, the primary bottleneck is the wire or cable connecting the home and small office to the network. It is here that cable and local telephone incumbents control the market. It is here that the high costs of digging trenches, pulling fiber, and getting wires through and into walls pose a prohibitive barrier to competition. And it is here, in the last mile, that unlicensed wireless approaches now offer the greatest promise to deliver a common physical infrastructure of first and last resort, owned by its users, shared as a commons, and offering no entity a bottleneck from which to control who gets to say what to whom.

As discussed in chapter 6, from the end of World War I and through the mid-twenties, improvements in the capacity of expensive transmitters and a series of strategic moves by the owners of the core patents in radio transmission led to the emergence of the industrial model of radio communications that typified the twentieth century. Radio came to be dominated by a small number of professional, commercial networks, based on high-capital-cost transmitters. These were supported by a regulatory framework tailored to making the primary model of radio utilization for most Americans passive reception, with simple receivers, of commercial programming delivered with high-powered transmitters. This industrial model, which assumed large-scale capital investment in the core of the network and small-scale investments at the edges, optimized for receiving what is generated at the core, imprinted on wireless communications systems both at the level of design and

at the level of regulation. When mobile telephony came along, it replicated the same model, using relatively cheap handsets oriented toward an infrastructure-centric deployment of towers. The regulatory model followed Hoover's initial pattern and perfected it. A government agency strictly controlled who may place a transmitter, where, with what antenna height, and using what power. The justification was avoidance of interference. The presence of strict licensing was used as the basic assumption in the engineering of wireless systems throughout this period. Since 1959, economic analysis of wireless regulation has criticized this approach, but only on the basis that it inefficiently regulated the legal right to construct a wireless system by using strictly regulated spectrum licenses, instead of creating a market in "spectrum use" rights./12 This critique kept the basic engineering assumptions stable - for radio to be useful, a high-powered transmitter must be received by simple receivers. Given this engineering assumption, someone had to control the right to emit energy in any range of radio frequencies. The economists wanted the controller to be a property owner with a flexible, transferable right. The regulators wanted it to be a licensee subject to regulatory oversight and approval by the FCC.

As chapter 3 explained, by the time that legislatures in the United States and around the world had begun to accede to the wisdom of the economists' critique, it had been rendered obsolete by technology. In particular, it had been rendered obsolete by the fact that the declining cost of computation and the increasing sophistication of communications protocols among end-user devices in a network made possible new, sharing-based solutions to the problem of how to allow users to communicate without wires. Instead of having a regulation-determined exclusive right to transmit, which may or may not be subject to market reallocation, it is possible to have a market in smart radio equipment owned by individuals. These devices have the technical ability to share capacity and cooperate in the creation of wireless carriage capacity. These radios can, for example, cooperate by relaying each other's messages or temporarily "lending" their antennae to neighbors to help them decipher messages of senders, without anyone having exclusive use of the spectrum. Just as PCs can cooperate to create a supercomputer in SETI@Home by sharing their computation, and a global-scale, peer-to-peer data-storage and retrieval system by sharing their hard drives, computationally intensive radios can share their capacity to produce a local wireless broadband infrastructure. Open wireless networks allow users to install their own wireless device - much like the WiFi devices that have become popular. These devices then search automatically for neighbors with similar capabilities, and self-configure into a high-speed wireless data network. Reaching this goal does not, at this point, require significant technological innovation. The technology is there, though it does require substantial engineering effort to implement. The economic incentives to develop such devices are fairly straightforward. Users already require wireless local networks. They will gain added utility from extending their range for themselves, which would be coupled with the possibility of sharing with others to provide significant wide-area network capacity for whose availability they need not rely on any particular provider. Ultimately, it would be a way for users to circumvent the monopoly last mile and recapture some of the rents they currently pay. Equipment manufacturers obviously have an incentive to try to cut into the rents captured by the broadband monopoly/oligopoly by offering an equipment-embedded alternative.

My point here is not to consider the comparative efficiency of a market in wireless licenses and a market in end-user equipment designed for sharing channels that no one owns. It is to highlight the

implications of the emergence of a last mile that is owned by no one in particular, and is the product of cooperation among neighbors in the form of, "I'll carry your bits if you carry mine." At the simplest level, neighbors could access locally relevant information directly, over a wide-area network. More significant, the fact that users in a locality coproduced their own last-mile infrastructure would allow commercial Internet providers to set up Internet points of presence anywhere within the "cloud" of the locale. The last mile would be provided not by these competing Internet service providers, but by the cooperative efforts of the residents of local neighborhoods. Competitors in providing the "middle mile" - the connection from the last mile to the Internet cloud - could emerge, in a way that they cannot if they must first lay their own last mile all the way to each home. The users, rather than the middle-mile providers, shall have paid the capital cost of producing the local transmission system - their own cooperative radios. The presence of a commons-based, coproduced last mile alongside the proprietary broadband network eliminates the last mile as a bottleneck for control over who speaks, with what degree of ease, and with what types of production values and interactivity.

The development of open wireless networks, owned by their users and focused on sophisticated general-purpose devices at their edges also offers a counterpoint to the emerging trend among mobile telephony providers to offer a relatively limited and controlled version of the Internet over the phones they sell. Some wireless providers are simply offering mobile Internet connections throughout their networks, for laptops. Others, however, are using their networks to allow customers to use their ever-more-sophisticated phones to surf portions of the Web. These latter services diverge in their styles. Some tend to be limited, offering only a set of affiliated Web sites rather than genuine connectivity to the Internet itself with a general-purpose device. Sprint's "News" offerings, for example, connects users to CNNtoGo, ABCNews.com, and the like, but will not enable a user to reach the blogosphere to upload a photo of protesters being manhandled, for example. So while mobility in principle increases the power of the Web, and text messaging puts e-mail-like capabilities everywhere, the effect of the implementations of the Web on phones is more ambiguous. It could be more like a Web-enabled reception device than a genuinely active node in a multidirectional network. Widespread adoption of open wireless networks would give mobile phone manufacturers a new option. They could build into the mobile telephones the ability to tap into open wireless networks, and use them as general-purpose access points to the Internet. The extent to which this will be a viable option for the mobile telephone manufacturers depends on how much the incumbent mobile telephone service providers, those who purchased their licenses at high-priced auctions, will resist this move. Most users buy their phones from their providers, not from general electronic equipment stores. Phones are often tied to specific providers in ways that users are not able to change for themselves. In these conditions, it is likely that mobile providers will resist the competition from free open wireless systems for "data minutes" by refusing to sell dual-purpose equipment. Worse, they may boycott manufacturers who make mobile phones that are also general-purpose Web-surfing devices over open wireless networks. How that conflict will go, and whether users would be willing to carry a separate small device to enable them to have open Internet access alongside their mobile phone, will determine the extent to which the benefits of open wireless networks will be transposed into the mobile domain. Normatively, that outcome has significant implications. From the perspective of the citizen watchdog function, ubiquitous availability of capture, rendering, and communication capabilities are important. From the perspective of personal

autonomy as informed action in context, extending openness to mobile units would provide significant advantages to allow individuals to construct their own information environment on the go, as they are confronting decisions and points of action in their daily lives.

Municipal Broadband Initiatives

One alternative path for the emergence of basic physical information transport infrastructure on a nonmarket model is the drive to establish municipal systems. These proposed systems would not be commons-based in the sense that they would not be created by the cooperative actions of individuals without formal structure. They would be public, like highways, sidewalks, parks, and sewage systems. Whether they are, or are not, ultimately to perform as commons would depend on how they would be regulated. In the United States, given the First Amendment constraints on government preferring some speech to other speech in public fora, it is likely that municipal systems would be managed as commons. In this regard, they would have parallel beneficial characteristics to those of open wireless systems. The basic thesis underlying municipal broadband initiatives is similar to that which has led some municipalities to create municipal utilities or transportation hubs. Connectivity has strong positive externalities. It makes a city's residents more available for the information economy and the city itself a more attractive locale for businesses. Most of the efforts have indeed been phrased in these instrumental terms. The initial drive has been the creation of municipal fiber-to-the-home networks. The town of Bristol, Virginia, is an example. It has a population of slightly more than seventeen thousand. Median household income is 68 percent of the national median. These statistics made it an unattractive locus for early broadband rollout by incumbent providers. However, in 2003, Bristol residents had one of the most advanced residential fiber-to-the-home networks in the country, available for less than forty dollars a month. Unsurprisingly, therefore, the city had broadband penetration rivaling many of the top U.S. markets with denser and wealthier populations. The "miracle" of Bristol is that the residents of the town, fed up with waiting for the local telephone and cable companies, built their own, municipally owned network. Theirs has become among the most ambitious and successful of more than five hundred publicly owned utilities in the United States that offer high-speed Internet, cable, and telephone services to their residents. Some of the larger cities - Chicago and Philadelphia, most prominently - are moving as of this writing in a similar direction. The idea in Chicago is that basic "dark fiber" - that is, the physical fiber going to the home, but without the electronics that would determine what kinds of uses the connectivity could be put to - would be built by the city. Access to use this entirely neutral, high-capacity platform would then be open to anyone - commercial and noncommercial alike. The drive in Philadelphia emphasizes the other, more recently available avenue - wireless. The quality of WiFi and the widespread adoption of wireless techniques have moved other municipalities to adopt wireless or mixed-fiber wireless strategies. Municipalities are proposing to use publicly owned facilities to place wireless points of access around the town, covering the area in a cloud of connectivity and providing open Internet access from anywhere in the city. Philadelphia's initiative has received the widest public attention, although other, smaller cities are closer to having a wireless cloud over the city already.

The incumbent broadband providers have not taken kindly to the municipal assault on their monopoly (or oligopoly) profits. When the city of Abilene, Texas, tried to offer municipal

broadband service in the late-1990s, Southwestern Bell (SBC) persuaded the Texas legislature to pass a law that prohibited local governments from providing high-speed Internet access. The town appealed to the FCC and the Federal Court of Appeals in Washington, D.C. Both bodies held that when Congress passed the 1996 Telecommunications Act, and said that, "no state . . . regulation . . . may prohibit . . . the ability of any entity to provide . . . telecommunications service," municipalities were not included in the term "any entity." As the D.C. Circuit put it, "any" might have some significance "depending on the speaker's tone of voice," but here it did not really mean " *any* entity," only some. And states could certainly regulate the actions of municipalities, which are treated in U.S. law as merely their subdivisions or organs.[/13](#) Bristol, Virginia, had to fight off similar efforts to prohibit its plans through state law before it was able to roll out its network. In early 2004, the U.S. Supreme Court was presented with the practice of state preemption of municipal broadband efforts and chose to leave the municipalities to fend for themselves. A coalition of Missouri municipalities challenged a Missouri law that, like the Texas law, prohibited them from stepping in to offer their citizens broadband service. The Court of the Appeals for the Eighth Circuit agreed with the municipalities. The 1996 Act, after all, was intended precisely to allow anyone to compete with the incumbents. The section that prohibited states from regulating the ability of "any entity" to enter the telecommunications service market precisely anticipated that the local incumbents would use their clout in state legislatures to thwart the federal policy of introducing competition into the local loop. Here, the incumbents were doing just that, but the Supreme Court reversed the Eighth Circuit decision. Without dwelling too much on the wisdom of allowing citizens of municipalities to decide for themselves whether they want a municipal system, the court issued an opinion that was technically defensible in terms of statutory interpretation, but effectively invited the incumbent broadband providers to put their lobbying efforts into persuading state legislators to prohibit municipal efforts.[/14](#) After Philadelphia rolled out its wireless plan, it was not long before the Pennsylvania legislature passed a similar law prohibiting municipalities from offering broadband. While Philadelphia's plan itself was grandfathered, future expansion from a series of wireless "hot spots" in open area to a genuine municipal network will likely be challenged under the new state law. Other municipalities in Pennsylvania are entirely foreclosed from pursuing this option. In this domain, at least as of 2005, the incumbents seem to have had some substantial success in containing the emergence of municipal broadband networks as a significant approach to eliminating the bottleneck in local network infrastructure.

Devices

The second major component of the physical layer of the networked environment is comprised of the devices people use to compute and communicate. Personal computers, handhelds, game consoles, and to a lesser extent, but lurking in the background, televisions, are the primary relevant devices. In the United States, personal computers are the overwhelmingly dominant mode of connectivity. In Europe and Japan, mobile handheld devices occupy a much larger space. Game consoles are beginning to provide an alternative computationally intensive device, and Web-TV has been a background idea for a while. The increasing digitization of both over-the-air and cable broadcast makes digital TV a background presence, if not an immediate alternative avenue, to Internet communications. None of these devices are constructed by a commons - in the way that open wireless networks, free software, or peer-produced content can be. Personal computers,

however, are built on open architecture, using highly standardized commodity components and open interfaces in an enormously competitive market. As a practical matter, therefore, PCs provide an open-platform device. Handhelds, game consoles, and digital televisions, on the other hand, use more or less proprietary architectures and interfaces and are produced in a less-competitive market - not because there is no competition among the manufacturers, but because the distribution chain, through the service providers, is relatively controlled. The result is that configurations and features can more readily be customized for personal computers. New uses can be developed and implemented in the hardware without permission from any owner of a manufacturing or distribution outlet. As handhelds grow in their capabilities, and personal computers collapse in size, the two modes of communicating are bumping into each other's turf. At the moment, there is no obvious regulatory push to nudge one or the other out. Observing the evolution of these markets therefore has less to do with policy. As we look at these markets, however, it is important to recognize that the outcome of this competition is not normatively neutral. The capabilities made possible by personal computers underlie much of the social and economic activity described throughout this book. Proprietary handhelds, and even more so, game consoles and televisions, are, presently at least, platforms that choreograph their use. They structure their users' capabilities according to design requirements set by their producers and distributors. A physical layer usable with general-purpose computers is one that is pliable and open for any number of uses by individuals, in a way that a physical layer used through more narrowly scripted devices is not.

The major regulatory threat to the openness of personal computers comes from efforts to regulate the use of copyrighted materials. This question is explored in greater depth in the context of discussing the logical layer. Here, I only note that peer-to-peer networks, and what Fisher has called "promiscuous copying" on the Internet, have created a perceived threat to the very existence of the major players in the industrial cultural production system - Hollywood and the recording industry. These industries are enormously adept at driving the regulation of their business environment - the laws of copyright, in particular. As the threat of copying and sharing of their content by users increased, these industries have maintained a steady pressure on Congress, the courts, and the executive to ratchet up the degree to which their rights are enforced. As we will see in looking at the logical and content layers, these efforts have been successful in changing the law and pushing for more aggressive enforcement. They have not, however, succeeded in suppressing widespread copying. Copying continues, if not entirely unabated, certainly at a rate that was impossible a mere six years ago.

One major dimension of the effort to stop copying has been a drive to regulate the design of personal computers. Pioneered by Senator Fritz Hollings in mid-2001, a number of bills were drafted and lobbied for: the first was the Security Systems Standards and Certification Act; the second, Consumer Broadband and Digital Television Promotion Act (CBDTPA), was actually introduced in the Senate in 2002./15 The basic structure of these proposed statutes was that they required manufacturers to design their computers to be "trusted systems." The term "trusted," however, had a very odd meaning. The point is that the system, or computer, can be trusted to perform in certain predictable ways, irrespective of what its owner wishes. The impulse is trivial to explain. If you believe that most users are using their personal computers to copy films and music illegally, then you can think of these users as untrustworthy. In order to be able to distribute films

and music in the digital environment that is trustworthy, one must disable the users from behaving as they would choose to. The result is a range of efforts at producing what has derisively been called "the Fritz chip": legal mandates that systems be designed so that personal computers cannot run programs that are not certified properly to the chip. The most successful of these campaigns was Hollywood's achievement in persuading the FCC to require manufacturers of all devices capable of receiving digital television signals from the television set to comply with a particular "trusted system" standard. This "broadcast flag" regulation was odd in two distinct ways. First, the rule-making documents show quite clearly that this was a rule driven by Hollywood, not by the broadcasters. This is unusual because the industries that usually play a central role in these rule makings are those regulated by the FCC, such as broadcasters and cable systems. Second, the FCC was not, in fact, regulating the industries that it normally has jurisdiction to regulate. Instead, the rule applied to any device that could use digital television signals *after* they had already been received in the home. In other words, they were regulating practically every computer and digital-video-capable consumer electronics device imaginable. The Court of Appeals ultimately indeed struck down the regulation as wildly beyond the agency's jurisdiction, but the broadcast flag nonetheless is the closest that the industrial information economy incumbents have come to achieving regulatory control over the design of computers.

The efforts to regulate hardware to fit the distribution model of Hollywood and the recording industry pose a significant danger to the networked information environment. The core design principle of general-purpose computers is that they are open for varied uses over time, as their owners change their priorities and preferences. It is this general-purpose character that has allowed personal computers to take on such varied roles since their adoption in the 1980s. The purpose of the Fritz chip-style laws is to make computing devices less flexible. It is to define a range of socially, culturally, and economically acceptable uses of the machines that are predicted by the legislature and the industry actors, and to implement factory-defined capabilities that are not flexible, and do not give end users the freedom to change the intended use over time and to adapt to changing social and economic conditions and opportunities.

The political economy of this regulatory effort, and similar drives that have been more successful in the logical and content layers, is uncharacteristic of American politics. Personal computers, software, and telecommunications services are significantly larger industries than Hollywood and the recording industry. Verizon alone has roughly similar annual revenues to the entire U.S. movie industry. Each one of the industries that the content industries have tried to regulate has revenues several times greater than do the movie and music industries combined. The relative successes of Hollywood and the recording industry in regulating the logical and content layers, and the viability of their efforts to pass a Fritz chip law, attest to the remarkable cultural power of these industries and to their lobbying prowess. The reason is likely historical. The software and hardware industries in particular have developed mostly outside of the regulatory arena; only around 2002 did they begin to understand that what goes on in Washington could really hurt them. The telecommunications carriers, which are some of the oldest hands at the regulatory game, have had some success in preventing regulations that would force them to police their users and limit Internet use. However, the bulk of their lobbying efforts have been aimed elsewhere. The institutions of higher education, which have found themselves under attack for not policing their students' use of peer-to-peer

networks, have been entirely ineffective at presenting their cultural and economic value and the importance of open Internet access to higher education, as compared to the hypothetical losses of Hollywood and the recording industry. Despite the past successes of these entertainment-industry incumbents, two elements suggest that physical device regulation of the CBDPTA form will not follow the same successful path of similar legislation at the logical layer, the DMCA of 1998. The first element is the fact that, unlike in 1998, the technology industries have now realized that Hollywood is seeking to severely constrain their design space. Industries with half a trillion dollars a year in revenues tend to have significant pull in American and international lawmaking bodies, even against industries, like movies and sound recording, that have high cultural visibility but no more than seventy-five billion dollars a year in revenues. The second is that in 1998, there were very few public advocacy organizations operating in the space of intellectual property and trying to play watchdog and to speak for the interests of users. By 2004, a number of organizations dedicated to users' rights in the digital environment emerged to make that conflict clear. The combination of well-defined business interests with increasing representation of user interests creates a political landscape in which it will be difficult to pass sweeping laws to limit the flexibility of personal computers. The most recent iteration of the Fritz chip agenda, the Inducing Infringement of Copyrights Act of 2004 was indeed defeated, for the time being, by a coalition of high-technology firms and people who would have formerly been seen as left-of-center media activists.

Regulation of device design remains at the frontier of the battles over the institutional ecology of the digital environment. It is precisely ubiquitous access to basic, general-purpose computers, as opposed to glorified televisions or telephone handsets, that lies at the very heart of the networked information economy. And it is therefore precisely ubiquitous access to such basic machines that is a precondition to the improvements in freedom and justice that we can see emerging in the digital environment.

The Logical Layer

At the logical layer, most of the efforts aimed to secure a proprietary model and a more tightly controlled institutional ecology follow a similar pattern to the efforts to regulate device design. They come from the needs of the content-layer businesses - Hollywood and the recording industry, in particular. Unlike the physical transmission layer, which is historically rooted in a proprietary but regulated organizational form, most of the logical layer of the Internet has its roots in open, nonproprietary protocols and standards. The broad term "logical layer" combines a wide range of quite different functionalities. The most basic logical components - the basic protocols and standards for Internet connectivity - have from the beginning of the Internet been open, unowned, and used in common by all Internet users and applications. They were developed by computer scientists funded primarily with public money. The basic Internet Protocol (IP) and Transmission Control Protocol (TCP) are open for all to use. Most of the basic standards for communicating were developed in the IETF, a loosely defined standards-setting body that works almost entirely on a meritocratic basis - a body that Michael Froomkin once suggested is the closest earthly approximation of Habermas's ideal speech situation. Individual computer engineers contributed irrespective of formal status or organizational affiliation, and the organization ran on the principle that Dave Clark termed "rough consensus and running code." The World Wide Web protocols and

authoring conventions HTTP and HTML were created, and over the course of their lives, shepherded by Tim Berners Lee, who has chosen to dedicate his efforts to making the Web a public good rather than cashing in on his innovation. The sheer technical necessity of these basic protocols and the cultural stature of their achievement within the engineering community have given these open processes and their commonslike institutional structure a strong gravitational pull on the design of other components of the logical layer, at least insofar as it relates to the communication side of the Internet.

This basic open model has been in constant tension with the proprietary models that have come to use and focus on the Internet in the past decade. By the mid-1990s, the development of graphical-user interfaces to the Web drove Internet use out of universities and into homes. Commercial actors began to look for ways to capture the commercial value of the human potential of the World Wide Web and the Internet, while Hollywood and the recording industry saw the threat of one giant worldwide copying machine looming large. At the same time, the Clinton administration's search of "third-way" liberal agenda manifested in these areas as a commitment to "let the private sector lead" in deployment of the Internet, and an "intellectual property" policy based on extreme protectionism for the exclusive-rights-dependent industries aimed, in the metaphors of that time, to get cars on the information superhighway or help the Internet become a celestial jukebox. The result was a series of moves designed to make the institutional ecology of the Internet more conducive to the proprietary model.

The Digital Millennium Copyright Act of 1998

No piece of legislation more clearly represents the battle over the institutional ecology of the digital environment than the pompously named Digital Millennium Copyright Act of 1998 (DMCA). The DMCA was the culmination of more than three years of lobbying and varied efforts, both domestically in the United States and internationally, over the passage of two WIPO treaties in 1996. The basic worldview behind it, expressed in a 1995 white paper issued by the Clinton administration, was that in order for the National Information Infrastructure (NII) to take off, it had to have "content," and that its great promise was that it could deliver the equivalent of thousands of channels of entertainment. This would only happen, however, if the NII was made safe for delivery of digital content without making it easily copied and distributed without authorization and without payment. The two core recommendations of that early road map were focused on regulating technology and organizational responsibility. First, law was to regulate the development of technologies that might defeat any encryption or other mechanisms that the owners of copyrighted materials would use to prevent use of their works. Second, Internet service providers were to be held accountable for infringements made by their users, so that they would have an incentive to police their systems. Early efforts to pass this agenda in legislation were resisted, primarily by the large telecommunications service providers. The Baby Bells - U.S. regional telephone companies that were created from the breakup of AT&T (Ma Bell) in 1984, when the telecommunications company was split up in order to introduce a more competitive structure to the telecom industry - also played a role in partly defeating implementation of this agenda in the negotiations toward new WIPO treaties in 1996, treaties that ultimately included a much-muted version of the white paper agenda. Nonetheless, the following year saw significant lobbying for

"implementing legislation" to bring U.S. law in line with the requirements of the new WIPO treaties. This new posture placed the emphasis of congressional debates on national industrial policy and the importance of strong protection to the export activities of the U.S. content industries. It was enough to tip the balance in favor of passage of the DMCA. The Internet service provider liability portions bore the marks of a hard-fought battle. The core concerns of the telecommunications companies were addressed by creating an explicit exemption for pure carriage of traffic. Furthermore, providers of more sophisticated services, like Web hosting, were provided immunity from liability for simple failure to police their system actively. In exchange, however, service providers were required to respond to requests by copyright owners by immediately removing materials that the copyright owners deemed infringing. This was the provision under which Diebold forced Swarthmore to remove the embarrassing e-mail records from the students' Web sites. The other, more basic, element of the DMCA was the anticircumvention regime it put in place. Pamela Samuelson has described the anticircumvention provisions of the DMCA as the result of a battle between Hollywood and Silicon Valley. At the time, unlike the telecommunications giants who were born of and made within the regulatory environment, Silicon Valley did not quite understand that what happened in Washington, D.C., could affect its business. The Act was therefore an almost unqualified victory for Hollywood, moderated only by a long list of weak exemptions for various parties that bothered to show up and lobby against it.

The central feature of the DMCA, a long and convoluted piece of legislation, is its anticircumvention and antidevice provisions. These provisions made it illegal to use, develop, or sell technologies that had certain properties. Copyright owners believed that it would be possible to build strong encryption into media products distributed on the Internet. If they did so successfully, the copyright owners could charge for digital distribution and users would not be able to make unauthorized copies of the works. If this outcome was achieved, the content industries could simply keep their traditional business model - selling movies or music as discrete packages - at lower cost, and with a more refined ability to extract the value users got from using their materials. The DMCA was intended to make this possible by outlawing technologies that would allow users to get around, or circumvent, the protection measures that the owners of copyrighted materials put in place. At first blush, this proposition sounds entirely reasonable. If you think of the content of a music file as a home, and of the copy protection mechanism as its lock, then all the DMCA does is prohibit the making and distributing of burglary tools. This is indeed how the legislation was presented by its supporters. From this perspective, even the relatively draconian consequences spelled out in the DMCA's criminal penalties seem defensible.

There are two distinct problems with this way of presenting what the DMCA does. First, copyrights are far from coextensive with real property. There are many uses of existing works that are permissible to all. They are treated in copyright law like walking on the sidewalk or in a public park is treated in property law, not like walking across the land of a neighbor. This is true, most obviously, for older works whose copyright has expired. This is true for certain kinds of uses of a work, like quoting it for purposes of criticism or parody. Encryption and other copy-protection techniques are not limited by the definition of legal rights. They can be used to protect all kinds of digital files - whether their contents are still covered by copyright or not, and whether the uses that users wish to make of them are privileged or not. Circumvention techniques, similarly, can be used

to circumvent copy-protection mechanisms for purposes both legitimate and illegitimate. A barbed wire cutter, to borrow Boyle's metaphor, could be a burglary tool if the barbed wire is placed at the property line. However, it could equally be a tool for exercising your privilege if the private barbed wire has been drawn around public lands or across a sidewalk or highway. The DMCA prohibited all wire cutters, even though there were many uses of these technologies that could be used for legal purposes. Imagine a ten-year-old girl doing her homework on the history of the Holocaust. She includes in her multimedia paper a clip from Steven Spielberg's film, *Schindler's List*, in which a little girl in red, the only color image on an otherwise black-and-white screen, walks through the pandemonium of a deportation. In her project, the child painstakingly superimposes her own face over that of the girl in the film for the entire sequence, frame by frame. She calls the paper, "My Grandmother." There is little question that most copyright lawyers (not retained by the owner of the movie) would say that this use would count as a "fair use," and would be privileged under the Copyright Act. There is also little question that if *Schindler's List* was only available in encrypted digital form, a company would have violated the DMCA if it distributed a product that enabled the girl to get around the encryption in order to use the snippet she needed, and which by traditional copyright law she was permitted to use. It is in the face of this concern about overreaching by those who employ technological protection measures that Julie Cohen argued for the "right to hack" - to circumvent code that impedes one's exercise of one's privileged uses.

The second problem with the DMCA is that its definitions are broad and malleable. Simple acts like writing an academic paper on how the encryption works, or publishing a report on the Web that tells users where they can find information about how to circumvent a copy-protection mechanism could be included in the definition of providing a circumvention device. Edward Felten is a computer scientist at Princeton. As he was preparing to publish an academic paper on encryption, he received a threatening letter from the Recording Industry Association of America (RIAA), telling him that publication of the paper constituted a violation of the DMCA. The music industry had spent substantial sums on developing encryption for digital music distribution. In order to test the system before it actually entrusted music with this wrapper, the industry issued a public challenge, inviting cryptographers to try to break the code. Felten succeeded in doing so, but did not continue to test his solutions because the industry required that, in order to continue testing, he sign a nondisclosure agreement. Felten is an academic, not a businessperson. He works to make knowledge public, not to keep it secret. He refused to sign the nondisclosure agreement, and prepared to publish his initial findings, which he had made without entering any nondisclosure agreement. As he did so, he received the RIAA's threatening letter. In response, he asked a federal district court to declare that publication of his findings was not a violation of the DMCA. The RIAA, realizing that trying to silence academic publication of a criticism of the weakness of its approach to encryption was not the best litigation stance, moved to dismiss the case by promising it would never bring suit./16

Another case did not end so well for the defendant. It involved a suit by the eight Hollywood studios against a hacker magazine, *2600*. The studios sought an injunction prohibiting *2600* from making available a program called DeCSS, which circumvents the copy-protection scheme used to control access to DVDs, named CSS. CSS prevents copying or any use of DVDs unauthorized by the vendor. DeCSS was written by a fifteen-year-old Norwegian named Jon Johanson, who claimed

(though the district court discounted his claim) to have written it as part of an effort to create a DVD player for GNU/Linux-based machines. A copy of DeCSS, together with a story about it was posted on the *2600* site. The industry obtained an injunction against *2600*, prohibiting not only the posting of DeCSS, but also its linking to other sites that post the program - that is, telling users where they can get the program, rather than actually distributing a circumvention program. That decision may or may not have been correct on the merits. There are strong arguments in favor of the proposition that making DVDs compatible with GNU/Linux systems is a fair use. There are strong arguments that the DMCA goes much farther than it needs to in restricting speech of software programmers and Web authors, and so is invalid under the First Amendment. The court rejected these arguments.

The point here is not, however, to revisit the legal correctness of that decision, but to illustrate the effects of the DMCA as an element in the institutional ecology of the logical layer. The DMCA is intended as a strong legal barrier to certain technological paths of innovation at the logical layer of the digital environment. It is intended specifically to preserve the "thing-" or "goods"-like nature of entertainment products - music and movies, in particular. As such, it is intended to, and does to some extent, shape the technological development toward treating information and culture as finished goods, rather than as the outputs of social and communications processes that blur the production-consumption distinction. It makes it more difficult for individuals and nonmarket actors to gain access to digital materials that the technology, the market, and the social practices, left unregulated, would have made readily available. It makes practices of cutting and pasting, changing and annotating existing cultural materials harder to do than the technology would have made possible. I have argued elsewhere that when Congress self-consciously makes it harder for individuals to use whatever technology is available to them, to speak as they please and to whomever they please, in the interest of some public goal (in this case, preservation of Hollywood and the recording industry for the public good), it must justify its acts under the First Amendment. However, the important question is not one of U.S. constitutional law.

The more general claim, true for any country that decides to enforce a DMCA-like law, is that prohibiting technologies that allow individuals to make flexible and creative uses of digital cultural materials burdens the development of the networked information economy and society. It burdens individual autonomy, the emergence of the networked public sphere and critical culture, and some of the paths available for global human development that the networked information economy makes possible. All these losses will be incurred in expectation of improvements in creativity, even though it is not at all clear that doing so would actually improve, even on a simple utilitarian calculus, the creative production of any given country or region. Passing a DMCA-type law will not by itself squelch the development of nonmarket and peer production. Indeed, many of these technological and social-economic developments emerged and have flourished after the DMCA was already in place. It does, however, represent a choice to tilt the institutional ecology in favor of industrial production and distribution of cultural packaged goods, at the expense of commons-based relations of sharing information, knowledge, and culture. Twentieth-century cultural materials provide the most immediate and important source of references and images for contemporary cultural creation. Given the relatively recent provenance of movies, recorded music, and photography, much of contemporary culture was created in these media. These basic materials for

the creation of contemporary multimedia culture are, in turn, encoded in formats that cannot simply be copied by hand, as texts might be even in the teeth of technical protection measures. The capacity to copy mechanically is a necessary precondition for the capacity to quote and combine existing materials of these kinds into new cultural statements and conversational moves. Preserving the capacity of industrial cultural producers to maintain a hermetic seal on the use of materials to which they own copyright can be bought only at the cost of disabling the newly emerging modes of cultural production from quoting and directly building upon much of the culture of the last century.

The Battle over Peer-to-Peer Networks

The second major institutional battle over the technical and social trajectory of Internet development has revolved around peer-to-peer (p2p) networks. I offer a detailed description of it here, but not because I think it will be the make-it-or-break-it of the networked information economy. If any laws have that determinative a power, they are the Fritz chip and DMCA. However, the peer-to-peer legal battle offers an excellent case study of just how difficult it is to evaluate the effects of institutional ecology on technology, economic organization, and social practice.

Peer-to-peer technologies as a global phenomenon emerged from Napster and its use by tens of millions of users around the globe for unauthorized sharing of music files. In the six years since their introduction, p2p networks have developed robust and impressive technical capabilities. They have been adopted by more than one hundred million users, and are increasingly applied to uses well beyond music sharing. These developments have occurred despite a systematic and aggressive campaign of litigation and criminal enforcement in a number of national systems against both developers and users. Technically, p2p networks are algorithms that run on top of the Internet and allow users to connect directly from one end user's machine to another. In theory, that is how the whole Internet works - or at least how it worked when there were a small number of computers attached to it. In practice, most users connect through an Internet service provider, and most content available for access on the Internet was available on a server owned and operated by someone distinct from its users. In the late 1990s, there were rudimentary utilities that allowed one user to access information stored on the computer of another, but no widely used utility allowed large numbers of individuals to search each other's hard drives and share data directly from one user to another. Around 1998-1999, early Internet music distribution models, like MP3.com, therefore provided a centralized distribution point for music. This made them highly vulnerable to legal attack. Shawn Fanning, then eighteen years old, was apparently looking for ways to do what teenagers always do - share their music with friends - in a way that would not involve a central point of storing and copying. He developed Napster - the first major, widely adopted p2p technology. Unlike MP3.com, users of Napster could connect their computers directly - one person could download a song stored on the computer of another without mediation. All that the Napster site itself did, in addition to providing the end-user software, was to provide a centralized directory of which songs resided on which machine. There is little disagreement in the literature that it is an infringement under U.S. copyright law for any given user to allow others to duplicate copyrighted music from his or her computer to theirs. The centralizing role of Napster in facilitating these

exchanges, alongside a number of ill-considered statements by some of its principals, were enough to render the company liable for contributory copyright infringement.

The genie of p2p technology and the social practice of sharing music, however, were already out of the bottle. The story of the following few years, to the extent that one can tell a history of the present and the recent past, offers two core insights. First, it shows how institutional design can be a battleground over the conditions of cultural production in the digital environment. Second, it exposes the limits of the extent to which the institutional ecology can determine the ultimate structure of behavior at a moment of significant and rapid technological and social perturbation. Napster's judicial closure provided no real respite for the recording industry. As Napster was winding down, Gnutella, a free software alternative, had already begun to replace it. Gnutella did not depend on any centralized component, not even to facilitate search. This meant that there was no central provider. There was no firm against which to bring action. Even if there were, it would be impossible to "shut down" use of the program. Gnutella was a freestanding program that individual users could install. Once installed, its users could connect to anyone else who had installed the program, without passing through any choke point. There was no central server to shut down. Gnutella had some technical imperfections, but these were soon overcome by other implementations of p2p. The most successful improvement over Gnutella was the FastTrack architecture, now used by Kazaa, Grokster, and other applications, including some free software applications. It improves on the search capabilities of Gnutella by designating some users as "supernodes," which store information about what songs are available in their "neighborhood." This avoids Gnutella's primary weakness, the relatively high degree of network overhead traffic. The supernodes operate on an ad hoc basis. They change based on whose computer is available with enough storage and bandwidth. They too, therefore, provide no litigation target. Other technologies have developed to speed up or make more robust the distribution of files, including BitTorrent, eDonkey and its free-software relative eMule, and many others. Within less than two years of Napster's closure, more people were using these various platforms to share files than Napster had users at its height. Some of these new firms found themselves again under legal assault - both in the United States and abroad.

As the technologies grew and developed, and as the legal attacks increased, the basic problem presented by the litigation against technology manufacturers became evident. Peer-to-peer techniques can be used for a wide range of uses, only some of which are illegal. At the simplest level, they can be used to distribute music that is released by an increasing number of bands freely. These bands hope to get exposure that they can parley into concert performances. As recorded music from the 1950s begins to fall into the public domain in Europe and Australia, golden oldies become another legitimate reason to use p2p technologies. More important, p2p systems are being adapted to different kinds of uses. Chapter 7 discusses how FreeNet is being used to disseminate subversive documents, using the persistence and robustness of p2p networks to evade detection and suppression by authoritarian regimes. BitTorrent was initially developed to deal with the large file transfers required for free software distributions. BitTorrent and eDonkey were both used by the Swarthmore students when their college shut down their Internet connection in response to Diebold's letter threatening action under the service provider liability provisions of the DMCA. The founders of KaZaa have begun to offer an Internet telephony utility, Skype, which allows users to

make phone calls from one computer to another for free, and from their computer to the telephone network for a small fee. Skype is a p2p technology.

In other words, p2p is developing as a general approach toward producing distributed data storage and retrieval systems, just as open wireless networks and distributed computing are emerging to take advantage of personal devices to produce distributed communications and computation systems, respectively. As the social and technological uses of p2p technologies grow and diversify, the legal assault on all p2p developers becomes less sustainable - both as a legal matter and as a social-technical matter. KaZaa was sued in the Netherlands, and moved to Australia. It was later subject to actions in Australia, but by that time, the Dutch courts found the company not to be liable to the music labels. Grokster, a firm based in the United States, was initially found to have offered a sufficiently diverse set of capabilities, beyond merely facilitating copyright infringements, that the Court of Appeals for the Ninth Circuit refused to find it liable simply for making and distributing its software. The Supreme Court reversed that holding, however, returning the case to the lower courts to find, factually, whether Grokster had actual intent to facilitate illegal copying. [/17] Even if Grokster ultimately loses, the FastTrack network architecture will not disappear; clients (that is, end user software) will continue to exist, including free software clients. Perhaps it will be harder to raise money for businesses located within the United States to operate in this technological space, because the new rule announced by the Supreme Court in *Grokster* raises the risk of litigation for innovators in the p2p space. However, as with encryption regulation in the mid-1990s, it is not clear that the United States can unilaterally prevent the development of technology for which there is worldwide demand and with regard to whose development there is globally accessible talent.

How important more generally are these legal battles to the organization of cultural production in the networked environment? There are two components to the answer: The first component considers the likely effect of the legal battles on the development and adoption of the technology and the social practice of promiscuous copying. In this domain, law seems unlikely to prevent the continued development of p2p technologies. It has, however, had two opposite results. First, it has affected the path of the technological evolution in a way that is contrary to the industry interests but consistent with increasing distribution of the core functions of the logical layer. Second, it seems to have dampened somewhat the social practice of file sharing. The second component assumes that a range of p2p technologies will continue to be widely adopted, and that some significant amount of sharing will continue to be practiced. The question then becomes what effect this will have on the primary cultural industries that have fought this technology - movies and recorded music. Within this new context, music will likely change more radically than movies, and the primary effect will be on the accreditation function - how music is recognized and adopted by fans. Film, if it is substantially affected, will likely be affected largely by a shift in tastes.

MP3.com was the first major music distribution site shut down by litigation. From the industry's perspective, it should have represented an entirely unthreatening business model. Users paid a subscription fee, in exchange for which they were allowed to download music. There were various quirks and kinks in this model that made it unattractive to the music industry at the time: the industry did not control this major site, and therefore had to share the rents from the music, and more important, there was no effective control over the music files once downloaded. However,

from the perspective of 2005, MP3.com was a vastly more manageable technology for the sound recording business model than a free software file-sharing client. MP3.com was a single site, with a corporate owner that could be (and was) held responsible. It controlled which user had access to what files - by requiring each user to insert a CD into the computer to prove that he or she had bought the CD - so that usage could in principle be monitored and, if desired, compensation could be tied to usage. It did not fundamentally change the social practice of choosing music. It provided something that was more like a music-on-demand jukebox than a point of music sharing. As a legal matter, MP3.com's infringement was centered on the fact that it stored and delivered the music from this central server instead of from the licensed individual copies. In response to the shutdown of MP3.com, Napster redesigned the role of the centralized mode, and left storage in the hands of users, keeping only the directory and search functions centralized. When Napster was shut down, Gnutella and later FastTrack further decentralized the system, offering a fully decentralized, ad hoc reconfigurable cataloging and search function. Because these algorithms represent architecture and a protocol-based network, not a particular program, they are usable in many different implementations. This includes free software programs like MLDonkey - which is a nascent file-sharing system that is aimed to run simultaneously across most of the popular file-sharing networks, including FastTrack, BitTorrent, and Overnet, the eDonkey network. These programs are now written by, and available from, many different jurisdictions. There is no central point of control over their distribution. There is no central point through which to measure and charge for their use. They are, from a technical perspective, much more resilient to litigation attack, and much less friendly to various possible models of charging for downloads or usage. From a technological perspective, then, the litigation backfired. It created a network that is less susceptible to integration into an industrial model of music distribution based on royalty payments per user or use.

It is harder to gauge, however, whether the litigation was a success or a failure from a social-practice point of view. There have been conflicting reports on the effects of file sharing and the litigation on CD sales. The recording industry claimed that CD sales were down because of file sharing, but more independent academic studies suggested that CD sales were not independently affected by file sharing, as opposed to the general economic downturn. [18] The Pew project on Internet and American Life user survey data suggests that the litigation strategy against individual users has dampened the use of file sharing, though file sharing is still substantially more common among users than paying for files from the newly emerging pay-per-download authorized services. In mid-2003, the Pew study found that 29 percent of Internet users surveyed said they had downloaded music files, identical to the percentage of users who had downloaded music in the first quarter of 2001, the heyday of Napster. Twenty-one percent responded that they allow others to download from their computer.[19] This meant that somewhere between twenty-six and thirty-five million adults in the United States alone were sharing music files in mid-2003, when the recording industry began to sue individual users. Of these, fully two-thirds expressly stated that they did not care whether the files they downloaded were or were not copyrighted. By the end of 2003, five months after the industry began to sue individuals, the number of respondents who admitted to downloading music dropped by half. During the next few months, these numbers increased slightly to twenty-three million adults, remaining below the mid-2003 numbers in absolute terms and more so in terms of percentage of Internet users. Of those who had at one point downloaded, but had stopped, roughly a third said that the threat of suit was the reason they had stopped file sharing. [20] During this same

period, use of pay online music download services, like iTunes, rose to about 7 percent of Internet users. Sharing of all kinds of media files - music, movies, and games - was at 23 percent of adult Internet users. These numbers do indeed suggest that, in the aggregate, music downloading is reported somewhat less often than it was in the past. It is hard to tell how much of this reduction is due to actual behavioral change as compared to an unwillingness to self-report on behavior that could subject one to litigation. It is impossible to tell how much of an effect the litigation has had specifically on sharing by younger people - teenagers and college students - who make up a large portion of both CD buyers and file sharers. Nonetheless, the reduction in the total number of self-reported users and the relatively steady percentage of total Internet users who share files of various kinds suggest that the litigation does seem to have had a moderating effect on file sharing as a social practice. It has not, however, prevented file sharing from continuing to be a major behavioral pattern among one-fifth to one-quarter of Internet users, and likely a much higher proportion in the most relevant populations from the perspective of the music and movie industries - teenagers and young adults.

From the perspective of understanding the effects of institutional ecology, then, the still-raging battle over peer-to-peer networks presents an ambiguous picture. One can speculate with some degree of confidence that, had Napster not been stopped by litigation, file sharing would have been a much wider social practice than it is today. The application was extremely easy to use; it offered a single network for all file-sharing users, thereby offering an extremely diverse and universal content distribution network; and for a brief period, it was a cultural icon and a seemingly acceptable social practice. The period of regrouping that followed its closure; the imperfect interfaces of early Gnutella clients; the relative fragmentation of file sharing into a number of networks, each with a smaller coverage of content than was present; and the fear of personal litigation risk are likely to have limited adoption. On the other hand, in the longer run, the technological developments have created platforms that are less compatible with the industrial model, and which would be harder to integrate into a stable settlement for music distribution in the digital environment.

Prediction aside, it is not immediately obvious why peer-to-peer networks contribute to the kinds of nonmarket production and creativity that I have focused on as the core of the networked information economy. At first blush, they seem simply to be mechanisms for fans to get industrially produced recorded music without paying musicians. This has little to do with democratization of creativity. To see why p2p networks nonetheless are a part of the development of a more attractive cultural production system, and how they can therefore affect the industrial organization of cultural production, we can look first at music, and then, independently, at movies. The industrial structure of each is different, and the likely effects of p2p networks are different in each case.

Recorded music began with the phonograph - a packaged good intended primarily for home consumption. The industry that grew around the ability to stamp and distribute records divided the revenue structure such that artists have been paid primarily from live public performances and merchandizing. Very few musicians, including successful recording artists, make money from recording royalties. The recording industry takes almost all of the revenues from record and CD sales, and provides primarily promotion and distribution. It does not bear the capital cost of the initial musical creation; artists do. With the declining cost of computation, that cost has become

relatively low, often simply a computer owned by artists themselves, much as they own their instruments. Because of this industrial structure, peer-to-peer networks are a genuine threat to displacing the entire recording industry, while leaving musicians, if not entirely unaffected, relatively insulated from the change and perhaps mildly better off. Just as the recording industry stamps CDs, promotes them on radio stations, and places them on distribution chain shelves, p2p networks produce the physical and informational aspects of a music distribution system. However, p2p networks do so collaboratively, by sharing the capacity of their computers, hard drives, and network connections. Filtering and accreditation, or "promotion," are produced on the model that Eben Moglen called "anarchist distribution." Jane's friends and friends of her friends are more likely to know exactly what music would make her happy than are recording executives trying to predict which song to place, on which station and which shelf, to expose her to exactly the music she is most likely to buy in a context where she would buy it. File-sharing systems produce distribution and "promotion" of music in a social-sharing modality. Alongside peer-produced music reviews, they could entirely supplant the role of the recording industry.

Musicians and songwriters seem to be relatively insulated from the effects of p2p networks, and on balance, are probably affected positively. The most comprehensive survey data available, from mid-2004, shows that 35 percent of musicians and songwriters said that free downloads have helped their careers. Only 5 percent said it has hurt them. Thirty percent said it increased attendance at concerts, 21 percent that it helped them sell CDs and other merchandise, and 19 percent that it helped them gain radio playing time. These results are consistent with what one would expect given the revenue structure of the industry, although the study did not separate answers out based on whether the respondent was able to live entirely or primarily on their music, which represented only 16 percent of the respondents to the survey. In all, it appears that much of the actual flow of revenue to artists - from performances and other sources - is stable. This is likely to remain true even if the CD market were entirely displaced by peer-to-peer distribution. Musicians will still be able to play for their dinner, at least not significantly less so than they can today. Perhaps there will be fewer millionaires. Perhaps fewer mediocre musicians with attractive physiques will be sold as "geniuses," and more talented musicians will be heard than otherwise would have, and will as a result be able to get paying gigs instead of waiting tables or "getting a job." But it would be silly to think that music, a cultural form without which no human society has existed, will cease to be in our world if we abandon the industrial form it took for the blink of a historical eye that was the twentieth century. Music was not born with the phonograph, nor will it die with the peer-to-peer network. The terms of the debate, then, are about cultural policy; perhaps about industrial policy. Will we get the kind of music we want in this system, whoever "we" are? Will American recording companies continue to get the export revenue streams they do? Will artists be able to live from making music? Some of these arguments are serious. Some are but a tempest in a monopoly-rent teapot. It is clear that a technological change has rendered obsolete a particular mode of distributing information and culture. Distribution, once the sole domain of market-based firms, now can be produced by decentralized networks of users, sharing instantiations of music they deem attractive with others, using equipment they own and generic network connections. This distribution network, in turn, allows a much more diverse range of musicians to reach much more finely grained audiences than were optimal for industrial production and distribution of mechanical instantiations of music in vinyl or CD formats. The legal battles reflect an effort by an incumbent industry to preserve its very

lucrative business model. The industry has, to this point, delayed the transition to peer-based distribution, but it is unclear for how long or to what extent it will be successful in preventing the gradual transition to user-based distribution.

The movie industry has a different industrial structure and likely a different trajectory in its relations to p2p networks. First and foremost, movies began as a relatively high capital cost experience good. Making a movie, as opposed to writing a song, was something that required a studio and a large workforce. It could not be done by a musician with a guitar or a piano. Furthermore, movies were, throughout most of their history, collective experience goods. They were a medium for public performance experienced outside of the home, in a social context. With the introduction of television, it was easy to adapt movie revenue structure by delaying release of films to television viewing until after demand for the movie at the theater declined, as well as to develop their capabilities into a new line of business - television production. However, theatrical release continued to be the major source of revenue. When video came along, the movie industry cried murder in the Sony Betamax case, but actually found it quite easy to work videocassettes into yet another release window, like television, and another medium, the made-for-video movie. Digital distribution affects the distribution of cultural artifacts as packaged goods for home consumption. It does not affect the social experience of going out to the movies. At most, it could affect the consumption of the twenty-year-old mode of movie distribution: videos and DVDs. As recently as the year 2000, when the Hollywood studios were litigating the DeCSS case, they represented to the court that home video sales were roughly 40 percent of revenue, a number consistent with other reports. [21] The remainder, composed of theatrical release revenues and various television releases, remains reasonably unthreatened as a set of modes of revenue capture to sustain the high-production value, high-cost movies that typify Hollywood. Forty percent is undoubtedly a large chunk, but unlike the recording industry, which began with individually owned recordings, the movie industry preexisted videocassettes and DVDs, and is likely to outlive them even if p2p networks were to eliminate that market entirely, which is doubtful.

The harder and more interesting question is whether cheap high-quality digital video-capture and editing technologies combined with p2p networks for efficient distribution could make film a more diverse medium than it is now. The potential hypothetical promise of p2p networks like BitTorrent is that they could offer very robust and efficient distribution networks for films outside the mainstream industry. Unlike garage bands and small-scale music productions, however, this promise is as yet speculative. We do not invest in public education for film creation, as we do in the teaching of writing. Most of the raw materials out of which a culture of digital capture and amateur editing could develop are themselves under copyright, a subject we return to when considering the content layer. There are some early efforts, like atomfilms.com, at short movie distribution. The technological capabilities are there. It is possible that if films older than thirty or even fifty years were released into the public domain, they would form the raw material out of which a new cultural production practice would form. If it did, p2p networks would likely play an important role in their distribution. However, for now, although the sound recording and movie industries stand shoulder to shoulder in the lobbying efforts, their circumstances and likely trajectory in relation to file sharing are likely quite different.

The battles over p2p and the DMCA offer some insight into the potential, but also the limits, of tweaking the institutional ecology. The ambition of the industrial cultural producers in both cases was significant. They sought to deploy law to shape emerging technologies and social practices to make sure that the business model they had adopted for the technologies of film and sound recording continued to work in the digital environment. Doing so effectively would require substantial elimination of certain lines of innovation, like certain kinds of decryption and p2p networks. It would require outlawing behavior widely adopted by people around the world - social sharing of most things that they can easily share - which, in the case of music, has been adopted by tens of millions of people around the world. The belief that all this could be changed in a globally interconnected network through the use of law was perhaps naïve. Nonetheless, the legal efforts have had some impact on social practices and on the ready availability of materials for free use. The DMCA may not have made any single copyright protection mechanism hold up to the scrutiny of hackers and crackers around the Internet. However, it has prevented circumvention devices from being integrated into mainstream platforms, like the Windows operating system or some of the main antivirus programs, which would have been "natural" places for them to appear in consumer markets. The p2p litigation did not eliminate the p2p networks, but it does seem to have successfully dampened the social practice of file sharing. One can take quite different views of these effects from a policy perspective. However, it is clear that they are self-conscious efforts to tweak the institutional ecology of the digital environment in order to dampen the most direct threats it poses for the twentieth-century industrial model of cultural production. In the case of the DMCA, this is done at the direct cost of making it substantially harder for users to make creative use of the existing stock of audiovisual materials from the twentieth century - materials that are absolutely central to our cultural self-understanding at the beginning of the twenty-first century. In the case of p2p networks, the cost to nonmarket production is more indirect, and may vary across different cultural forms. The most important long-term effect of the pressure that this litigation has put on technology to develop decentralized search and retrieval systems may, ultimately and ironically, be to improve the efficiency of radically decentralized cultural production and distribution, and make decentralized production more, rather than less, robust to the vicissitudes of institutional ecology.

The Domain Name System: From Public Trust to the Fetishism of Mnemonics

Not all battles over the role of property-like arrangements at the logical layer originate from Hollywood and the recording industry. One of the major battles outside of the ambit of the copyright industries concerned the allocation and ownership of domain names. At stake was the degree to which brand name ownership in the material world could be leveraged into attention on the Internet. Domain names are alphanumeric mnemonics used to represent actual Internet addresses of computers connected to the network. While 130.132.51.8 is hard for human beings to remember, www.yale.edu is easier. The two strings have identical meaning to any computer connected to the Internet - they refer to a server that responds to World Wide Web queries for Yale University's main site. Every computer connected to the Internet has a unique address, either permanent or assigned by a provider for the session. That requires that someone distribute addresses - both numeric and mnemonic. Until 1992, names and numbers were assigned on a purely first-come, first-served basis by Jon Postel, one of the very first developers of the Internet, under U.S. government contract. Postel also ran a computer, called the root server, to which all computers

would turn to ask the numeric address of letters.mnemonic.edu, so they could translate what the human operator remembered as the address into one their machine could use. Postel called this system "the Internet Assigned Numbers Authority, IANA," whose motto he set as, "Dedicated to preserving the central coordinating functions of the global Internet for the public good." In 1992, Postel got tired of this coordinating job, and the government contracted it to a private firm called Network Solutions, Inc., or NSI. As the number of applications grew, and as the administration sought to make this system pay for itself, NSI was allowed in 1995 to begin to charge fees for assigning names and numbers. At about the same time, widespread adoption of a graphical browser made using the World Wide Web radically simpler and more intuitive to the uninitiated. These two developments brought together two forces to bear on the domain name issue - each with a very different origin and intent. The first force consisted of the engineers who had created and developed the Internet, led by Postel, who saw the domain name space to be a public trust and resisted its commercialization by NSI. The second force consisted of trademark owners and their lawyers, who suddenly realized the potential for using control over domain names to extend the value of their brand names to a new domain of trade - e-commerce. These two forces placed the U.S. government under pressure to do two things: (1) release the monopoly that NSI - a for-profit corporation - had on the domain name space, and (2) find an efficient means of allowing trademark owners to control the use of alphanumeric strings used in their trademarks as domain names. Postel initially tried to "take back the root" by asking various regional domain name servers to point to his computer, instead of to the one maintained by NSI in Virginia. This caused uproar in the government, and Postel was accused of attacking and hijacking the Internet! His stature and passion, however, placed significant weight on the side of keeping the naming system as an open public trust. That position came to an abrupt end with his death in 1996. By late 1996, a self-appointed International Ad Hoc Committee (IAHC) was formed, with the blessing of the Internet Society (ISOC), a professional membership society for individuals and organizations involved in Internet planning. IAHC's membership was about half intellectual property lawyers and half engineers. In February 1997, IAHC came out with a document called the gTLD-MoU (generic top-level domain name memorandum of understanding). Although the product of a small group, the gTLD-MoU claimed to speak for "The Internet Community." Although it involved no governments, it was deposited "for signature" with the International Telecommunications Union (ITU). Dutifully, some 226 organizations - Internet services companies, telecommunications providers, consulting firms, and a few chapters of the ISOC signed on. Section 2 of the gTLD-MoU, announcing its principles, reveals the driving forces of the project. While it begins with the announcement that the top-level domain space "is a public resource and is subject to the public trust," it quickly commits to the principle that "the current and future Internet name space stakeholders can benefit most from a self-regulatory and market-oriented approach to Internet domain name registration services." This results in two policy principles: (1) commercial competition in domain name registration by releasing the monopoly NSI had, and (2) protecting trademarks in the alphanumeric strings that make up the second-level domain names. The final, internationalizing component of the effort - represented by the interests of the WIPO and ITU bureaucracies - was attained by creating a Council of Registrars as a Swiss corporation, and creating special relationships with the ITU and the WIPO.

None of this institutional edifice could be built without the U.S. government. In early 1998, the administration responded to this ferment with a green paper, seeking the creation of a private,

nonprofit corporation registered in the United States to take on management of the domain name issue. By its own terms, the green paper responded to concerns of the domain name registration monopoly and of trademark issues in domain names, first and foremost, and to some extent to increasing clamor from abroad for a voice in Internet governance. Despite a cool response from the European Union, the U.S. government proceeded to finalize a white paper and authorize the creation of its preferred model - the private, nonprofit corporation. Thus was born the Internet Corporation for Assigned Names and Numbers (ICANN) as a private, nonprofit California corporation. Over time, it succeeded in large measure in loosening NSI's monopoly on domain name registration. Its efforts on the trademark side effectively created a global preemptive property right. Following an invitation in the U.S. government's white paper for ICANN to study the proper approach to trademark enforcement in the domain name space, ICANN and WIPO initiated a process that began in July 1998 and ended in April 1999. As Froomkin describes his experience as a public-interest expert in this process, the process feigned transparency and open discourse, but was in actuality an opaque staff-driven drafting effort./22 The result was a very strong global property right available to trademark owners in the alphanumeric strings that make up domain names. This was supported by binding arbitration. Because it controlled the root server, ICANN could enforce its arbitration decisions worldwide. If ICANN decides that, say, the McDonald's fast-food corporation and not a hypothetical farmer named Old McDonald owned www.mcdonalds.com, all computers in the world would be referred to the corporate site, not the personal one. Not entirely satisfied with the degree to which the ICANN-WIPO process protected their trademarks, some of the major trademark owners lobbied the U.S. Congress to pass an even stricter law. This law would make it easier for the owners of commercial brand names to obtain domain names that include their brand, whether or not there was any probability that users would actually confuse sites like the hypothetical Old McDonald's with that of the fast-food chain.

The degree to which the increased appropriation of the domain name space is important is a function of the extent to which the cultural practice of using human memory to find information will continue to be widespread. The underlying assumption of the value of trademarked alphanumeric strings as second-level domain names is that users will approach electronic commerce by typing in "www.brandname.com" as their standard way of relating to information on the Net. This is far from obviously the most efficient solution. In physical space, where collecting comparative information on price, quality, and so on is very costly, brand names serve an important informational role. In cyberspace, where software can compare prices, and product-review services that link to vendors are easy to set up and cheap to implement, the brand name becomes an encumbrance on good information, not its facilitator. If users are limited, for instance, to hunting around as to whether information they seek is on www.brandname.com, www.brand= name.com, or www.brand.net, name recognition from the real world becomes a bottleneck to e-commerce. And this is precisely the reason why owners of established marks sought to assure early adoption of trademarks in domain names - it assures users that they can, in fact, find their accustomed products on the Web without having to go through search algorithms that might expose them to comparison with pesky start-up competitors. As search engines become better and more tightly integrated into the basic browser functionality, the idea that a user who wants to buy from Delta Airlines would simply type "www.delta.com," as opposed to plugging "delta airlines" into an integrated search toolbar and getting the airline as a first hit becomes quaint. However, quaint inefficient cultural

practices can persist. And if this indeed is one that will persist, then the contours of the property right matter. As the law has developed over the past few years, ownership of a trademark that includes a certain alphanumeric string almost always gives the owner of the trademark a preemptive right in using the letters and numbers incorporated in that mark as a domain name.

Domain name disputes have fallen into three main categories. There are cases of simple arbitrage. Individuals who predicted that having a domain name with the brand name in it would be valuable, registered such domain names aplenty, and waited for the flat-footed brand name owners to pay them to hand over the domain. There is nothing more inefficient about this form of arbitrage than any other. The arbitrageurs "reserved" commercially valuable names so they could be auctioned, rather than taken up by someone who might have a non-negotiable interest in the name - for example, someone whose personal name it was. These arbitrageurs were nonetheless branded pirates and hijackers, and the consistent result of all the cases on domain names has been that the corporate owners of brand names receive the domain names associated with their brands without having to pay the arbitrageurs. Indeed, the arbitrageurs were subject to damage judgments. A second kind of case involved bona fide holders of domain names that made sense for them, but were nonetheless shared with a famous brand name. One child nicknamed "Pokey" registered "pokey.org," and his battle to keep that name against a toy manufacturer that sold a toy called "pokey" became a poster child for this type of case. Results have been more mixed in this case, depending on how sympathetic the early registrant was. The third type of case - and in many senses, most important from the perspective of freedom to participate not merely as a consumer in the networked environment, but as a producer - involves those who use brand names to draw attention to the fact that they are attacking the owner of the brand. One well-known example occurred when Verizon Wireless was launched. The same hacker magazine involved in the DeCSS case, *2600*, purchased the domain name "verizonreallysucks.com" to poke fun at Verizon. In response to a letter requiring that they give up the domain name, the magazine purchased the domain name "VerizonShouldSpendMoreTimeFixingItsNetworkAndLessMoneyOnLawyers.com." These types of cases have again met with varying degrees of sympathy from courts and arbitrators under the ICANN process, although it is fairly obvious that using a brand name in order to mock and criticize its owner and the cultural meaning it tries to attach to its mark is at the very core of fair use, cultural criticism, and free expression.

The point here is not to argue for one type of answer or another in terms of trademark law, constitutional law, or the logic of ICANN. It is to identify points of pressure where the drive to create proprietary rights is creating points of control over the flow of information and the freedom to make meaning in the networked environment. The domain name issue was seen by many as momentous when it was new. ICANN has drawn a variety of both yearnings and fears as a potential source of democratic governance for the Internet or a platform for U.S. hegemony. I suspect that neither of these will turn out to be true. The importance of property rights in domain names is directly based on the search practices of users. Search engines, directories, review sites, and referrals through links play a large role in enabling users to find information they are interested in. Control over the domain name space is unlikely to provide a real bottleneck that will prevent both commercial competitors and individual speakers from drawing attention to their competition or criticism. However, the battle is indicative of the efforts to use proprietary rights in a particular

element of the institutional ecology of the logical layer - trademarks in domain names - to tilt the environment in favor of the owners of famous brand names, and against individuals, noncommercial actors, and smaller, less-known competitors.

The Browser Wars

A much more fundamental battle over the logical layer has occurred in the browser wars. Here, the "institutional" component is not formal institutions, like laws or regulations, but technical practice institutions - the standards for Web site design. Unlike on the network protocol side, the device side of the logical layer - the software running personal computers - was thoroughly property-based by the mid-1990s. Microsoft's dominance in desktop operating systems was well established, and there was strong presence of other software publishers in consumer applications, pulling the logical layer toward a proprietary model. In 1995, Microsoft came to perceive the Internet and particularly the World Wide Web as a threat to its control over the desktop. The user-side Web browser threatened to make the desktop a more open environment that would undermine its monopoly. Since that time, the two pulls - the openness of the nonproprietary network and the closed nature of the desktop - have engaged in a fairly energetic tug-of-war over the digital environment. This push-me-pull-you game is played out both in the domain of market share, where Microsoft has been immensely successful, and in the domain of standard setting, where it has been only moderately successful. In market share, the story is well known and has been well documented in the Microsoft antitrust litigation. Part of the reason that it is so hard for a new operating system to compete with Microsoft's is that application developers write first, and sometimes only, for the already-dominant operating system. A firm investing millions of dollars in developing a new piece of photo-editing software will usually choose to write it so that it works with the operating system that has two hundred million users, not the one that has only fifteen million users. Microsoft feared that Netscape's browser, dominant in the mid-1990s, would come to be a universal translator among applications - that developers could write their applications to run on the browser, and the browser would handle translation across different operating systems. If that were to happen, Microsoft's operating system would have to compete on intrinsic quality. Windows would lose the boost of the felicitous feedback effect, where more users mean more applications, and this greater number of applications in turn draws more new users, and so forth. To prevent this eventuality, Microsoft engaged in a series of practices, ultimately found to have violated the antitrust laws, aimed at getting a dominant majority of Internet users to adopt Microsoft's Internet Explorer (IE). Illegal or not, these practices succeeded in making IE the dominant browser, overtaking the original market leader, Netscape, within a short number of years. By the time the antitrust case was completed, Netscape had turned browser development over to the open-source development community, but under licensing conditions sufficiently vague so that the project generated little early engagement. Only around 2001-2002, did the Mozilla browser development project get sufficient independence and security for developers to begin to contribute energetically. It was only in late 2004, early 2005, that Mozilla Firefox became the first major release of a free software browser that showed promise of capturing some user-share back from IE.

Microsoft's dominance over the operating system and browser has not, as a practical matter, resulted in tight control over the information flow and use on the Internet. This is so for three reasons. First,

the TCP/IP protocol is more fundamental to Internet communications. It allows any application or content to run across the network, as long as it knows how to translate itself into very simple packets with standard addressing information. To prevent applications from doing this over basic TCP/IP would make the Microsoft operating system substantially crippling to many applications developers, which brings us to the second reason. Microsoft's dominance depends to a great extent on the vastly greater library of applications available to run on Windows. To make this library possible, Microsoft makes available a wide range of application program interfaces that developers can use without seeking Microsoft's permission. As a strategic decision about what enhances its core dominance, Microsoft may tilt the application development arena in its favor, but not enough to make it too hard for most applications to be implemented on a Windows platform. While not nearly as open as a genuinely open-source platform, Windows is also a far cry from a completely controlled platform, whose owner seeks to control all applications that are permitted to be developed for, and all uses that can be made of, its platform. Third, while IE controls much of the browser market share, Microsoft has not succeeded in dominating the standards for Web authoring. Web browser standard setting happens on the turf of the mythic creator of the Web - Tim Berners Lee. Lee chairs the W3C, a nonprofit organization that sets the standard ways in which Web pages are authored so that they have a predictable appearance on the browser's screen. Microsoft has, over the years, introduced various proprietary extensions that are not part of the Web standard, and has persuaded many Web authors to optimize their Web sites to IE. If it succeeds, it will have wrested practical control over standard setting from the W3C. However, as of this writing, Web pages generally continue to be authored using mostly standard, open extensions, and anyone browsing the Internet with a free software browser, like any of the Mozilla family, will be able to read and interact with most Web sites, including the major e-commerce sites, without encountering nonstandard interfaces optimized for IE. At a minimum, these sites are able to query the browser as to whether or not it is IE, and serve it with either the open standard or the proprietary standard version accordingly.

Free Software

The role of Mozilla in the browser wars points to the much more substantial and general role of the free software movement and the open-source development community as major sources of openness, and as a backstop against appropriation of the logical layer. In some of the most fundamental uses of the Internet-Web-server software, Web-scripting software, and e-mail servers - free or open-source software has a dominant user share. In others, like the operating system, it offers a robust alternative sufficiently significant to prevent enclosure of an entire component of the logical layer. Because of its licensing structure and the fact that the technical specifications are open for inspection and use by anyone, free software offers the most completely open, commons-based institutional and organizational arrangement for any resource or capability in the digital environment. Any resource in the logical layer that is the product of a free software development project is institutionally designed to be available for nonmarket, nonproprietary strategies of use. The same openness, however, makes free software resistant to control. If one tries to implement a constraining implementation of a certain function - for example, an audio driver that will not allow music to be played without proper authorization from a copyright holder - the openness of the code for inspection will allow users to identify what, and how, the software is constraining. The same institutional framework will allow any developer to "fix" the problem and change the way the

software behaves. This is how free and open-source software is developed to begin with. One cannot limit access to the software - for purposes of inspection and modification - to developers whose behavior can be controlled by contract or property and still have the software be "open source" or free. As long as free software can provide a fully implemented alternative to the computing functionalities users want, perfect enclosure of the logical layer is impossible. This openness is a boon for those who wish the network to develop in response to a wide range of motivations and practices. However, it presents a serious problem for anyone who seeks to constrain the range of uses made of the Internet. And, just as they did in the context of trusted systems, the incumbent industrial culture producers - Hollywood and the recording industry - would, in fact, like to control how the Internet is used and how software behaves.

Software Patents

Throughout most of its history, software has been protected primarily by copyright, if at all. Beginning in the early 1980s, and culminating formally in the late 1990s, the Federal Circuit, the appellate court that oversees the U.S. patent law, made clear that software was patentable. The result has been that software has increasingly become the subject of patent rights. There is now pressure for the European Union to pass a similar reform, and to internationalize the patentability of software more generally. There are a variety of policy questions surrounding the advisability of software patents. Software development is a highly incremental process. This means that patents tend to impose a burden on a substantial amount of future innovation, and to reward innovation steps whose qualitative improvement over past contributions may be too small to justify the discontinuity represented by a patent grant. Moreover, innovation in the software business has flourished without patents, and there is no obvious reason to implement a new exclusive right in a market that seems to have been enormously innovative without it. Most important, software components interact with each other constantly. Sometimes interoperating with a certain program may be absolutely necessary to perform a function, not because the software is so good, but because it has become the standard. The patent then may extend to the very functionality, whereas a copyright would have extended only to the particular code by which it was achieved. The primary fear is that patents over standards could become major bottlenecks.

From the perspective of the battle over the institutional ecology, free software and open-source development stand to lose the most from software patents. A patent holder may charge a firm that develops dependent software in order to capture rents. However, there is no obvious party to charge for free software development. Even if the patent owner has a very open licensing policy - say, licensing the patent nonexclusively to anyone without discrimination for $10,000 - most free software developers will not be able to play. IBM and Red Hat may pay for licenses, but the individual contributor hacking away at his or her computer, will not be able to. The basic driver of free software innovation is easy ubiquitous access to the state of the art, coupled with diverse motivations and talents brought to bear on a particular design problem. If working on a problem requires a patent license, and if any new development must not only write new source code, but also avoid replicating a broad scope patent or else pay a large fee, then the conditions for free software development are thoroughly undermined. Free software is responsible for some of the most basic and widely used innovations and utilities on the Internet today. Software more generally is heavily

populated by service firms that do not functionally rely on exclusive rights, copyrights, or patents. Neither free software nor service-based software development need patents, and both, particularly free and open-source software, stand to be stifled significantly by widespread software patenting. As seen in the case of the browser war, in the case of Gnutella, and the much more widely used basic utilities of the Web - Apache server software, a number of free e-mail servers, and the Perl scripting language - free and open-source software developers provide central chunks of the logical layer. They do so in a way that leaves that layer open for anyone to use and build upon. The drive to increase the degree of exclusivity available for software by adopting patents over and above copyright threatens the continued vitality of this development methodology. In particular, it threatens to take certain discrete application areas that may require access to patented standard elements or protocols out of the domain of what can be done by free software. As such, it poses a significant threat to the availability of an open logical layer for at least some forms of network use.

The Content Layer

The last set of resources necessary for information production and exchange is the universe of existing information, knowledge, and culture. The battle over the scope, breadth, extent, and enforcement of copyright, patent, trademarks, and a variety of exotic rights like trespass to chattels or the right to link has been the subject of a large legal literature. Instead of covering the entire range of enclosure efforts of the past decade or more, I offer a set of brief descriptions of the choices being made in this domain. The intention is not to criticize or judge the intrinsic logic of any of these legal changes, but merely to illustrate how all these toggles of institutional ecology are being set in favor of proprietary strategies, at the expense of nonproprietary producers.

Copyright

The first domain in which we have seen a systematic preference for commercial producers that rely on property over commons-based producers is in copyright. This preference arises from a combination of expansive interpretations of what rights include, a niggardly interpretive attitude toward users' privileges, especially fair use, and increased criminalization. These have made copyright law significantly more industrial-production friendly than it was in the past or than it need be from the perspective of optimizing creativity or welfare in the networked information economy, rather than rent-extraction by incumbents.

Right to Read. Jessica Litman early diagnosed an emerging new "right to read." [/23](#) The basic right of copyright, to control copying, was never seen to include the right to control who reads an existing copy, when, and how many times. Once a user bought a copy, he or she could read it many times, lend it to a friend, or leave it on the park bench or in the library for anyone else to read. This provided a coarse valve to limit the deadweight loss associated with appropriating a public good like information. As a happenstance of computer technology, reading on a screen involves making a temporary copy of a file onto the temporary memory of the computer. An early decision of the Ninth Circuit Court of Appeals, *MAI Systems*, treated RAM (random-access memory) copies of this sort as "copies" for purposes of copyright. [/24](#) This position, while weakly defended, was not later challenged or rejected by other courts. Its result is that every act of reading on a screen involves

"making a copy" within the meaning of the Copyright Act. As a practical matter, this interpretation expands the formal rights of copyright holders to cover any and all computer-mediated uses of their works, because no use can be made with a computer without at least formally implicating the right to copy. More important than the formal legal right, however, this universal baseline claim to a right to control even simple reading of one's copyrighted work marked a change in attitude. Justified later through various claims - such as the efficiency of private ordering or of price discrimination - it came to stand for a fairly broad proposition: Owners should have the right to control all valuable uses of their works. Combined with the possibility and existence of technical controls on actual use and the DMCA's prohibition on circumventing those controls, this means that copyright law has shifted. It existed throughout most of its history as a regulatory provision that reserved certain uses of works for exclusive control by authors, but left other, not explicitly constrained uses free. It has now become a law that gives rights holders the exclusive right to control any computer-mediated use of their works, and captures in its regulatory scope all uses that were excluded from control in prior media.

Fair Use Narrowed. Fair use in copyright was always a judicially created concept with a large degree of uncertainty in its application. This uncertainty, coupled with a broader interpretation of what counts as a commercial use, a restrictive judicial view of what counts as fair, and increased criminalization have narrowed its practical scope.

First, it is important to recognize that the theoretical availability of the fair-use doctrine does not, as a practical matter, help most productions. This is due to a combination of two factors: (1) fair-use doctrine is highly fact specific and uncertain in application, and (2) the Copyright Act provides large fixed statutory damages, even if there is no actual damage to the copyright owner. Lessig demonstrated this effect most clearly by working through an example of a documentary film. [/25](#) A film will not be distributed without liability insurance. Insurance, in turn, will not be issued without formal clearance, or permission, from the owner of each copyrighted work, any portion of which is included in the film, even if the amount used is trivially small and insignificant to the documentary. A five-second snippet of a television program that happened to play on a television set in the background of a sequence captured in documentary film can therefore prevent distribution of the film, unless the filmmaker can persuade the owner of that program to grant rights to use the materials. Copyright owners in such television programs may demand thousands of dollars for even such a minimal and incidental use of "their" images. This is not because a court would ultimately find that using the image as is, with the tiny fraction of the television program in the background, was not covered by fair use. It probably would be a fair use. It is because insurance companies and distributors would refuse to incur the risk of litigation.

Second, in the past few years, even this uncertain scope has been constricted by expanding the definitions of what counts as interference with a market and what counts as a commercial use. Consider the *Free Republic* case. In that case, a political Web site offered a forum for users to post stories from various newspapers as grist for a political discussion of their contents or their slant. The court held that because newspapers may one day sell access to archived articles, and because some users may read some articles on the Web forum instead of searching and retrieving them from the newspapers' archive, the use interfered with a potential market. Moreover, because Free Republic received donations from users (although it did not require them) and exchanged advertising

arrangements with other political sites, the court treated the site as a "commercial user," and its use of newspaper articles to facilitate political discussion of them "a commercial use." These factors enabled the court to hold that posting an article from a medium-daily newspapers-whose existence does not depend on copyright, in a way that may one day come to have an effect on uncertain future revenues, which in any case would be marginal to the business model of the newspapers, was not a fair use even when done for purposes of political commentary.

Criminalization. Copyright enforcement has also been substantially criminalized in the past few years. Beginning with the No Electronic Theft Act (NET Act) in 1997 and later incorporated into the DMCA, criminal copyright has recently become much more expansive than it was until a few years ago. Prior to passage of the NET Act, only commercial pirates-those that slavishly made thousands of copies of video or audiocassettes and sold them for profit - would have qualified as criminal violators of copyright. Criminal liability has now been expanded to cover private copying and free sharing of copyrighted materials whose cumulative nominal price (irrespective of actual displaced demand) is quite low. As criminal copyright law is currently written, many of the tens of millions using p2p networks are felons. It is one thing when the recording industry labels tens of millions of individuals in a society "pirates" in a rhetorical effort to conform social norms to its members' business model. It is quite another when the state brands them felons and fines or imprisons them. Litman has offered the most plausible explanation of this phenomenon. [26] As the network makes low-cost production and exchange of information and culture easier, the large-scale commercial producers are faced with a new source of competition - volunteers, people who provide information and culture for free. As the universe of people who can threaten the industry has grown to encompass more or less the entire universe of potential customers, the plausibility of using civil actions to force individuals to buy rather than share information goods decreases. Suing all of one's intended customers is not a sustainable business model. In the interest of maintaining the business model that relies on control over information goods and their sale as products, the copyright industry has instead enlisted criminal enforcement by the state to prevent the emergence of such a system of free exchange. These changes in formal law have, in what is perhaps a more important development, been coupled with changes in the Justice Department's enforcement policy, leading to a substantial increase in the shadow of criminal enforcement in this area. [27]

Term Extension. The change in copyright law that received the most widespread public attention was the extension of copyright term in the Sonny Bono Copyright Term Extension Act of 1998. The statute became cause celebre in the early 2000s because it was the basis of a major public campaign and constitutional challenge in the case of *Eldred v. Ashcroft*. [28] The actual marginal burden of this statute on use of existing materials could be seen as relatively small. The length of copyright protection was already very long - seventy-five years for corporate-owned materials, life of the author plus fifty for materials initially owned by human authors. The Sonny Bono Copyright Term Extension Act increased these two numbers to ninety-five and life plus seventy, respectively. The major implication, however, was that the Act showed that retroactive extension was always available. As materials that were still valuable in the stocks of Disney, in particular, came close to the public domain, their lives would be extended indefinitely. The legal challenge to the statute brought to public light the fact that, as a practical matter, almost the entire stock of twentieth-century culture and beyond would stay privately owned, and its copyright would be renewed indefinitely. For video

and sound recordings, this meant that almost the entire universe of materials would never become part of the public domain; would never be available for free use as inputs into nonproprietary production. The U.S. Supreme Court upheld the retroactive extension. The inordinately long term of protection in the United States, initially passed under the pretext of "harmonizing" the length of protection in the United States and in Europe, is now being used as an excuse to "harmonize" the length of protection for various kinds of materials - like sound recordings - that actually have shorter terms of protection in Europe or other countries, like Australia. At stake in all these battles is the question of when, if ever, will Errol Flynn's or Mickey Mouse's movies, or Elvis's music, become part of the public domain? When will these be available for individual users on the same terms that Shakespeare or Mozart are available? The implication of *Eldred* is that they may never join the public domain, unless the politics of term-extension legislation change.

No de Minimis Digital Sampling. A narrower, but revealing change is the recent elimination of digital sampling from the universe of *ex ante* permissible actions, even when all that is taken is a tiny snippet. The case is recent and has not been generalized by other courts as of this writing. However, it offers insight into the mind-set of judges who are confronted with digital opportunities, and who in good faith continue to see the stakes as involving purely the organization of a commercial industry, rather than defining the comparative scope of commercial industry and nonmarket commons-based creativity. Courts seem blind to the effects of their decisions on the institutional ecology within which nonproprietary, individual, and social creation must live. In *Bridgeport Music, Inc.*, the Sixth Circuit was presented with the following problem: The defendant had created a rap song.[/29] In making it, he had digitally copied a two-second guitar riff from a digital recording of a 1970s song, and then looped and inserted it in various places to create a completely different musical effect than the original. The district court had decided that the amount borrowed was so small as to make the borrowing de minimis - too little for the law to be concerned with. The Court of Appeals, however, decided that it would be too burdensome for courts to have to decide, on a case-by-case basis, how much was too little for law to be concerned with. Moreover, it would create too much uncertainty for recording companies; it is, as the court put it, "cheaper to license than to litigate."[/30] The court therefore held that any digital sampling, no matter how trivial, could be the basis of a copyright suit. Such a bright-line rule that makes all direct copying of digital bits, no matter how small, an infringement, makes digital sound recordings legally unavailable for noncommercial, individually creative mixing. There are now computer programs, like Garage Band, that allow individual users to cut and mix existing materials to create their own music. These may not result in great musical compositions. But they may. That, in any event, is not their point. They allow users to have a very different relationship to recorded music than merely passively listening to finished, unalterable musical pieces. By imagining that the only parties affected by copyright coverage of sampling are recording artists who have contracts with recording studios and seek to sell CDs, and can therefore afford to pay licensing fees for every two-second riff they borrow, the court effectively outlawed an entire model of user creativity. Given how easy it is to cut, paste, loop, slow down, and speed up short snippets, and how creatively exhilarating it is for users - young and old - to tinker with creating musical compositions with instruments they do not know how to play, it is likely that the opinion has rendered illegal a practice that will continue, at least for the time being. Whether the social practice will ultimately cause the law to change or vice versa is more difficult to predict.

Contractual Enclosure: Click-Wrap Licenses and the Uniform Computer Information Transactions Act (UCITA)

Practically all academic commentators on copyright law - whether critics or proponents of this provision or that - understand copyright to be a public policy accommodation between the goal of providing incentives to creators and the goal of providing efficiently priced access to both users and downstream creators. Ideally, it takes into consideration the social costs and benefits of one settlement or another, and seeks to implement an optimal trade-off. Beginning in the 1980s, software and other digital goods were sold with "shrink-wrap licenses." These were licenses to use the software, which purported to apply to mass-market buyers because the buyer would be deemed to have accepted the contract by opening the packaging of the software. These practices later transmuted online into click-wrap licenses familiar to most anyone who has installed software and had to click "I Agree" once or more before the software would install. Contracts are not bound by the balance struck in public law. Licensors can demand, and licensees can agree to, almost any terms. Among the terms most commonly inserted in such licenses that restrict the rights of users are prohibitions on reverse engineering, and restrictions on the use of raw data in compilations, even though copyright law itself does not recognize rights in data. As Mark Lemley showed, most courts prior to the mid-1990s did not enforce such terms.[31] Some courts refused to enforce shrink-wrap licenses in mass-market transactions by relying on state contract law, finding an absence of sufficient consent or an unenforceable contract of adhesion. Others relied on federal preemption, stating that to the extent state contract law purported to enforce a contract that prohibited fair use or otherwise protected material in the public domain - like the raw information contained in a report - it was preempted by federal copyright law that chose to leave this material in the public domain, freely usable by all. In 1996, in *ProCD v. Zeidenberg*, the Seventh Circuit held otherwise, arguing that private ordering would be more efficient than a single public determination of what the right balance was.[32]

The following few years saw substantial academic debate as to the desirability of contractual opt-outs from the public policy settlement. More important, the five years that followed saw a concerted effort to introduce a new part to the Uniform Commercial Code (UCC) - a model commercial law that, though nonbinding, is almost universally adopted at the state level in the United States, with some modifications. The proposed new UCC Article 2B was to eliminate the state law concerns by formally endorsing the use of standard shrink-wrap licenses. The proposed article generated substantial academic and political heat, ultimately being dropped by the American Law Institute, one of the main sponsors of the UCC. A model law did ultimately pass under the name of the Uniform Computer Information Transactions Act (UCITA), as part of a less universally adopted model law effort. Only two states adopted the law - Virginia and Maryland. A number of other states then passed anti-UCITA laws, which gave their residents a safe harbor from having UCITA applied to their click-wrap transactions.

The reason that *ProCD* and UCITA generated so much debate was the concern that click-wrap licenses were operating in an inefficient market, and that they were, as a practical matter, displacing the policy balance represented by copyright law. Mass-market transactions do not represent a genuine negotiated agreement, in the individualized case, as to what the efficient contours of

permissions are for the given user and the given information product. They are, rather, generalized judgments by the vendor as to what terms are most attractive for it that the market will bear. Unlike rival economic goods, information goods sold at a positive price in reliance on copyright are, by definition, priced above marginal cost. The information itself is nonrival. Its marginal cost is zero. Any transaction priced above the cost of communication is evidence of some market power in the hands of the provider, used to price based on value and elasticity of demand, not on marginal cost. Moreover, the vast majority of users are unlikely to pay close attention to license details they consider to be boilerplate. This means there is likely significant information shortfall on the part of consumers as to the content of the licenses, and the sensitivity of demand to overreaching contract terms is likely low. This is not because consumers are stupid or slothful, but because the probability that either they would be able to negotiate out from under a standard provision, or a court would enforce against them a truly abusive provision is too low to justify investing in reading and arguing about contracts for all but their largest purchases. In combination, these considerations make it difficult to claim as a general matter that privately set licensing terms would be more efficient than the publicly set background rules of copyright law./33 The combination of mass-market contracts enforced by technical controls over use of digital materials, which in turn are protected by the DMCA, threatens to displace the statutorily defined public domain with a privately defined realm of permissible use./34 This privately defined settlement would be arrived at in non-negotiated mass-market transactions, in the presence of significant information asymmetries between consumers and vendors, and in the presence of systematic market power of at least some degree.

Trademark Dilution

As discussed in chapter 8, the centrality of commercial interaction to social existence in early-twenty-first-century America means that much of our core iconography is commercial in origin and owned as a trademark. Mickey, Barbie, Playboy, or Coke are important signifiers of meaning in contemporary culture. Using iconography is a central means of creating rich, culturally situated expressions of one's understanding of the world. Yet, as Boyle has pointed out, now that we treat flag burning as a constitutionally protected expression, trademark law has made commercial icons the sole remaining venerable objects in our law. Trademark law permits the owners of culturally significant images to control their use, to squelch criticism, and to define exclusively the meaning that the symbols they own carry.

Three factors make trademark protection today more of a concern as a source of enclosure than it might have been in the past. First is the introduction of the federal Anti-Dilution Act of 1995. Second is the emergence of the brand as the product, as opposed to a signifier for the product. Third is the substantial reduction in search and other information costs created by the Net. Together, these three factors mean that owned symbols are becoming increasingly important as cultural signifiers, are being enclosed much more extensively than before precisely as cultural signifiers, and with less justification beyond the fact that trademarks, like all exclusive rights, are economically valuable to their owners.

In 1995, Congress passed the first federal Anti-Dilution Act. Though treated as a trademark protection law, and codifying doctrines that arose in trademark common law, antidilution is a

fundamentally different economic right than trademark protection. Traditional trademark protection is focused on preventing consumer confusion. It is intended to assure that consumers can cheaply tell the difference between one product and another, and to give producers incentives to create consistent quality products that can be associated with their trademark. Trademark law traditionally reflected these interests. Likelihood of consumer confusion was the sine qua non of trademark infringement. If I wanted to buy a Coca-Cola, I did not want to have to make sure I was not buying a different dark beverage in a red can called Coca-Gola. Infringement actions were mostly limited to suits among competitors in similar relevant markets, where confusion could occur. So, while trademark law restricted how certain symbols could be used, it was so only as among competitors, and only as to the commercial, not cultural, meaning of their trademark. The antidilution law changes the most relevant factors. It is intended to protect famous brand names, irrespective of a likelihood of confusion, from being diluted by use by others. The association between a particular corporation and a symbol is protected for its value to that corporation, irrespective of the use. It no longer regulates solely competitors to the benefit of competition. It prohibits many more possible uses of the symbol than was the case under traditional trademark law. It applies even to noncommercial users where there is no possibility of confusion. The emergence of this antidilution theory of exclusivity is particularly important as brands have become the product itself, rather than a marker for the product. Nike and Calvin Klein are examples: The product sold in these cases is not a better shoe or shirt - the product sold is the brand. And the brand is associated with a cultural and social meaning that is developed purposefully by the owner of the brand so that people will want to buy it. This development explains why dilution has become such a desirable exclusive right for those who own it. It also explains the cost of denying to anyone the right to use the symbol, now a signifier of general social meaning, in ways that do not confuse consumers in the traditional trademark sense, but provide cultural criticism of the message signified.

Ironically, the increase in the power of trademark owners to control uses of their trademark comes at a time when its functional importance as a mechanism for reducing search costs is declining. Traditional trademark's most important justification was that it reduced information collection costs and thereby facilitated welfare-enhancing trade. In the context of the Internet, this function is significantly less important. General search costs are lower. Individual items in commerce can provide vastly greater amounts of information about their contents and quality. Users can use machine processing to search and sift through this information and to compare views and reviews of specific items. Trademark has become less, rather than more, functionally important as a mechanism for dealing with search costs. When we move in the next few years to individual-item digital marking, such as with RFID (radio frequency identification) tags, all the relevant information about contents, origin, and manufacture down to the level of the item, as opposed to the product line, will be readily available to consumers in real space, by scanning any given item, even if it is not otherwise marked at all. In this setting, where the information qualities of trademarks will significantly decline, the antidilution law nonetheless assures that owners can control the increasingly important cultural meaning of trademarks. Trademark, including dilution, is subject to a fair use exception like that of copyright. For the same reasons as operated in copyright, however, the presence of such a doctrine only ameliorates, but does not solve, the limits that a broad exclusive right places on the capacity of nonmarket-oriented creative uses of materials - in this case, culturally meaningful symbols.

Database Protection

In 1991, in *Feist Publications, Inc. v. Rural Tel. Serv. Co.*, the Supreme Court held that raw facts in a compilation, or database, were not covered by the Copyright Act. The constitutional clause that grants Congress the power to create exclusive rights for authors, the Court held, required that works protected were original with the author. The creative element of the compilation - its organization or selectivity, for example, if sufficiently creative - could therefore be protected under copyright law. However, the raw facts compiled could not. Copying data from an existing compilation was therefore not "piracy"; it was not unfair or unjust; it was purposefully privileged in order to advance the goals of the constitutional power to make exclusive grants - the advancement of progress and creative uses of the data./35 A few years later, the European Union passed a Database Directive, which created a discrete and expansive right in raw data compilations./36 The years since the Court decided *Feist* have seen repeated efforts by the larger players in the database publishing industry to pass similar legislation in the United States that would, as a practical matter, overturn *Feist* and create exclusive private rights in the raw data in compilations. "Harmonization" with Europe has been presented as a major argument in favor of this law. Because the *Feist* Court based its decision on limits to the constitutional power to create exclusive rights in raw information, efforts to protect database providers mostly revolved around an unfair competition law, based in the Commerce Clause, rather than on precisely replicating the European right. In fact, however, the primary draft that has repeatedly been introduced walks, talks, and looks like a property right.

Sustained and careful work, most prominently by Jerome Reichman and Paul Uhlir, has shown that the proposed database right is unnecessary and detrimental, particularly to scientific research. /37 Perhaps no example explains this point better than the "natural experiment" that Boyle has pointed to, and which the United States and Europe have been running over the past decade or so. The United States has formally had no exclusive right in data since 1991. Europe has explicitly had such a right since 1996. One would expect that both the European Union and the United States would look to the comparative effects on the industries in both places when the former decides whether to keep its law, and the latter decides whether to adopt one like it. The evidence is reasonably consistent and persuasive. Following the *Feist* decision, the U.S. database industry continued to grow steadily, without a blip. The "removal" of the property right in data by *Feist* had no effect on growth. Europe at the time had a much smaller database industry than did the United States, as measured by the number of databases and database companies. Maurer, Hugenholz, and Onsrud showed that, following the introduction of the European sui generis right, each country saw a one-time spike in the number of databases and new database companies, but this was followed within a year or two by a decline to the levels seen before the Directive, which have been fairly stagnant since the early 1990s./38 Another study, more specifically oriented toward the appropriate policy for government-collected data, compared the practices of Europe - where government agencies are required to charge what the market will bear for access to data they collect - and the United States, where the government makes data it collects freely available at the cost of reproduction, as well as for free on the Web. That study found that the secondary uses of data, including commercial- and noncommercial-sector uses - such as, for example, markets in commercial risk management and meteorological services - contributed vastly more to the economy of the United States because of secondary uses of freely accessed government weather data than

equivalent market sectors in Europe were able to contribute to their respective economies. [/39](#) The evidence suggests, then, that the artificial imposition of rents for proprietary data is suppressing growth in European market-based commercial services and products that rely on access to data, relative to the steady growth in the parallel U.S. markets, where no such right exists. It is trivial to see that a cost structure that suppresses growth among market-based entities that would at least partially benefit from being able to charge more for their outputs would have an even more deleterious effect on nonmarket information production and exchange activities, which are burdened by the higher costs and gain no benefit from the proprietary rights.

There is, then, mounting evidence that rights in raw data are unnecessary to create a basis for a robust database industry. Database manufacturers rely on relational contracts - subscriptions to continuously updated databases - rather than on property-like rights. The evidence suggests that, in fact, exclusive rights are detrimental to various downstream industries that rely on access to data. Despite these fairly robust observations from a decade of experience, there continues to be a threat that such a law will pass in the U.S. Congress. This continued effort to pass such a law underscores two facts. First, much of the legislation in this area reflects rent seeking, rather than reasoned policy. Second, the deeply held belief that "more property-like rights will lead to more productivity" is hard to shake, even in the teeth of both theoretical analysis and empirical evidence to the contrary.

Linking and Trespass to Chattels: New Forms of Information Exclusivity

Some litigants have turned to state law remedies to protect their data indirectly, by developing a common-law, trespass-to-server form of action. The primary instance of this trend is *eBay v. Bidder's Edge*, a suit by the leading auction site against an aggregator site. Aggregators collect information about what is being auctioned in multiple locations, and make the information about the items available in one place so that a user can search eBay and other auction sites simultaneously. The eventual bidding itself is done on the site that the item's owner chose to make his or her item available, under the terms required by that site. The court held that the automated information collection process - running a computer program that automatically requests information from the server about what is listed on it, called a spider or a bot - was a "trespass to chattels." [/40](#) This ancient form of action, originally intended to apply to actual taking or destruction of goods, mutated into a prohibition on unlicensed automated searching. The injunction led to Bidder's Edge closing its doors before the Ninth Circuit had an opportunity to review the decision. A common-law decision like *eBay v. Bidder's Edge* creates a common-law exclusive private right in information by the back door. In principle, the information itself is still free of property rights. Reading it mechanically - an absolute necessity given the volume of the information and its storage on magnetic media accessible only by mechanical means - can, however, be prohibited as "trespass." The practical result would be equivalent to some aspects of a federal exclusive private right in raw data, but without the mitigating attributes of any exceptions that would be directly introduced into legislation. It is still too early to tell whether cases such as these ultimately will be considered preempted by federal copyright law, [/41](#) or perhaps would be limited by first amendment law on the model of *New York Times v. Sullivan*.[/42](#)

Beyond the roundabout exclusivity in raw data, trespass to chattels presents one instance of a broader question that is arising in application of both common-law and statutory provisions. At

stake is the legal control over information about information, like linking and other statements people make about the availability and valence of some described information. Linking - the mutual pointing of many documents to each other - is the very core idea of the World Wide Web. In a variety of cases, parties have attempted to use law to control the linking practices of others. The basic structure of these cases is that A wants to tell users M and N about information presented by B. The meaning of a link is, after all, "here you can read information presented by someone other than me that I deem interesting or relevant to you, my reader." Someone, usually B, but possibly some other agent C, wants to control what M and N know or do with regard to the information B is presenting. B (or C) then sues A to prevent A from linking to the information on B's site.

The simplest instance of such a case involved a service that Microsoft offered - sidewalk.com - that provided access to, among other things, information on events in various cities. If a user wanted a ticket to the event, the sidewalk site linked that user directly to a page on ticketmaster.com where the user could buy a ticket. Ticketmaster objected to this practice, preferring instead that sidewalk.com link to its home page, in order to expose the users to all the advertising and services Ticketmaster provided, rather than solely to the specific service sought by the user referred by sidewalk.com. At stake in these linking cases is who will control the context in which certain information is presented. If deep linking is prohibited, Ticketmaster will control the context - the other movies or events available to be seen, their relative prominence, reviews, and so forth. The right to control linking then becomes a right to shape the meaning and relevance of one's statements for others. If the choice between Ticketmaster and Microsoft as controllers of the context of information may seem of little normative consequence, it is important to recognize that the right to control linking could easily apply to a local library, or church, or a neighbor as they participate in peer-producing relevance and accreditation of the information to which they link.

The general point is this: On the Internet, there are a variety of ways that some people can let others know about information that exists somewhere on the Web. In doing so, these informers loosen someone else's control over the described information - be it the government, a third party interested in limiting access to the information, or the person offering the information. In a series of instances over the past half decade or more we have seen attempts by people who control certain information to limit the ability of others to challenge that control by providing information about the information. These are not cases in which a person without access to information is seeking affirmative access from the "owner" of information. These are cases where someone who dislikes what another is saying about particular information is seeking the aid of law to control what other parties can say to each other about that information. Understood in these terms, the restrictive nature of these legal moves in terms of how they burden free speech in general, and impede the freedom of anyone, anywhere, to provide information, relevance, and accreditation, becomes clear. The *eBay v. Bidder's Edge* case suggests one particular additional aspect. While much of the political attention focuses on formal "intellectual property"-style statutes passed by Congress, in the past few years we have seen that state law and common-law doctrine are also being drafted to create areas of exclusivity and boundaries on the free use of information. These efforts are often less well informed, and because they were arrived at ad hoc, often without understanding that they are actually forms of regulating information production and exchange, they include none of the balancing privileges or limitations of rights that are so common in the formal statutory frameworks.

International "Harmonization"

One theme that has repeatedly appeared in the discussion of databases, the DMCA, and term extension, is the way in which "harmonization" and internationalization of exclusive rights are used to ratchet up the degree of exclusivity afforded rights holders. It is trite to point out that the most advanced economies in the world today are information and culture exporters. This is true of both the United States and Europe. Some of the cultural export industries - most notably Hollywood, the recording industry, some segments of the software industry, and pharmaceuticals - have business models that rely on the assertion of exclusive rights in information. Both the United States and the European Union, therefore, have spent the past decade and a half pushing for ever-more aggressive and expansive exclusive rights in international agreements and for harmonization of national laws around the world toward the highest degrees of protection. Chapter 9 discusses in some detail why this was not justified as a matter of economic rationality, and why it is deleterious as a matter of justice. Here, I only note the characteristic of internationalization and harmonization as a one-way ratchet toward ever-expanding exclusivity.

Take a simple provision like the term of copyright protection. In the mid-1990s, Europe was providing for many works (but not all) a term of life of the author plus seventy years, while the United States provided exclusivity for the life of the author plus fifty. A central argument for the Sonny Bono Copyright Term Extension Act of 1998 was to "harmonize" with Europe. In the debates leading up to the law, one legislator actually argued that if our software manufacturers had a shorter term of copyright, they would be disadvantaged relative to the European firms. This argument assumes, of course, that U.S. software firms could stay competitive in the software business by introducing nothing new in software for seventy-five years, and that it would be the loss of revenues from products that had not been sufficiently updated for seventy-five years to warrant new copyright that would place them at a disadvantage. The newly extended period created by the Sonny Bono Copyright Term Extension Act is, however, longer in some cases than the protection afforded in Europe. Sound recordings, for example, are protected for fifty years in Europe. The arguments are now flowing in the opposite direction-harmonization toward the American standard for all kinds of works, for fear that the recordings of Elvis or the Beatles will fall into the European public domain within a few paltry years. "Harmonization" is never invoked to de-escalate exclusivity - for example, as a reason to eliminate the European database right in order to harmonize with the obviously successful American model of no protection, or to shorten the length of protection for sound recordings in the United States.

International agreements also provide a fertile forum for ratcheting up protection. Lobbies achieve a new right in a given jurisdiction - say an extension of term, or a requirement to protect technological protection measures on the model of the DMCA. The host country, usually the United States, the European Union, or both, then present the new right for treaty approval, as the United States did in the context of the WIPO treaties in the mid-1990s. Where this fails, the United States has more recently begun to negotiate bilateral free trade agreements (FTAs) with individual nations. The structure of negotiation is roughly as follows: The United States will say to Thailand, or India, or whoever the trading partner is: If you would like preferential treatment of your core export, say textiles or rice, we would like you to include this provision or that in your

domestic copyright or patent law. Once this is agreed to in a number of bilateral FTAs, the major IP exporters can come back to the multilateral negotiations and claim an emerging international practice, which may provide more exclusivity than their then applicable domestic law. With changes to international treaties in hand, domestic resistance to legislation can be overcome, as we saw in the United States when the WIPO treaties were used to push through Congress the DMCA anticircumvention provisions that had failed to pass two years earlier. Any domestic efforts to reverse and limit exclusivity then have to overcome substantial hurdles placed by the international agreements, like the agreement on Trade Related Aspects of Intellectual Property (TRIPS). The difficulty of amending international agreements to permit a nation to decrease the degree of exclusivity it grants copyright or patent holders becomes an important one-way ratchet, preventing de-escalation.

Countervailing Forces

As this very brief overview demonstrates, most of the formal institutional moves at the content layer are pushing toward greater scope and reach for exclusive rights in the universe of existing information, knowledge, and cultural resources. The primary countervailing forces in the content layer are similar to the primary countervailing forces in the logical layer - that is, social and cultural push - back against exclusivity. Recall how central free software and the open, cooperative, nonproprietary standard-setting processes are to the openness of the logical layer. In the content layer, we are seeing the emergence of a culture of free creation and sharing developing as a countervailing force to the increasing exclusivity generated by the public, formal lawmaking system. The Public Library of Science discussed in chapter 9 is an initiative of scientists who, frustrated with the extraordinarily high journal costs for academic journals, have begun to develop systems for scientific publication whose outputs are immediately and freely available everywhere. The Creative Commons is an initiative to develop a series of licenses that allow individuals who create information, knowledge, and culture to attach simple licenses that define what others may, or may not, do with their work. The innovation represented by these licenses relative to the background copyright system is that they make it trivial for people to give others permission to use their creations. Before their introduction, there were no widely available legal forms to make it clear to the world that it is free to use my work, with or without restrictions. More important than the institutional innovation of Creative Commons is its character as a social movement. Under the moniker of the "free culture" movement, it aims to encourage widespread adoption of sharing one's creations with others. What a mature movement like the free software movement, or nascent movements like the free culture movement and the scientists' movement for open publication and open archiving are aimed at is the creation of a legally self-reinforcing domain of open cultural sharing. They do not negate property-like rights in information, knowledge, and culture. Rather, they represent a self-conscious choice by their participants to use copyrights, patents, and similar rights to create a domain of resources that are free to all for common use.

Alongside these institutionally instantiated moves to create a self-reinforcing set of common resources, there is a widespread, global culture of ignoring exclusive rights. It is manifest in the widespread use of file-sharing software to share copyrighted materials. It is manifest in the widespread acclaim that those who crack copy-protection mechanisms receive. This culture has

developed a rhetoric of justification that focuses on the overreaching of the copyright industries and on the ways in which the artists themselves are being exploited by rights holders. While clearly illegal in the United States, there are places where courts have sporadically treated participation in these practices as copying for private use, which is exempted in some countries, including a number of European countries. In any event the sheer size of this movement and its apparent refusal to disappear in the face of lawsuits and public debate present a genuine countervailing pressure against the legal tightening of exclusivity. As a practical matter, efforts to impose perfect private ordering and to limit access to the underlying digital bits in movies and songs through technical means have largely failed under the sustained gaze of the community of computer scientists and hackers who have shown its flaws time and again. Moreover, the mechanisms developed in response to a large demand for infringing file-sharing utilities were the very mechanisms that were later available to the Swarthmore students to avoid having the Diebold files removed from the Internet and that are shared by other censorship-resistant publication systems. The tools that challenge the "entertainment-as-finished-good" business model are coming into much wider and unquestionably legitimate use. Litigation may succeed in dampening use of these tools for copying, but also creates a heightened political awareness of information-production regulation. The same students involved in the Diebold case, radicalized by the lawsuit, began a campus "free culture" movement. It is difficult to predict how this new political awareness will play out in a political arena - the making of copyrights, patents, and similar exclusive rights - that for decades has functioned as a technical backwater that could never invoke a major newspaper editorial, and was therefore largely controlled by the industries whose rents it secured.

The Problem of Security

This book as a whole is dedicated to the emergence of commons-based information production and its implications for liberal democracies. Of necessity, the emphasis of this chapter too is on institutional design questions that are driven by the conflict between the industrial and networked information economies. Orthogonal to this conflict, but always relevant to it, is the perennial concern of communications policy with security and crime. Throughout much of the 1990s, this concern manifested primarily as a conflict over encryption. The "crypto-wars," as they were called, revolved around the FBI's efforts to force industry to adopt technology that had a backdoor - then called the "Clipper Chip" - that would facilitate wiretapping and investigation. After retarding encryption adoption in the United States for almost a decade, the federal government ultimately decided that trying to hobble security in most American systems (that is, forcing everyone to adopt weaker encryption) in order to assure that the FBI could better investigate the failures of security that would inevitably follow use of such weak encryption was a bad idea. The fact that encryption research and business was moving overseas - giving criminals alternative sources for obtaining excellent encryption tools while the U.S. industry fell behind - did not help the FBI's cause. The same impulse is to some extent at work again, with the added force of the post-9/11 security mind-set.

One concern is that open wireless networks are available for criminals to hide their tracks - the criminal uses someone else's Internet connection using their unencrypted WiFi access point, and when the authorities successfully track the Internet address back to the WiFi router, they find an

innocent neighbor rather than the culprit. This concern has led to some proposals that manufacturers of WiFi routers set their defaults so that, out of the box, the router is encrypted. Given how "sticky" defaults are in technology products, this would have enormously deleterious effects on the development of open wireless networks. Another concern is that free and open-source software reveals its design to anyone who wants to read it. This makes it easier to find flaws that could be exploited by attackers and nearly impossible to hide purposefully designed weaknesses, such as susceptibility to wiretapping. A third is that a resilient, encrypted, anonymous peer-to-peer network, like FreeNet or some of the major p2p architectures, offers the criminals or terrorists communications systems that are, for all practical purposes, beyond the control of law enforcement and counterterrorism efforts. To the extent that they take this form, security concerns tend to support the agenda of the proprietary producers.

However, security concerns need not support proprietary architectures and practices. On the wireless front, there is a very wide range of anonymization techniques available for criminals and terrorists who use the Internet to cover their tracks. The marginally greater difficulty that shutting off access to WiFi routers would impose on determined criminals bent on covering their tracks is unlikely to be worth the loss of an entire approach toward constructing an additional last-mile loop for local telecommunications. One of the core concerns of security is the preservation of network capacity as a critical infrastructure. Another is assuring communications for critical security personnel. Open wireless networks that are built from ad hoc, self-configuring mesh networks are the most robust design for a local communications loop currently available. It is practically impossible to disrupt local communications in such a network, because these networks are designed so that each router will automatically look for the next available neighbor with which to make a network. These systems will self-heal in response to any attack on communications infrastructure as a function of their basic normal operational design. They can then be available both for their primary intended critical missions and for first responders as backup data networks, even when main systems have been lost - as they were, in fact, lost in downtown Manhattan after the World Trade Center attack. To imagine that security is enhanced by eliminating the possibility that such a backup local communications network will emerge in exchange for forcing criminals to use more anonymizers and proxy servers instead of a neighbor's WiFi router requires a very narrow view of security. Similarly, the same ease of study that makes flaws in free software observable to potential terrorists or criminals makes them available to the community of developers, who quickly shore up the defenses of the programs. Over the past decade, security flaws in proprietary programs, which are not open to inspection by such large numbers of developers and testers, have been much more common than security breaches in free software. Those who argue that proprietary software is more secure and allows for better surveillance seem to be largely rehearsing the thought process that typified the FBI's position in the Clipper Chip debate.

More fundamentally, the security concerns represent a lack of ease with the great freedom enabled by the networked information environment. Some of the individuals who can now do more alone and in association with others want to do harm to the United States in particular, and to advanced liberal market-based democracies more generally. Others want to trade Nazi memorabilia or child pornography. Just as the Internet makes it harder for authoritarian regimes to control their populations, so too the tremendous openness and freedom of the networked environment requires

new ways of protecting open societies from destructive individuals and groups. And yet, particularly in light of the systematic and significant benefits of the networked information economy and its sharing-based open production practices to the core political commitments of liberal democracies, preserving security in these societies by eliminating the technologies that can support improvements in the very freedom being protected is perverse. Given Abu Ghraib and Guantanamo Bay, however, squelching the emergence of an open networked environment and economy hardly seems to be the most glaring of self-defeating moves in the war to protect freedom and human dignity in liberal societies. It is too early to tell whether the security urge will ultimately weigh in on the side of the industrial information economy incumbents, or will instead follow the path of the crypto-wars, and lead security concerns to support the networked information economy's ability to provide survivable, redundant, and effective critical infrastructures and information production and exchange capabilities. If the former, this impulse may well present a formidable obstacle to the emergence of an open networked information environment.

Notes

1. Paul Starr, *The Creation of the Media: Political Origins of Modern Communications* (New York: Basic Books, 2004).

2. Ithiel de Sola-Pool, *Technologies of Freedom* (Cambridge, MA: Belknap Press, 1983), 91-100.

3. *Bridgeport Music, Inc. v. Dimension Films*, 2004 U.S. App. LEXIS 26877.

4. Other layer-based abstractions have been proposed, most effectively by Lawrence Solum and Minn Chung, *The Layers Principle: Internet Architecture and the Law*, University of San Diego Public Law Research Paper No. 55. Their model more closely hews to the OSI layers, and is tailored to being more specifically usable for a particular legal principle-never regulate at a level lower than you need to. I seek a higherlevel abstraction whose role is not to serve as a tool to constrain specific rules, but as a map for understanding the relationships between diverse institutional elements as they relate to the basic problem of how information is produced and exchanged in society.

5. The first major treatment of this phenomenon was Michael Froomkin, "The Internet as a Source of Regulatory Arbitrage" (1996), http://www.law.miami.edu/froomkin/articles/arbitr.htm .

6. Jonathan Krim, "AOL Blocks Spammers' Web Sites," *Washington Post*, March 20, 2004, p. A01; also available at http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&contentId=A9449-2004Mar19&notFound=true .

7. FCC Report on High Speed Services, December 2003 (Appendix to Fourth 706 Report NOI).

8. 216 F.3d 871 (9th Cir. 2000).

9. *National Cable and Telecommunications Association v. Brand X Internet Services* (decided June 27, 2005).

10. *Turner Broad. Sys. v. FCC*, 512 U.S. 622 (1994) and *Turner Broad. Sys. v. FCC*, 520 U.S. 180 (1997).

11. *Chesapeake & Potomac Tel. Co. v. United States*, 42 F.3d 181 (4th Cir. 1994); *Comcast Cablevision of Broward County, Inc. v. Broward County*, 124 F. Supp. 2d 685, 698 (D. Fla., 2000).

12. The locus classicus of the economists' critique was Ronald Coase, "The Federal Communications Commission," *Journal of Law and Economics* 2 (1959): 1. The best worked-out version of how these property rights would look remains Arthur S. De Vany et al., "A Property System for Market Allocation of the Electromagnetic Spectrum: A Legal-Economic-Engineering Study," *Stanford Law Review* 21 (1969): 1499.

13. *City of Abilene, Texas v. Federal Communications Commission*, 164 F3d 49 (1999).

14. *Nixon v. Missouri Municipal League*, 541 U.S. 125 (2004).

15. Bill Number S. 2048, 107th Congress, 2nd Session.

16. *Felten v. Recording Indust. Assoc. of America Inc.*, No. CV- 01-2669 (D.N.J. June 26, 2001).

17. *Metro-Goldwyn-Mayer v. Grokster, Ltd.* (decided June 27, 2005).

18. See Felix Oberholzer and Koleman Strumpf, "The Effect of File Sharing on Record Sales" (working paper), http://www.unc.edu/cigar/papers/FileSharing_March2004.pdf .

19. Mary Madden and Amanda Lenhart, "Music Downloading, File-Sharing, and Copyright" (Pew, July 2003), http://www.pewinternet.org/pdfs/PIP_Copyright_Memo.pdf .

20. Lee Rainie and Mary Madden, "The State of Music Downloading and File-Sharing Online" (Pew, April 2004), http://www.pewinternet.org/pdfs/PIP_Filesharing_April_04.pdf .

21. See 111 F.Supp.2d at 310, fns. 69-70; *PBS Frontline* report, http://www.pbs.org/wgbh/pages/frontline/shows/hollywood/business/windows.html .

22. A. M. Froomkin, "Semi-Private International Rulemaking: Lessons Learned from the WIPO Domain Name Process," http://www.personal.law.miami.edu/froomkin/articles/TPRC99.pdf .

23. Jessica Litman, "The Exclusive Right to Read," *Cardozo Arts and Entertainment Law Journal* 13 (1994): 29.

24. *MAI Systems Corp. v. Peak Computer, Inc.*, 991 F.2d 511 (9th Cir. 1993).

25. Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (New York: Penguin Press, 2004).

26. Jessica Litman, "Electronic Commerce and Free Speech," *Journal of Ethics and Information Technology* 1 (1999): 213.

27. See Department of Justice Intellectual Property Policy and Programs, http://www.usdoj.gov/criminal/cybercrime/ippolicy.html .

28. *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

29. *Bridgeport Music, Inc. v. Dimension Films*, 383 F.3d 390 (6th Cir.2004).

30. 383 F3d 390, 400.

31. Mark A. Lemley, "Intellectual Property and Shrinkwrap Licenses," *Southern California Law Review* 68 (1995): 1239, 1248-1253.

32. 86 F.3d 1447 (7th Cir. 1996).

33. For a more complete technical explanation, see Yochai Benkler, "An Unhurried View of Private Ordering in Information Transactions," *Vanderbilt Law Review* 53 (2000): 2063.

34. James Boyle, "Cruel, Mean or Lavish? Economic Analysis, Price Discrimination and Digital Intellectual Property," *Vanderbilt Law Review* 53 (2000); Julie E. Cohen, "Copyright and the Jurisprudence of Self-Help," *Berkeley Technology Law Journal* 13 (1998): 1089; Niva Elkin-Koren, "Copyright Policy and the Limits of Freedom of Contract," *Berkeley Technology Law Journal* 12 (1997): 93.

35. *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, 499 U.S. 340, 349-350 (1991).

36. Directive No. 96/9/EC on the legal protection of databases, 1996 O.J. (L 77) 20.

37. J. H. Reichman and Paul F. Uhlir, "Database Protection at the Crossroads: Recent Developments and Their Impact on Science and Technology," *Berkeley Technology Law Journal* 14 (1999): 793; Stephen M. Maurer and Suzanne Scotchmer, "Database Protection: Is It Broken and Should We Fix It?" *Science* 284 (1999): 1129.

38. See Stephen M. Maurer, P. Bernt Hugenholtz, and Harlan J. Onsrud, "Europe's Database Experiment," *Science* 294 (2001): 789; Stephen M. Maurer, "Across Two Worlds: Database Protection in the U.S. and Europe," paper prepared for Industry Canada's Conference on Intellectual Property and Innovation in the Knowledge-Based Economy, May 23-24 2001.

39. Peter Weiss, "Borders in Cyberspace: Conflicting Public Sector Information Policies and their Economic Impacts" (U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, February 2002).

40. *eBay, Inc. v. Bidder's Edge, Inc.*, 2000 U.S. Dist. LEXIS 13326 (N.D.Cal. 2000).

41. The preemption model could be similar to the model followed by the Second Circuit in *NBA v. Motorola*, 105 F.3d 841 (2d Cir. 1997), which restricted state misappropriation claims to narrow bounds delimited by federal policy embedded in the Copyright Act. This might require actual proof that the bots have stopped service, or threaten the service's very existence.

42. *New York Times v. Sullivan*, 376 U.S. 254, 266 (1964).

Chapter 12
Conclusion: The Stakes of Information Law and Policy

Complex modern societies have developed in the context of mass media and industrial information economy. Our theories of growth and innovation assume that industrial models of innovation are dominant. Our theories about how effective communications in complex societies are achieved center on market-based, proprietary models, with a professional commercial core and a dispersed, relatively passive periphery. Our conceptions of human agency, collective deliberation, and common culture in these societies are embedded in the experience and practice of capital-intensive information and cultural production practices that emphasize proprietary, market-based models and starkly separate production from consumption. Our institutional frameworks reflect these conceptual models of information production and exchange, and have come, over the past few years, to enforce these conceptions as practiced reality, even when they need not be.

This book began with four economic observations. First, the baseline conception that proprietary strategies are dominant in our information production system is overstated. The education system, from kindergarten to doctoral programs, is thoroughly infused with nonproprietary motivations, social relations, and organizational forms. The arts and sciences are replete with voluntarism and actions oriented primarily toward social-psychological motivations rather than market appropriation. Political and theological discourses are thoroughly based in nonmarket forms and motivations. Perhaps most surprisingly, even industrial research and development, while market oriented, is in most industries not based on proprietary claims of exclusion, but on improved efficiencies and customer relations that can be captured and that drive innovation, without need for proprietary strategies of appropriation. Despite the continued importance of nonproprietary production in information as a practical matter, the conceptual nuance required to acknowledge its importance ran against the grain of the increasingly dominant thesis that property and markets are the roots of all growth and productivity. Partly as a result of the ideological and military conflict with Communism, partly as a result of the theoretical elegance of a simple and tractable solution, policy makers and their advisers came to believe toward the end of the twentieth century that property in information and innovation was like property in wristwatches and automobiles. The more clearly you defined and enforced it, and the closer it was to perfect exclusive rights, the more production you would get. The rising dominance of this conceptual model combined with the rent-seeking lobbying of industrial-model producers to underwrite a fairly rapid and substantial tipping of the institutional ecology of innovation and information production in favor of proprietary models. The U.S. patent system was overhauled in the early 1980s, in ways that strengthened and broadened the reach and scope of exclusivity. Copyright was vastly expanded in the mid-1970s, and again in the latter 1990s. Trademark was vastly expanded in the 1990s. Other associated rights were created and strengthened throughout these years.

The second economic point is that these expansions of rights operate, as a practical matter, as a tax on nonproprietary models of production in favor of the proprietary models. It makes access to information resources more expensive for all, while improving appropriability only for some. Introducing software patents, for example, may help some of the participants in the one-third of the software industry that depends on sales of finished software items. But it clearly raises the costs without increasing benefits for the two-thirds of the industry that is service based and relational. As a practical matter, the substantial increases in the scope and reach of exclusive rights have adversely affected the operating conditions of nonproprietary producers. Universities have begun to seek patents and pay royalties, impeding the sharing of information that typified past practice. Businesses

that do not actually rely on asserting patents for their business model have found themselves amassing large patent portfolios at great expense, simply to fend off the threat of suit by others who would try to hold them up. Older documentary films, like *Eyes on the Prize*, have been hidden from public view for years, because of the cost and complexity of clearing the rights to every piece of footage or trademark that happens to have been captured by the camera. New documentaries require substantially greater funding than would have been necessary to pay for their creation, because of the costs of clearing newly expanded rights.

The third economic observation is that the basic technologies of information processing, storage, and communication have made nonproprietary models more attractive and effective than was ever before possible. Ubiquitous low-cost processors, storage media, and networked connectivity have made it practically feasible for individuals, alone and in cooperation with others, to create and exchange information, knowledge, and culture in patterns of social reciprocity, redistribution, and sharing, rather than proprietary, market-based production. The basic material capital requirements of information production are now in the hands of a billion people around the globe who are connected to each other more or less seamlessly. These material conditions have given individuals a new practical freedom of action. If a person or group wishes to start an information-production project for any reason, that group or person need not raise significant funds to acquire the necessary capital. In the past, the necessity to obtain funds constrained information producers to find a market-based model to sustain the investment, or to obtain government funding. The funding requirements, in turn, subordinated the producers either to the demands of markets, in particular to mass-market appeal, or to the agendas of state bureaucracies. The networked information environment has permitted the emergence to much greater significance of the nonmarket sector, the nonprofit sector, and, most radically, of individuals.

The fourth and final economic observation describes and analyzes the rise of peer production. This cluster of phenomena, from free and open-source software to *Wikipedia* and SETI@Home, presents a stark challenge to conventional thinking about the economics of information production. Indeed, it challenges the economic understanding of the relative roles of market-based and nonmarket production more generally. It is important to see these phenomena not as exceptions, quirks, or ephemeral fads, but as indications of a fundamental fact about transactional forms and their relationship to the technological conditions of production. It is a mistake to think that we have only two basic free transactional forms - property-based markets and hierarchically organized firms. We have three, and the third is social sharing and exchange. It is a widespread phenomenon - we live and practice it every day with our household members, coworkers, and neighbors. We coproduce and exchange economic goods and services. But we do not count these in the economic census. Worse, we do not count them in our institutional design. I suggest that the reason social production has been shunted to the peripheries of the advanced economies is that the core economic activities of the economies of steel and coal required large capital investments. These left markets, firms, or state-run enterprises dominant. As the first stage of the information economy emerged, existing information and human creativity - each a "good" with fundamentally different economic characteristics than coal or steel - became important inputs. The organization of production nevertheless followed an industrial model, because information production and exchange itself still required high capital costs - a mechanical printing press, a broadcast station, or later, an IBM mainframe. The current networked stage of the information economy emerged when the barrier of high capital costs was removed. The total capital cost of communication and creation did not necessarily decline. Capital investment, however, became widely distributed in small dollops, owned by individuals connected in a network. We came to a stage where the core economic activities of the most advanced economies - the production and processing of information - could

be achieved by pooling physical capital owned by widely dispersed individuals and groups, who have purchased the capital means for personal, household, and small-business use. Then, human creativity and existing information were left as the main remaining core inputs. Something new and radically different started to happen. People began to apply behaviors they practice in their living rooms or in the elevator - "Here, let me lend you a hand," or "What did you think of last night's speech?" - to production problems that had, throughout the twentieth century, been solved on the model of Ford and General Motors. The rise of peer production is neither mysterious nor fickle when viewed through this lens. It is as rational and efficient given the objectives and material conditions of information production at the turn of the twenty-first century as the assembly line was for the conditions at the turn of the twentieth. The pooling of human creativity and of computation, communication, and storage enables nonmarket motivations and relations to play a much larger role in the production of the information environment than it has been able to for at least decades, perhaps for as long as a century and a half.

A genuine shift in the way we produce the information environment that we occupy as individual agents, as citizens, as culturally embedded creatures, and as social beings goes to the core of our basic liberal commitments. Information and communications are core elements of autonomy and of public political discourse and decision making. Communication is the basic unit of social existence. Culture and knowledge, broadly conceived, form the basic frame of reference through which we come to understand ourselves and others in the world. For any liberal political theory - any theory that begins with a focus on individuals and their freedom to be the authors of their own lives in connection with others - the basic questions of how individuals and communities come to know and evaluate are central to the project of characterizing the normative value of institutional, social, and political systems. Independently, in the context of an information- and innovation-centric economy, the basic components of human development also depend on how we produce information and innovation, and how we disseminate its implementations. The emergence of a substantial role for nonproprietary production offers discrete strategies to improve human development around the globe. Productivity in the information economy can be sustained without the kinds of exclusivity that have made it difficult for knowledge, information, and their beneficial implementations to diffuse beyond the circles of the wealthiest nations and social groups. We can provide a detailed and specific account of why the emergence of nonmarket, nonproprietary production to a more significant role than it had in the industrial information economy could offer improvements in the domains of both freedom and justice, without sacrificing - indeed, while improving - productivity.

From the perspective of individual autonomy, the emergence of the networked information economy offers a series of identifiable improvements in how we perceive the world around us, the extent to which we can affect our perceptions of the world, the range of actions open to us and their possible outcomes, and the range of cooperative enterprises we can seek to enter to pursue our choices. It allows us to do more for and by ourselves. It allows us to form loose associations with others who are interested in a particular outcome they share with us, allowing us to provide and explore many more diverse avenues of learning and speaking than we could achieve by ourselves or in association solely with others who share long-term strong ties. By creating sources of information and communication facilities that no one owns or exclusively controls, the networked information economy removes some of the most basic opportunities for manipulation of those who depend on information and communication by the owners of the basic means of communications and the producers of the core cultural forms. It does not eliminate the possibility that one person will try to act upon another as object. But it removes the structural constraints that make it impossible to communicate at all without being subject to such action by others.

From the perspective of democratic discourse and a participatory republic, the networked information economy offers a genuine reorganization of the public sphere. Except in the very early stages of a small number of today's democracies, modern democracies have largely developed in the context of mass media as the core of their public spheres. A systematic and broad literature has explored the basic limitations of commercial mass media as the core of the public sphere, as well as it advantages. The emergence of a networked public sphere is attenuating, or even solving, the most basic failings of the mass-mediated public sphere. It attenuates the power of the commercial mass-media owners and those who can pay them. It provides an avenue for substantially more diverse and politically mobilized communication than was feasible in a commercial mass media with a small number of speakers and a vast number of passive recipients. The views of many more individuals and communities can be heard. Perhaps most interestingly, the phenomenon of peer production is now finding its way into the public sphere. It is allowing loosely affiliated individuals across the network to fulfill some of the basic and central functions of the mass media. We are seeing the rise of nonmarket, distributed, and collaborative investigative journalism, critical commentary, and platforms for political mobilization and organization. We are seeing the rise of collaborative filtering and accreditation, which allows individuals engaged in public discourse to be their own source of deciding whom to trust and whose words to question.

A common critique of claims that the Internet improves democracy and autonomy is centered on information overload and fragmentation. What we have seen emerging in the networked environment is a combination of self-conscious peer-production efforts and emergent properties of large systems of human beings that have avoided this unhappy fate. We have seen the adoption of a number of practices that have made for a reasonably navigable and coherent information environment without re-creating the mass-media model. There are organized nonmarket projects for producing filtering and accreditation, ranging from the Open Directory Project to mailing lists to like-minded people, like MoveOn.org. There is a widespread cultural practice of mutual pointing and linking; a culture of "Here, see for yourself, I think this is interesting." The basic model of observing the judgments of others as to what is interesting and valuable, coupled with exercising one's own judgment about who shares one's interests and whose judgment seems to be sound has created a pattern of linking and usage of the Web and the Internet that is substantially more ordered than a cacophonous free-for-all, and less hierarchically organized and controlled by few than was the mass-media environment. It turns out that we are not intellectual lemmings. Given freedom to participate in making our own information environment, we neither descend into Babel, nor do we replicate the hierarchies of the mass-mediated public spheres to avoid it.

The concepts of culture and society occupy more tenuous positions in liberal theory than autonomy and democracy. As a consequence, mapping the effects of the changes in information production and exchange on these domains as aspects of liberal societies is more complex. As to culture, the minimum that we can say is that the networked information environment is rendering culture more transparent. We all "occupy" culture; our perceptions, views, and structures of comprehension are all always embedded in culture. And yet there are degrees to which this fact can be rendered more or less opaque to us as inhabitants of a culture. In the networked information environment, as individuals and groups use their newfound autonomy to engage in personal and collective expression through existing cultural forms, these forms become more transparent - both through practice and through critical examination. The mass-media television culture encouraged passive consumption of polished, finished goods. The emergence of what might be thought of as a newly invigorated folk culture - created by and among individuals and groups, rather than by professionals for passive consumption - provides both a wider set of cultural forms and practices and a better-educated or better-practiced community of "readers" of culture. From the perspective of a

liberal theory unwilling simply to ignore the fact that culture structures meaning, personal values, and political conceptions, the emergence of a more transparent and participatory cultural production system is a clear improvement over the commercial, professional mass culture of the twentieth century. In the domain of social relations, the degree of autonomy and the loose associations made possible by the Internet, which play such an important role in the gains for autonomy, democracy, and a critical culture, have raised substantial concerns about how the networked environment will contribute to a further erosion of community and solidarity. As with the Babel objection, however, it appears that we are not using the Internet further to fragment our social lives. The Internet is beginning to replace twentieth-century remote media - television and telephone. The new patterns of use that we are observing as a result of this partial displacement suggest that much of network use focuses on enhancing and deepening existing real-world relations, as well as adding new online relations. Some of the time that used to be devoted to passive reception of standardized finished goods through a television is now reoriented toward communicating and making together with others, in both tightly and loosely knit social relations. Moreover, the basic experience of treating others, including strangers, as potential partners in cooperation contributes to a thickening of the sense of possible social bonds beyond merely co-consumers of standardized products. Peer production can provide a new domain of reasonably thick connection with remote others.

The same capabilities to make information and knowledge, to innovate, and to communicate that lie at the core of the gains in freedom in liberal societies also underlie the primary advances I suggest are possible in terms of justice and human development. From the perspective of a liberal conception of justice, the possibility that more of the basic requirements of human welfare and the capabilities necessary to be a productive, self-reliant individual are available outside of the market insulates access to these basic requirements and capabilities from the happenstance of wealth distribution. From a more substantive perspective, information and innovation are central components of all aspects of a rich meaning of human development. Information and innovation are central to human health - in the production and use of both food and medicines. They are central to human learning and the development of the knowledge any individual needs to make life richer. And they are, and have for more than fifty years been known to be, central to growth of material welfare. Along all three of these dimensions, the emergence of a substantial sector of nonmarket production that is not based on exclusivity and does not require exclusion to feed its own engine contributes to global human development. The same economic characteristics that make exclusive rights in information a tool that imposes barriers to access in advanced economies make these rights a form of tax on technological latecomers. What most poor and middle-income countries lack is not human creativity, but access to the basic tools of innovation. The cost of the material requirements of innovation and information production is declining rapidly in many domains, as more can be done with ever-cheaper computers and communications systems. But exclusive rights in existing innovation tools and information resources remain a significant barrier to innovation, education, and the use of information-embedded tools and goods in low- and middle-income countries. As new strategies for the production of information and knowledge are making their outputs available freely for use and continuing innovation by everyone everywhere, the networked information economy can begin to contribute significantly to improvements in human development. We already see free software and free and open Internet standards playing that role in information technology sectors. We are beginning to see it take form in academic publishing, raw information, and educational materials, like multilingual encyclopedias, around the globe. More tentatively, we are beginning to see open commons-based innovation models and peer production emerge in areas of agricultural research and bioagricultural innovation, as well as, even more tentatively, in the area of biomedical research. These are still very early examples of what can be

produced by the networked information economy, and how it can contribute, even if only to a limited extent, to the capacity of people around the globe to live a long and healthy, well-educated, and materially adequate life.

If the networked information economy is indeed a significant inflection point for modern societies along all these dimensions, it is so because it upsets the dominance of proprietary, market-based production in the sphere of the production of knowledge, information, and culture. This upset is hardly uncontroversial. It will likely result in significant redistribution of wealth, and no less importantly, power, from previously dominant firms and business models to a mixture of individuals and social groups on the one hand, and on the other hand businesses that reshape their business models to take advantage of, and build tools an platforms for, the newly productive social relations. As a practical matter, the major economic and social changes described here are not deterministically preordained by the internal logic of technological progress. What we see instead is that the happenstance of the fabrication technology of computation, in particular, as well as storage and communications, has created technological conditions conducive to a significant realignment of our information production and exchange system. The actual structure of the markets, technologies, and social practices that have been destabilized by the introduction of computer-communications networks is now the subject of a large-scale and diffuse institutional battle.

We are seeing significant battles over the organization and legal capabilities of the physical components of the digitally networked environment. Will all broadband infrastructures be privately owned? If so, how wide a margin of control will owners have to prefer some messages over others? Will we, to the contrary, permit open wireless networks to emerge as an infrastructure of first and last resort, owned by its users and exclusively controlled by no one? The drives to greater private ownership in wired infrastructure, and the push by Hollywood and the recording industry to require digital devices mechanically to comply with exclusivity-respecting standards are driving the technical and organizational design toward a closed environment that would be more conducive to proprietary strategies. Open wireless networks and the present business model of the large and successful device companies - particularly, personal computers - to use open standards push in the opposite direction. End-user equipment companies are mostly focused on making their products as valuable as possible to their users, and are therefore oriented toward offering general-purpose platforms that can be deployed by their owners as they choose. These then become equally available for market-oriented as for social behaviors, for proprietary consumption as for productive sharing.

At the logical layer, the ethic of open standards in the technical community, the emergence of the free software movement and its apolitical cousin, open-source development practices, on the one hand, and the antiauthoritarian drives behind encryption hacking and some of the peer-to-peer technologies, on the other hand, are pushing toward an open logical layer available for all to use. The efforts of the content industries to make the Internet manageable - most visibly, the DMCA and the continued dominance of Microsoft over the desktop, and the willingness of courts and legislatures to try to stamp out copyright-defeating technologies even when these obviously have significant benefits to users who have no interest in copying the latest song in order not to pay for the CD - are the primary sources of institutional constraint on the freedom to use the logical resources necessary to communicate in the network.

At the content layer - the universe of existing information, knowledge, and culture - we are observing a fairly systematic trend in law, but a growing countertrend in society. In law, we see a

continual tightening of the control that the owners of exclusive rights are given. Copyrights are longer, apply to more uses, and are interpreted as reaching into every corner of valuable use. Trademarks are stronger and more aggressive. Patents have expanded to new domains and are given greater leeway. All these changes are skewing the institutional ecology in favor of business models and production practices that are based on exclusive proprietary claims; they are lobbied for by firms that collect large rents if these laws are expanded, followed, and enforced. Social trends in the past few years, however, are pushing in the opposite direction. These are precisely the trends of networked information economy, of nonmarket production, of an increased ethic of sharing, and an increased ambition to participate in communities of practice that produce vast quantities of information, knowledge, and culture for free use, sharing, and follow-on creation by others.

The political and judicial pressures to form an institutional ecology that is decidedly tilted in favor of proprietary business models are running head-on into the emerging social practices described throughout this book. To flourish, a networked information economy rich in social production practices requires a core common infrastructure, a set of resources necessary for information production and exchange that are open for all to use. This requires physical, logical, and content resources from which to make new statements, encode them for communication, and then render and receive them. At present, these resources are available through a mixture of legal and illegal, planned and unplanned sources. Some aspects come from the happenstance of the trajectories of very different industries that have operated under very different regulatory frameworks: telecommunications, personal computers, software, Internet connectivity, public- and private-sector information, and cultural publication. Some come from more or less widespread adoption of practices of questionable legality or outright illegality. Peer-to-peer file sharing includes many instances of outright illegality practiced by tens of millions of Internet users. But simple uses of quotations, clips, and mix-and-match creative practices that may, or, increasingly, may not, fall into the narrowing category of fair use are also priming the pump of nonmarket production. At the same time, we are seeing an ever-more self-conscious adoption of commons-based practices as a modality of information production and exchange. Free software, Creative Commons, the Public Library of Science, the new guidelines of the National Institutes of Health (NIH) on free publication of papers, new open archiving practices, librarian movements, and many other communities of practice are developing what was a contingent fact into a self-conscious social movement. As the domain of existing information and culture comes to be occupied by information and knowledge produced within these free sharing movements and licensed on the model of open-licensing techniques, the problem of the conflict with the proprietary domain will recede. Twentieth-century materials will continue to be a point of friction, but a sufficient quotient of twenty-first-century materials seem now to be increasingly available from sources that are happy to share them with future users and creators. If this social-cultural trend continues over time, access to content resources will present an ever-lower barrier to nonmarket production.

The relationship of institutional ecology to social practice is a complex one. It is hard to predict at this point whether a successful sustained effort on the part of the industrial information economy producers will succeed in flipping even more of the institutional toggles in favor of proprietary production. There is already a more significant social movement than existed in the 1990s in the United States, in Europe, and around the world that is resisting current efforts to further enclose the information environment. This social movement is getting support from large and wealthy industrial players who have reoriented their business model to become the platforms, toolmakers, and service providers for and alongside the emerging nonmarket sector. IBM, Hewlett Packard, and Cisco, for example, might stand shoulder to shoulder with a nongovernment organization (NGO) like Public Knowledge in an effort to block legislation that would require personal computers to

comply with standards set by Hollywood for copy protection. When Hollywood sued Grokster, the file-sharing company, and asked the Supreme Court to expand contributory liability of the makers of technologies that are used to infringe copyrights, it found itself arrayed against amicus briefs filed by Intel, the Consumer Electronics Association, and Verizon, SBC, AT&T, MCI, and Sun Microsystems, alongside briefs from the Free Software Foundation, and the Consumer Federation of America, Consumers Union, and Public Knowledge.

Even if laws that favor enclosure do pass in one, or even many jurisdictions, it is not entirely clear that law can unilaterally turn back a trend that combines powerful technological, social, and economic drivers. We have seen even in the area of peer-to-peer networks, where the arguments of the incumbents seemed the most morally compelling and where their legal successes have been the most complete, that stemming the tide of change is difficult - perhaps impossible. Bits are a part of a flow in the networked information environment, and trying to legislate that fact away in order to preserve a business model that sells particular collections of bits as discrete, finished goods may simply prove to be impossible. Nonetheless, legal constraints significantly shape the parameters of what companies and individuals decide to market and use. It is not hard to imagine that, were Napster seen as legal, it would have by now encompassed a much larger portion of the population of Internet users than the number of users who actually now use file-sharing networks. Whether the same moderate levels of success in shaping behavior can be replicated in areas where the claims of the incumbents are much more tenuous, as a matter of both policy and moral claims - such as in the legal protection of anticircumvention devices or the contraction of fair use - is an even harder question. The object of a discussion of the institutional ecology of the networked environment is, in any event, not prognostication. It is to provide a moral framework within which to understand the many and diverse policy battles we have seen over the past decade, and which undoubtedly will continue into the coming decade, that I have written this book.

We are in the midst of a quite basic transformation in how we perceive the world around us, and how we act, alone and in concert with others, to shape our own understanding of the world we occupy and that of others with whom we share it. Patterns of social practice, long suppressed as economic activities in the context of industrial economy, have now emerged to greater importance than they have had in a century and a half. With them, they bring the possibility of genuine gains in the very core of liberal commitments, in both advanced economies and around the globe. The rise of commons-based information production, of individuals and loose associations producing information in nonproprietary forms, presents a genuine discontinuity from the industrial information economy of the twentieth century. It brings with it great promise, and great uncertainty. We have early intimations as to how market-based enterprises can adjust to make room for this newly emerging phenomenon - IBM's adoption of open source, Second Life's adoption of user-created immersive entertainment, or Open Source Technology Group's development of a platform for Slashdot. We also have very clear examples of businesses that have decided to fight the new changes by using every trick in the book, and some, like injecting corrupt files into peer-to-peer networks, that are decidedly not in the book. Law and regulation form one important domain in which these battles over the shape of our emerging information production system are fought. As we observe these battles; as we participate in them as individuals choosing how to behave and what to believe, as citizens, lobbyists, lawyers, or activists; as we act out these legal battles as legislators, judges, or treaty negotiators, it is important that we understand the normative stakes of what we are doing.

We have an opportunity to change the way we create and exchange information, knowledge, and culture. By doing so, we can make the twenty-first century one that offers individuals greater autonomy, political communities greater democracy, and societies greater opportunities for cultural self-reflection and human connection. We can remove some of the transactional barriers to material opportunity, and improve the state of human development everywhere. Perhaps these changes will be the foundation of a true transformation toward more liberal and egalitarian societies. Perhaps they will merely improve, in well-defined but smaller ways, human life along each of these dimensions. That alone is more than enough to justify an embrace of the networked information economy by anyone who values human welfare, development, and freedom.