distinct field or locus of symbolic generation and differentiation. The distinction of "culture" from the regimentation and reification of science, politics, economics, or militarism is a distinctly modern phenomenon, the result of a process that Max Weber called *Wertausdifferenzierung*, or "value differentiation." Weber claimed that culture under the modern state and capitalist economy tends to foster oppositional poses as much as legitimizing ones. Under the political canopy of the twentieth-century industrial and welfare state, cultural politics was merely an adjunct to questions of resource distribution, but calling for resource distribution in a neoliberal context seems futile and is dismissed as counterproductive. Consequently, in recent years, Benshabib explains, cultural groups have been employing political strategies to assert recognition, rather than redistribution, although there can be redistributive consequences of cultural recognition. In a desperate, divided, Darwinian world economy, cultural recognition can seem as important as life itself. Attempts at forging a global public sphere discount the importance of cultural recognition in favor of procedural equality. Not that there is anything wrong with that; but failing to consider the visceral power of specific cultural claims is likely to exclude and alienate much of the postcolonial world.

With its powerful trends toward localization in search results and thus the customization of knowledge, Google's search functions actually reinforce the interests of the local-culture movements and thus inhibit rather than further the expansion of a genuine global civil society. However, several major aspects of Google's business have influenced the expansion of global civil society in its present form and have offered a glimpse of what a global public sphere might look like: YouTube, Blogger, and Google News. These are some of the main factors in the Googlization of the world. If the development of a global public sphere is a good thing and a goal to be pursued—and despite the obstacles to such a development that I've been analyzing, there are people and forces that would assert that it is—we need to ponder ways in which we can influence the Googlization of the world to achieve that end. One way to do that is to analyze further another major aspect of the Googlization of everything: the Googlization of knowledge.

Those of us who take liberalism and Enlightenment values seriously often quote Sir Francis Bacon's aphorism that "knowledge is power." But, as the historian Stephen Gaukroger argues, this is not a claim about knowledge: it is a claim about power. "Knowledge plays a hitherto unrecognized role in power," Gaukroger writes. "The model is not Plato but Machiavelli." Knowledge, in other words, is an instrument of the powerful. Access to knowledge gives access to that instrument of power, but merely having knowledge or using it does not automatically confer power. The powerful always have the ways and means to use knowledge toward their own ends.

However, expanding access to knowledge brings more people with more and different ends into the space where those ends can be made known, be advocated, and take their place on the agendas of nations and...
transnational movements alike. Indeed, advocates for increased access to knowledge have put that issue itself on the international agenda regarding questions ranging from access to patent medicines to access to proprietary software. The issue of access to knowledge is thus central to the prospects for expanding the public sphere and thereby contesting the claims of the powerful to all the instruments of power.

Much of human knowledge exists in the form of long arrays of text, what we still call books. We are dazzled and distracted by the new methods of transmitting and using this knowledge, but most of the best expressions of deep human thinking still rest on paper, bound with glue, nestled and protected by cloth covers, on the shelves of libraries around the world. How can we simultaneously preserve and extend that knowledge? How can we vet and judge its utility and truth? How can we connect the most people with the best knowledge? Google, of course, offers answers to those questions. It’s up to us to decide whether Google’s answers are good enough.

**SHUFFLING THE PAGES**

In May 2006, the *Wired* magazine contributor Kevin Kelly published in the *New York Times Magazine* his predictive account of flux and change in the book-publishing world. That article outlined what he claimed “will” (not “might” or “could”) happen to the book business and the practices of writing and reading under a new regime fostered by Google’s plan to scan millions of books from university and public libraries and offer searchable texts to Internet users. “So what happens when all the books in the world become a single liquid fabric of inter-connected words and ideas?” Kelly wrote. “First, works on the margins of popularity will find a small audience larger than the nearly-zero audience they usually have now. Second, the universal library will deepen our grasp of history, as every original document in the course of civilization is scanned and cross-linked. Third, the universal library of all books will cultivate a new sense of authority.”

Kelly suggested that the linkages of text to text, book-to-book, page to page, and passage to passage will fill the knowledge gaps that have made certain people winners and others losers. “If you can truly incorporate all texts—past and present, multilingual—on a particular subject,” he wrote, “then you can have a clearer sense of what we as a civilization, as species, do know and don’t know. The white spaces of our collective ignorance are highlighted, while the golden peaks of our knowledge are drawn with completeness. This degree of authority is only rarely achieved in scholarship today, but it will become routine.”

Such heady predictions of technological revolution have become so common, so accepted in our techno-fundamentalist culture, that even when John Updike criticized Kelly’s vision in an essay published a month later in the *New York Times Book Review*, he did not doubt that it would someday come to pass. Updike just lamented the change, musing about how wonderful his old bookstore haunts were for him and everyone else who strolled the streets of New York, Oxford, and Boston in the 1950s. His elitist comments served only to bolster the democratic credentials of Kelly and others who have been asserting that Google’s plan to scan millions of books would spread knowledge to those not as lucky as Updike.

As it turns out, universal access to book knowledge is proving not so easy to accomplish. Kelly’s predictions depend, of course, on the cooperation of one part of the system that he slights in his article: the copyright system. He mentions copyright as a mere nuisance: to acknowledge that a system built by lawyers might defeat one built by engineers would have run counter to his vision. In fact, when he wrote his article, it seemed entirely possible that the current American copyright system would crush Google’s plan to scan the entire collections of dozens of university libraries.

**THE GOOGLIZATION OF BOOKS**

For several years, Kelly’s vision for a universal digital library seemed to be approaching realization through a project known at different times
as Google Print, Google Book Search, and Google Books. The project foundered and then apparently recovered, thanks to the legal settlement that Google reached in October 2008 with the Association of American Publishers and the Authors' Guild. That settlement came after four years of argument over what copyright would look like in a digital age. It dodged the legal and philosophical questions at the heart of the dispute, and it set up a bold new system for book research and distribution that, instead of promoting access to knowledge, raised even more questions: the lack of competition, increased monopolization, and the increasing privatization of the information ecosystem.5

In 2004, Google began scanning and indexing millions of books from more than twenty-five university libraries. This service has been the subject of much hyperbolic speculation. On first learning of Google’s plans, legal scholars such as Lawrence Lessig claimed that they would radically democratize information for the public, not just for academics. Authors such as Cory Doctorow initially applauded Google Books for offering ways to connect interested readers to particular texts and thus prevent small books from getting lost in the mass market. And techno-libertarians such as Kelly celebrated the transformative nature of electronic texts, arguing that Google Books would allow users to connect disparate pieces of information as they saw fit, thus evading the tyranny of the book- cover and library catalog. These were expressions by true-believers in the potential of digital culture—when properly supported by a benevolent force such as Google—to transform, extend, and democratize knowledge. Publishers and authors, meanwhile, took a less rosy view, and two high-profile lawsuits were initiated against the program for copyright infringement.

Google Books has failed to live up to any of the exaggerated claims that its early proponents made for it. Not only has it failed to deliver on its promises, but along the way it has disrupted the copyright system and the economy of publishing. Google had hoped to take the modes and standards of Web copyright practice and apply them to books in the real world, where they do not fit. Once people discovered the contours and details of the settlement proposal engineered by publishers’ lawyers

and Google in the fall of 2008, they saw some big problems. Copyright and cyberlaw professors who had cheered Google’s bold embrace of the principle of fair use of copyrighted material realized that Google had actually designed a system that would give it important competitive advantages, making it too powerful within the economy and culture of books.6 When it was first announced, the Harvard law professor and copyright reform advocate Lawrence Lessig called the settlement “a good deal that could be the basis for something really fantastic.” But after considering all the debates and issues surrounding the settlement and Google’s plans, Lessig concluded that the settlement would not only fail to loosen up American copyright law but might even restrict and commercialize the flow of digital knowledge, and he withdrew his support for the project.8

More significantly, the head of one of the original Google library-partners, Harvard University Libraries, publicly declared that he opposed the project. The historian Robert Darnton had been a professor at Princeton University when Harvard reentered its partnership with Google. Once he became head of the library’s at Harvard, he began to question whether it was in the best interest of the university to contribute to the privatization of knowledge through Google. In February 2009, Darnton published an influential article in the New York Review of Books in which he declared Google’s efforts to control so much of our historical heritage a danger to the future of learning.9

In addition, the governments of France and Germany issued opinions that Google Books would give Google an unfair advantage in the market for out-of-print texts. Authors in China sued Google for infringing their copyrights by scanning their books without permission, prompting a rare apology from the company.10 In September 2009 the U.S. Department of Justice issued an opinion that the Google-Books ‘settlement would violate U.S. antitrust laws unless it were significantly redrawn. Google and the publishers withdrew the settlement to revise it and resubmit it for a hearing before a judge, which occurred in February 2010. Even the revised version failed to allay the Justice Department’s concerns that because the settlement would facilitate the sale of digital copies of these books through Google, the proposed system would
effectively make Google the sole vendor of most of the books published in the twentieth century.

As I finished the editing and updating of this book in August 2010, Judge Denny Chin still had under consideration the approval of the class-action settlement of the case between the publishers (and some authors) and Google. If Judge Chin approves the settlement, Google will be in a position to offer for sale millions of digital files of out-of-print books published in the twentieth century. In addition, Google would offer access to many millions of books that were never protected by copyright or whose copyright has expired. The settlement would facilitate a remarkable change in the relationship among books, readers, publishers, authors, libraries, and Google. Access to so many great works would be greater than anyone imagined just ten years ago. But American libraries would be commercialized, essentially hosting Google vending machines on their premises. Publishers and authors might make a little more money than they did before. Occasionally, a long-lost work might emerge to be a surprise best seller. But Google would assert itself as the mediator of the accessibility and affordability for this vast collection. No other firm could realistically hope to mount a competing service. Readers would seamlessly shift between the safe, anonymous, republican space of the public library and the commercialized environment of Google without a warning that their reading and browsing habits would be tracked. And, perhaps most costly, we might never be willing to design and fund high-quality, durable, publicly run, noncommercial services with the mission of spreading knowledge rather than selling books or placing advertisements.

If, on the other hand, Judge Chin rejects the settlement and puts the copyright lawsuits between authors and Google and publishers and Google back into court, then the entire project is doomed. Google, after trying to settle, could not convincingly and in good faith mount a defense against the accusations of copyright infringement. And publishers would have little incentive to renegotiate and settle on lesser terms than Google offered in the first place.

The most troubling aspect of the settlement goes beyond any of the legalities and specifics, and it has nothing to do with how we will find and experience books and knowledge in the next few decades. The scanning project that has been bringing the collected works of dozens of libraries into easy-to-use forms and the changes in policy and practice that would flow from the settlement are monumental in scope. The Google Books project is one of the most revolutionary information policy changes in a century or more. If approved, it would alter how we think about copyright, culture, books, history, access, and libraries. Yet the public has had no say in how it will be constructed and run. No public policymaking body oversaw its creation. No legislature considered the notion of creating what amounts to a compulsory-license system (through which the copyright holder is never asked beforehand if she agrees to the copying; instead the copier may assume the right to copy) to allow a company to scan copyrighted books by the millions.

The Google Books plan is a perfect example of public failure. The great national, public, and university libraries of the world never garnered the funds or the political will and vision needed to create a universal, digital delivery service like Google envisions. The public institutions failed to see and thus satisfy a desire—perhaps a need—for such a service. Google stepped in and declared that it could offer something close to universal access for no cost to the public. The catch, of course, was that it would have to be done on Google's terms, with no attention paid to long-term preservation needs or quality standards. Essentially, the Google Books project is a radical change in information policy executed by a class-action settlement. If it goes into effect, private law will determine public policy.

How did such a seemingly benign project balloon into the most controversial and risky effort Google ever initiated? Google's leaders may not have realized it at the time, but many people were growing wary of its increasing power over the global information ecosystem, and the details of its proposal to digitize millions of copyrighted books touched on some very controversial issues: copyright, competition, privacy, the privatization of public libraries, and the future of books themselves. Hanging over the promise of access to knowledge offered by Google Books is the specter of its opposite—restrictions on
open access to books, their contents, and the power that such access might help provide.

BOOK SEARCH, COPYRIGHT, AND THE FREE RIDE

In an op-ed piece in the *New York Times* in October 2009, Google’s cofounder Sergey Brin defended the Books program and declared that Google was interested in digitizing books because such a project fit the idealistic mission of the firm. “Because books are such an important part of the world’s collective knowledge and cultural heritage, Larry Page, the co-founder of Google, first proposed that we digitize all books a decade ago, when we were a fledgling startup,” Brin wrote. “At the time, it was viewed as so ambitious and challenging a project that we were unable to attract anyone to work on it. But five years later, in 2004, Google Books (then called Google Print) was born, allowing users to search hundreds of thousands of books. Today, they number over 10 million and counting.”

Brin lamented that the project had attracted lawsuits from publishers and a few wealthy authors, but he wrote that the settlement was in the best interest of everyone—including the public. “While we [Google and the publishers that sued Google] have had disagreements, we have a common goal—to unlock the wisdom held in the enormous number of out-of-print books, while fairly compensating the rights holders,” Brin wrote. “As a result, we were able to work together to devise a settlement that accomplishes our shared vision. While this settlement is a win-win for authors, publishers and Google, the real winners are the readers who will now have access to a greatly expanded world of books.” Brin also presented the project as a way to preserve the knowledge of centuries from the perils of physical harm, such as fire and flood.

Oddly, Brin wrote this piece without conceding that the quality of Google’s document scans was too poor to serve the aims of preservation. In many cases, human hands obscure the text in Google Books images, and pages are missing or blurry. The quality of Google’s scanned images is far below that of library-run digital preservation efforts. More interesting, though, is Brin’s failure to mention the fact that Google Books is a revenue-generating project for the company. It is not a public service. And Google is not a library.

Google may have been the biggest and most controversial player in the effort to digitize books, but it was hardly the first. The saga of digital books offered on the Web is tortured and long. Back in the early 1990s, several groups of tech-savvy bibliophiles began posting plain-text versions of classic works that have entered the public domain. Among the best-known of these services are Project Gutenberg and Eldritch Press. As public participation in the Web grew through the 1990s and more people expressed a desire to read substantial texts on mobile devices and laptops, these services grew in importance, but they suffered from several limitations. First, public-domain works were simply not in high demand in electronic form; second, the plain-text format made files portable and searchable but were often unattractive to read. Firms such as Random House had experimented with electronic versions of their popular books as early as 1994, but the early reading devices on which these works were offered either did not work well or were too expensive—or both. Meanwhile, as Amazon.com established itself as the leading retail outlet for printed books on the Web, it began offering a “Look Inside” feature, presenting electronic glimpses of tables of contents and samples of text to assist customers. But searching, researching, and acquiring access to the full texts of electronic works on Amazon remained impossible. Amazon was offering digital images of text purely as a sales technique, not as a public good.

Before embarking publicly on the massive scanning of library collections without permission, Google launched what it called its “partner program.” Inspired by Amazon’s success in book sales online, beginning in early 2003 Google began negotiating with commercial and academic publishers to secure digital rights for what was initially called Google Print. The terms of access to the millions of book-page images Google collected depended on the particular wishes of the publishers. Some titles offered nearly full-text access. Others offered only excerpts. In general, users could view only a few pages of a book at a time, and they could not copy, print, or download the images. The margins of
the pages Google offered contained links to sources where a user could purchase the books, as well as bibliographic information and links to the publishers’ sites.

Then, in December 2004, Google shocked publishers and the public by announcing its plans to digitize millions of bound books from five major English-language libraries. The libraries’ initial contributions in 2004 were as follows.

- Harvard University libraries: 40,000 public-domain books during the pilot phase of the project, with the possibility of extension. The library has more than 15 million volumes.
- Stanford University libraries: hundreds of thousands of public-domain books, with the possibility of extending the program to cover the entire collection of 7.6 million books.
- University of Michigan at Ann Arbor: all 7.8 million books in the collection, even those under copyright.
- Oxford University: all books published before 1900. The library holds a total of 6.5 million books.
- The New York Public Library: between ten thousand and one hundred thousand public-domain volumes as part of the pilot project. The library holds 20 million volumes.

Over the next several months, dozens of other university libraries joined the project. These included the University of Wisconsin, the University of Virginia, and, most significant, the University of California system, which planned to scan more than 2.5 million books at a rate of three thousand volumes per day. In total, Google planned to add more than 7.7 million library volumes to its electronic index at an estimated cost of $10 per book. Most of the more recent library partners offered Google their special collections, as well as access to select volumes not included in the Michigan collection. In return for access to the books, Google promised to provide the libraries with electronic copies of the works they contributed to the project. However, in some of the more recent partnership agreements, Google held back from scanning certain works while they determined the status of lawsuits and the utility of the files for the Google project.

Under the original, unauthorized library-scanning project—which is distinct from the “partner” project authorized by publishers—search results and the user experience depended on the copyright status of the book. The company announced that for works published before 1923 (and thus mostly in the public domain in the United States), users would have access to the entire text. For works published since 1923 (and thus potentially still under copyright protection), the user would see the bibliographic information, as well as a few text excerpts (“snippets”) containing the term that the user had typed into the search box. Google claims that viewing the displayed results of copyrighted works is comparable to the “experience of flipping through a book in a bookstore.” As with the authorized “partner” content, Google provided links to allow users to buy books from numerous vendors, as well as targeted advertisements that depended on the nature of the book and possibly also the inferred interests of the searcher.

When major commercial publishers learned of this clandestine library-scanning project, their initial reactions were panicked, alarmist, and largely unwarranted. They expressed concerns that the Google project would threaten book sales and risk hacking and the widespread pirating of texts. Gradually it became clear that Google’s library project posed no threat to publishers’ core markets and projects. If anything, the project could have been a marketing boon: if the searches yielded books that met users’ needs, they were likely to purchase at least some of those works. Since then, it has become clear that publishers were most offended by the prospect of a wealthy corporation free-riding on their content to offer a commercial and potentially lucrative service without any regard to compensation or quality control. The publishers wanted a piece of the revenue—and some control over the manner of display and search results.

Copyright, which has traditionally protected the rights of authors and publishers to control the copying and distribution of their works, has rarely been used to govern ancillary markets for goods that enhance the value or utility of the copyrighted works. As the author and activist Cory Doctorow has pointed out, booksellers have never tried to extract licensing revenues from bookcase makers, bookmark makers, or eyeglass
producers. By analogy, a searchable, online, full-text index, similar to what Google had originally planned to offer, is a supplement to a book (and to book culture), not a substitute for it. However, creating such an index requires that Google make digital copies of the complete physical books, thus violating the fundamental provisions of the copyright act. So although the publishers’ complaints were hyperbolic, they might indeed have had the law on their side.

The conflict over Google’s bold initial library project raised questions that get to the heart of copyright. If Google and the publishers had not settled the publishers’ lawsuit and instead had pursued these questions through the courts, not only would we have witnessed some fascinating and important public discussions of the role, scope, purpose, and design of the copyright system, but we might also have seen some significant—perhaps radical—changes to it. Had the publishers prevailed, Google’s core mission and the openness of the Web would have been threatened—as would all the revenue that Google has accumulated from advertisements and capital markets. Had Google prevailed, we would have seen a serious shift of power in information markets, from analog firms devoted to creating and taxing scarcity by pricing and selling books to digital firms (like Google) designed to manage the abundance of information by collecting information about its users and selling advertising access to them. More directly, the peculiarly American notion of fair use of copyrighted material—and perhaps even the copying of entire works for clearly commercial purposes—would have been expanded and solidified. It would have represented a shift far beyond what Congress had ever imagined when it codified fair use in 1976, when the advent of the photocopier supposedly threatened commercial publishing with extinction. Fair use, in short, is a defense one may use in U.S. courts when accused of copyright infringement. One may argue that the use of the original material is small enough that it does not threaten the market for the original, or that the use is clearly in the service of journalism, criticism, research, or education. Nothing about fair use is clear and simple. Courts are supposed to consider fair-use arguments on a case-by-case basis. And there are very few certainties about how well such a defense would work. Fair use was developed to allow individuals to avoid going through the time and expense of securing permission to use copyrighted material when the public clearly benefits from the unauthorized use of it. Google, however, was making an argument about the general permissibility of its massive copying. Had Google pursued that argument and prevailed, fair use would have been a significantly stronger users’ right than it had been designed to be. Had Google lost in court, fair use on the Web might have been severely curtailed.

Instead, the settlement of October 2008 avoided any such revolutionary change to the law, yet it generated a new, hybrid set of rules to govern our information ecosystem and set the terms of access to our cultural heritage. Here are some of the major elements of the settlement:

- The members of the Authors Guild and the Association of American Publishers agreed to cease pursuing damages for copyright infringement.
- Google offered to pay $125 million to publishers to settle the case.
- Google undertook plans to establish and run a not-for-profit rights registry to allow rights holders to claim or establish control over out-of-print works. This registry was intended to serve as a database through which scholars and publishers could find rights holders in order to clear rights. Because no such registry existed previously, this provision had the potential to be a boon to research and publishing. In addition, it could help rights holders accruing royalties (meager though they might be) by exploiting a market that has never worked efficiently or effectively: that for reprints or selections from out-of-print works. Google was undertaking to do what the U.S. Copyright Office should have done years ago.
- Google agreed to offer (with strict controls on the ability to print and share) full-text copies of certain out-of-print books for sale as downloads.
- Google undertook to offer much better access to many out-of-print works still under copyright. Before the settlement, Google offered largely useless excerpts of these texts. The settlement provided for much richer and broader access.
- Google agreed to provide designated computer terminals in U.S. libraries that would offer free full-text, online viewing of millions
of out-of-print books. Google would forbid printing from these terminals, but users would be able to purchase electronic copies of the books from these terminals.

Compared with the severe limitations on user access to most twentieth-century works under the original model for Google Books, this new model promised to improve the service substantially. In addition, the settlement aimed to avoid the threat of the great copyright meltdown outlined above. Clearly both sides saw real risks in forcing a courtroom showdown. However, back when Google introduced the library-scanning project as part of the Books program, many copyright critics celebrated the fact that a big, rich, powerful company was taking a stand to strengthen fair use. That never happened. Fair use in the digital world is just as murky and unpredictable as it was the day before the settlement. But what about the problems and pitfalls of this settlement? Critics of Google Books still have serious concerns about it. Immediately after the announcement of the settlement, I asked Google’s legal department the following questions:

Isn’t this a tremendous antitrust problem? Google has essentially set up a huge compulsory licensing system without the legislation that usually makes such systems work. In addition, this proposed system excludes many publishers, such as university presses, and authors who are not members of the Authors’ Guild. More important, this system excludes the other major search engines and the one competitor Google has in the digital book race, the Open Content Alliance. Don’t these parties now have a very strong claim for an antitrust action?

The Google legal team did not believe that this agreement was structured in such a way as to exclude others from developing a competing service. The agreements with and about publishers, libraries, and the registry were all nonexclusive, as is typical of Google’s approach to competition in the Web business. The registry would be started with Google funds, but it would be an independent, nonprofit entity able to deal with the Open Content Alliance and other services without restriction from Google. Generally, Google’s lawyers did not see this service as presenting a typical antitrust problem. There are so many segments to the book market in the world, including real bookstores, online stores, such as Amazon.com, and used-book outlets, they claimed, that no single entity or sector can set prices for books (even out-of-print books) effectively. There are always competing sources, including libraries themselves?

But isn’t this a potential privacy nightmare for libraries? I asked. Will Google compile personally identifiable information from users of its free terminals (for example, by requiring them to log in to Google Docs or some other service)? Will Google collect search and usage data from these library terminals to “improve” searches? Will such data be open for study by publishers or media scholars? How long would Google retain such data if it were compiled?

The response from Google’s lawyers, in November 2008, exhibited a willingness to examine this potential problem. They indicated that much about the design of the program was yet to be determined. Google had not agreed to share personal information with publishers, but the company might share aggregate data collected through the service. And although Google had not yet designed the system, the legal department predicted that users would not have to log in to Google to use the public terminals. The legal department assured me that the company would “build in privacy protections” with the guidance and assistance of the library partners.

SELLING OUT LIBRARIES AND CORPORATE WELFARE

The main criticism of Google Books has always concerned the actions of the university libraries that have participated in this program, rather than Google itself or the effects of the program on libraries in general. The advantages to libraries of the settlement are twofold. First, they might face much less legal risk by permitting Google to scan books in their collections that are still protected by copyright (although future lawsuits by authors and publishers who live outside the United States, Canada, Australia, and the United Kingdom—the only countries covered by the settlement—remain a risk). Second, because Google has pledged