

This work is distributed as a Discussion Paper by the
STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 07-36

**Restricting Access to Books on the Internet:
Some Unanticipated Effects of U.S. Copyright Legislation**

By
Paul A. David,
Stanford University and Oxford Internet Institute
and
Jared Rubin
California State University, Fullerton

March 2008

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

The Stanford Institute for Economic Policy Research at Stanford University supports research bearing on economic and public policy issues. The SIEPR Discussion Paper Series reports on research and policy analysis conducted by researchers affiliated with the Institute. Working papers in this series reflect the views of the authors and not necessarily those of the Stanford Institute for Economic Policy Research or Stanford University.

**RESTRICTING ACCESS TO BOOKS ON THE INTERNET:
Some Unanticipated Effects of U.S. Copyright Legislation¹**

Paul A. David²

*Stanford University & The Oxford Internet Institute
& UNU-MERIT / Collaborative Creativity Group (Maastricht, NL)*

Jared Rubin

California State University, Fullerton

PUBLICATION NOTE

The article that was published under the same title in the *Review of Economic Research on Copyright Issues*, 2008, vol. 5(1), pp. 23-53 contains (authors' errata on pp. 40, 46, 47 that are corrected in this text.

ABSTRACT

One manifestation of the trend towards the strengthening of copyright protection that has been noticeable during the past two decades is the secular extension of the potential duration during which access to copyrightable materials remains legally restricted. Those restrictions carry clear implications for the current and prospective costs to readers seeking “on-line” availability of the affected content in digital form, via the Internet. This paper undertakes to quantify one aspect of these developments by providing readily understandable measures of the restrictive consequences of the successive modifications that were made in U.S. copyright laws during the second half of the twentieth century. Specifically, we present estimates of the past, present and future number of copyrighted books belonging to different publication-date “cohorts” whose entry into the public domain (and consequent accessibility in scanned on-line form) will thereby have been postponed. In some instances these deferrals of access due to legislative extensions of the duration of copyright protection are found to reach surprisingly far into the future, and to arise from the effects of interactions among the successive changes in the law that generally have gone unnoticed.

¹ The research and writing of this paper has been undertaken as part of the Project on “The Co-evolution of Information Technology and Copyright Law” in the Knowledge Networks and Institutions for Innovation Program (KNIIP) of the Stanford Institute for Economic Policy Research (SIEPR). This work drew encouragement at an early stage from Lawrence Lessig’s comments on the potential value of a systematic quantitative assessment of the consequences of successive legislative extensions of the duration of copyright protection. Financial support received under a Rockefeller Foundation grant awarded to Stanford University, and the sustained interest taken in this somewhat usual line of “culturally relevant” research by Joan Shigekawa, the Foundation’s Associate Director, are most gratefully acknowledged. We are grateful too for the helpful comments received from an anonymous referee.

² Contact Author: Paul A. David, Department of Economics, Stanford University, Stanford CA. 94305-6072; Email: pad@stanford.edu

1. INTRODUCTION

It is today a commonplace observation that the technical possibilities of accessing enormous global resources of cultural and scientific information have been and are continuing to be greatly augmented by spectacular, on-going advances in digital information and computer-mediated telecommunication technologies. At the same time, however, the concurrent evolution of the provisions of copyright law – not only in the U.S., but internationally – has evinced the seemingly inexorable tendency to expand the sphere of protection for intellectual property until the latter eventually might cover every possible use of information-goods that possibly could yield private economic benefits. A growing number of legal scholars and economic commentators lately have begun to point out that the other side of this trend in copyright law has been a drift towards restricting the potential social benefits that otherwise might flow from leaving such information-goods in the public domain, or providing legal means of facilitating their shared use.

One simple manifestation of the latter trend is found in the secular extension of the potential duration during which access to copyrightable materials remains restricted by the licensing terms that copyright owners can impose upon would-be users of their protected works. This paper undertakes to quantify this effect, providing readily understandable measures of the restrictive consequences of successive modifications of U.S. copyright statutes. Specifically, it presents estimates of the volume and time-distribution of published books whose entry into the public domain has thereby been postponed, quantitatively assessing the consequences of twentieth century modifications of the U.S. statutes on the public's access to material that was published in this country under copyright. These results provide a set of quantitative benchmarks against which it is both possible and informative to assess the extent of legislative awareness (as exhibited by contemporary testimony in Congressional committee hearings and debates, for example) regarding the calculable consequences that could be expected to flow from prospective changes in copyright statutes.

Whatever salutary effect the availability of a convenient means of providing such benchmarks may turn out to have on future standards of legislative and judicial deliberations in this area, its retrospective application contributes to a better understanding of the process that has brought the copyright regime to its present condition. At the very least, this paper lends greater concreteness to analyses of the trade-offs between public and private benefits, and between societal and individual costs that appear to have been at the heart of the political economy of recurring legislative reforms of the U.S. copyright efforts during the past half-century.

By design, our main quantitative findings regarding the cumulative effect of U.S. copyright legislation in delaying full access to a growing number of publications are quite transparent, and therefore amenable to accurate description in non-technical terms. On their face, the graphic presentation of our findings carry an important, broad message: the worthy dreams of technologists like Brin and Page at Google of creating easy and ubiquitous access to the contents of a global virtual library cannot be realized through digital information technology engineering ingenuity alone. This imparts a striking degree of concreteness to the contention that the detailed operations of legal and other elements of the institutional infrastructure need to be re-examined in view of the potent and persisting influences they exercise over the present and future access to information, its re-use and its re-combination in generating new knowledge and cultural goods.

The paper is organized as follows: Section 2 reviews the implementation of the plan to scan existing library holdings to provide an online library with digitized access to the contents of the world's printed books, starting with Google's "Libraries Project" and following the developments that ensued from the reactions of copyright holders and libraries. The course of changes in the duration of copyright protections in the U.S. affected by legislative acts during the twentieth century is examined in section 3, with particular attention paid to the cumulative effects of interactions among the changes made in the latter half of the century. Section 4 presents the methodology for quantifying the way in which the statutory provisions affected the timing of the return to public domain status of successive cohorts of books copyrighted (and whose copyrights were renewed) in the U.S. Section 5 discusses the findings with regard to the differential impact of each of the legislative acts by estimating under alternative counterfactual assumptions the volume of books that would be withheld from the public domain at successive dates – projecting these effects forward until 2027. We conclude the paper in Section 6 with a discussion of the findings' broader implications in regard to the interactions between technological changes and intellectual property law in the areas affected by copyrights.

2. BOOK-SCANNING – FROM GOOGLE'S "LIBRARIES PROJECT" TO THE OPEN CONTENT ALLIANCE

Much fanfare accompanied the announcement in December of 2004 that Google, the operator of the world's most popular Internet search service, had concluded an agreement with four leading research libraries in the U.S. and the Bodleian Library at Oxford to begin converting their holdings into digital files that would be "freely searchable over the Web". The *New York Times* story led by presenting the grand vision in suitably cautionary terms:

It may be only a step on a long road toward the long-predicted global virtual library. But the collaboration of Google and research institutions that also include Harvard, the University of Michigan, Stanford and the New York Public Library is a major stride in an ambitious Internet effort by various parties. The goal is to expand the Web beyond its current valuable, if eclectic, body of material and create a digital card catalog and searchable library for the world's books, scholarly papers and special collections (Markoff and Wyatt, 2004).

Stanford University's Librarian, Michael A. Keller, was quoted in this story as seeing the future more clearly, and in the more roseate hues that initially surrounded public commentary on the project: "[w]ithin two decades, most of the world's knowledge will be digitized and available, one hopes for free reading on the Internet, just as there is free reading in libraries today." Such hopes are now seen to have been unrealistically optimistic.

Although details of the agreements between Google and the universities have not been completely disclosed, and there are aspects of these contracts that are expected to remain private, a number of limitations of this undertaking already were evident in the fine print that appeared further down in the pages of the *New York Times* on December 14th.³ Not all the holdings of the

³ The agreement with the University of Michigan has been made public as required under the Michigan Freedom of Information Act. Under this agreement, the University is provided with a digital copy of all scanned books, but it can use these copies within the confines of U.S. Copyright Law (Band, 2006). Tom Garnett, the director of the Biodiversity Heritage Library, claimed that, "Google had a very restrictive agreement, and in all our discussions were unwilling to yield." These terms included the requirement that libraries put their own technology in place to

initial four research libraries (there are currently eighteen libraries and organizations involved in the project) were to be scanned: only some 40,000 at Harvard and only a limited number of works already in the public domain at the NYPL and Oxford (some fragile and rare book holdings at the NYPL and an unspecified number of pre-1900 publications in the Bodleian at Oxford). The main effort would involve the digitizing of eight million books (including those in-copyright) in the Stanford University Library and the seven million at Michigan. But under this program, only a portion of those titles would be made fully available on the Web. In the words of Elizabeth Edwards, a Stanford librarian: "Google will be responsible for determining what's in copyright and what's not if there are any questionable materials, and *copyright will drive what will be fully displayed*" (Edwards, 2005. Emphasis added).⁴

So, were they all to be digitized instantaneously today, just how many of those eight million books held by Stanford's library would become available for free reading on the Internet? A very approximate estimate made in 2005 put the figure at about ten percent, that being the rough portion of the University's total holdings that were published before 1923, and therefore clearly out of copyright.⁵ Even if this were a reasonably accurate guess, it would be useful to confirm it, and of interest to know the corresponding proportion of the seven or so million titles from the University of Michigan that are expected to be rendered fully accessible by this agreement.

But rather than having an answer that would reflect the history of book acquisition by Stanford University and the University of Michigan, what would the answer be if we were to consider the union of all the libraries in the U.S.? If we attend to the initial visionary statements of Librarian Keller, or of Messrs. Brin and Page, the eighteen libraries and organizations currently involved in the Google Libraries Project and their holdings represent no more than a beginning. Indeed, how many more books will come out of U.S. copyright in the course of the next two decades, and thus could be made available via the Web for unrestricted browsing, searching and downloading? That is another, down-to-earth question that could be answered for U.S. copyrighted books *in toto*. Furthermore, undertaking this line of inquiry will make it possible to take the additional step of quantifying the first-order effects of each among the several, successive changes in copyright law during the past century that – as Elizabeth Edwards has noted – "will drive what can be fully displayed on the Web." Alternatively, we can turn the matter around and try to answer a related but significantly different question: what has been the magnitude of the effects of successive revisions of the 1909 copyright statute upon the

block commercial services other than Google (Hafner, 2007). Other details have also emerged, notably in public discussions within the research library community. See, for example, Edwards (2005).

⁴ According to Elizabeth Edwards, for works in copyright, a "click-through" link would be provided to the appropriate Office of Copyright (Library of Congress) WorldCat record, from whence the would-be user could proceed to try to locate and contact the current holder of the copyright and obtain permission to secure a copy – from GooglePrint, or perhaps another such service. For an entertaining exposé of the realities that presently would stand between finding the book title on Google, or the original copyright registration information from OCLC and being able to discuss a license with the current owner of a 1930 copyright, consult Lessig (2004:222-223).

⁵ Federal government publications are placed in the public domain from the outset, and it is not clear whether the ten percent figure includes Stanford's extensive holding of such documents. A "ten percent" rule of thumb for the proportion of works ever-copyrighted works in the U.S. that are currently in the public domain appears to have emerged in casual discussions. For example, Chris Anderson, Editor-in-Chief of *Wired*, introduced a public Forum discussion held at the NY Public Library on November 17, 2005 by noting that "there are about 32 million books out there" and "maybe" as many as "three million of those are out of copyright...." (Author's Guild, 2006).

numbers of books and pamphlets that are *not* scheduled to become fully accessible on the Internet in each year of the coming quarter century?

Events have moved quickly enough to make trying to supply an answer to that question a matter of some practical relevance, as well as intrinsic interest. As Google's library-book scanning project has unfolded in the years since the December 2004 launch of the Google Print Library Project, a variety of issues have emerged to complicate the picture -- without altering that basic point regarding the critical role of copyright protections in affecting access. Google initially contemplated three different on-line presentations for three different kinds of books: books in the public domain, book in copyright whose publishers had given permission for Google scanning and display,⁶ and books in copyright for which publishers had not given such permission. In the latter case Google, displayed a fragment of the text, which was enough to elicit voluble public complaints from some of the affected authors and publishers. On August 12, 2005, Google announced a delay in the project in order to give publishers and other copyright holders the chance to "opt out" of having their copyrighted works scanned. This, however, did not forestall the filing of a lawsuit on September 20, 2005, in which three authors (including the Authors Guild, a group representing more than 8,000 published authors) claimed that Google had engaged in "massive copyright infringement." Google then resumed its scanning program (in November 2005), focusing mainly on the older works that were unambiguously out of copyright, and works that were always in the public domain.(Wyatt, 2005a, 2005b).⁷ A month later five major publishers sued Google requesting damages and injunctive relief, asserting that the "massive, wholesale and systematic copying of entire books still protected by copyright" infringes on the publishers' rights (Toobin, 2007).

Google's dominance of the movement to scan the world's books was short-lived: on October 3, 2005, the Open Content Alliance (OCA), an organization composed of corporations, nonprofit groups, and universities, announced plans to digitalize hundreds of thousands of out-of-copyright works. Hosted by Yahoo and joined by Microsoft, this consortium makes its books accessible to any search engine, a significantly less restrictive covenant than the one employed by Google (and indeed, recently by Microsoft), whose books are only available through a Google search.⁸ The OCA has avoided copyright trouble by concentrating primarily on public domain works as well as employing an "opt-in" strategy (as opposed to Google's "opt-out" strategy), in which members ask copyright holders for permission before digitalizing a work (Hafner, 2005a, 2005b). The relative freedom from restrictions and corporate influence has made the OCA an attractive alternative to Google for a number of libraries, including the Library of Congress, the National Archive in England, and the University of California, Berkeley (Hafner, 2007).

We may suppose that if this develops into a movement like that in "open source software", whereby the OCA and other organizations receive grants and gifts for scanning facilities that are provided for volunteers, the effect will be to create accessible content that could be searched online by generic (untied) web search engines, even those provided by Google and Microsoft. This could create another occasion to fill the would-be readers' screen with "related

⁶ This category of was were being digitalized under Google's "Partner Program", in which publishers "opt-in" copyrighted material to be scanned by Google and offered in "snippets" online.

⁷ For a broader discussion of the Google litigation, see Band (2006), Hanratty (2005), and Varian (2006). Much of the debate centers around whether Google's "opt-out" feature is consistent with the principles of fair use doctrine.

⁸ A year after joining the OCA, Microsoft placed a similar restriction prohibiting books that it digitalizes from being included in search engines other than its own (Hafner, 2007).

advertising”, but, in the process it would cannibalize that portion of the tied commercial download market for books.

One may regard this prospect from two different but complementary angles. From the viewpoint of Google and Microsoft, or other major firms, how big the market is at present, and how large it will become within the near future, is a question that should have a direct bearing on how much it is worth trying to keep hold of the “tied book-search and download” business. From the perspective of the book-reading consumers, on the other hand, the question can be phrased differently: how many books will be returning to the public domain, and might therefore become universally accessible under the terms envisaged by the Open Content Alliance – with the possible option to use higher cost “enhanced” commercial search services that might still be provided by Google and others? A first-order answer to both questions can be provided by the same set of numbers: the count of books that were copyrighted in the US and subsequently returned to the public domain, year-by-year, up to the present and looking forward into the 2020’s. In other words, what we need to know is the *degree* to which “copyright will drive what will be fully displayed”.

3. THE HISTORY OF CHANGES IN THE DURATION OF PROTECTION UNDER U.S. COPYRIGHT LAW

The U.S. Constitution clearly differentiates between property and creative property. While exclusive property rights exist for the former, the Constitution demands that Congress take back the rights to creative property after a limited time and place it in the public domain. The goal of copyright extends only as far as it takes to "promote the Progress of Science and useful Arts". Beyond this, copyright theoretically does not transfer monopoly status to ideas; it is intended to provide incentive to create, not rents for the author (Lessig, 2004).

An ideal copyright system would counterbalance the incentive to produce as much high quality work as possible against the benefits of having a large public domain. On the one hand, the opportunity cost of writing decreases when protection is extended. On the other hand, having a healthy public domain is an important public good, particularly for the "promotion of Progress". Academics, artists, and authors all have great interest in there being a bountiful public domain; the ability to use and transform ideas is essential to the progress and advancement of thought and culture. The framers of the Constitution understood this, and they ensured that all creative ideas would eventually fall into the public domain.

At the beginning of the twentieth century, the maximum copyright duration was 42 years (a 28-year initial term and a 14-year renewal term). By the end of the century, the duration of all copyrighted works was the author's life plus 70 years. Undoubtedly, the driving force behind this legislation was in favor of lengthening protection. The first act extending duration was the Copyright Act of 1909, which lengthened the renewal term to 28 years, making the maximum renewal period 56 years (an initial 28 year period followed by a 28 year renewal period).⁹ The 1909 Act dictated copyright duration until 1962, when Congress began a series of term extensions which have defined copyright law ever since. In the last half century, Congress has extended the copyright term eleven times, beginning with the Act of 1962, which kept copyrights in their renewal term that were set to expire on September 19, 1962 out of the public domain.¹⁰

⁹ Act of March 4, 1909, ch. 320, 35 Stat. 1075.

¹⁰ Act of 1962, and sequelae through to 1976: P.Ls. 87-668, 89-142, 90-141, 91-147, 91-555, 92-170, 93-573.

The effect of these changes was to set the maximum statutory duration awarded to newly copyrighted works (before 1978) to 75 (28+47) years.

These acts culminated in the Copyright Act of 1976,¹¹ which officially set the renewal term for all works published before January 1, 1978 to 47 years, and set the copyright duration for all those published after 1977 to the author's life plus 50 years. This system enabled U.S. copyright law to accord more closely with foreign laws and paved the way towards the U.S. eventually joining the Berne Convention (an international copyright agreement), which has a baseline requirement including a copyright term for works by individual authors of life of the author plus fifty years. The 1976 Act allowed the United States to accord to the Berne Convention (which it signed in 1989) primarily by changing the copyright system from a “conditional” one which premised the existence and continuation of copyright on compliance with formalities to an “unconditional” system in which a reduced set of voluntary formalities plays only a minor role.¹² Under the act, the renewal requirement was completely abolished and copyright was granted to all works at the moment of their fixation, whether the work had been registered or not.¹³

In 1992, Congress eliminated the renewal requirement for works published between 1964 and 1977, automatically granting these works renewal.¹⁴ What motivated Congress to remove this formality, in turn preventing thousands of works with little commercial value from joining the public domain? In the testimony before Congress, seven of the nine witnesses were in favor the Act, arguing that it would prevent tragedies that had occurred in the past where an author forgot to renew his copyright or a widow was unaware that renewal was necessary. For example, Barbara Ringer, the former Head of the Renewal and Assignment Section of the Copyright Office's Examining Division, testified:

As I write this statement I have a mental image of my office in the old Copyright Office ... and of the constant procession of tragedies that were played out there. Some of these tragedies were revealed in correspondence: renewal applications received too late or inquiries (some from Congressional offices) about what to do now that the first term had expired. Worse were the frantic phone calls; if there was still any time left in the 28th year it was the Office's policy to move heaven and earth to get the renewal registered in time, but for claims received too late the pain we felt in conveying this message was nothing compared to the reaction on the other end of the line (Subcommittee on Patents, 1991).

Ringer's testimony, as well as the emotional statements of a widow who lost royalties from her late husband's work due to a technicality, contributed to the view that the renewal requirement was an antiquated formality. The only testimony against the bill came from an old-movie store owner and from law professor L. Ray Patterson, who argued that the law would provide a benefit for a large group of persons who neither desire nor need it, that economically

¹¹ The Copyright Act of (October) 1976: P.L. 94-553 90 Stat. 2541, amending U.S.C. §17, in effect January 1, 1978.

¹² For more on the Berne Convention and the steps that the U.S. took to join in the late-1980s, see Sprigman (2004).

¹³ Under the 1976 Act, registration remained a prerequisite to the initiation of an infringement suit.

¹⁴ Copyright Amendments Act of 1992: P.L. 102-307, 106 Stat. 266, amending U.S.C. §17, sect. 304.

valuable works will be renewed, and that it was contrary to the constitutional purpose of copyright.

Still more recently, the news-reading public was made aware of the most recent change in the terms of copyright by the 1998 controversy over passage of the Sonny Bono Copyright Term Extension Act (CTEA),¹⁵ followed by litigation (*Eldred v. Ashcroft*) that was brought to the U.S. Supreme Court in an unsuccessful attempt to have the statute overturned on constitutional grounds.¹⁶ *Eldred v. Ashcroft* placed the economic arguments against retroactive copyright extension on a larger stage – indeed, numerous prominent law and economics professors, including a number of Nobel laureates, composed an amicus curiae brief espousing such arguments – but the CTEA was upheld as constitutional.

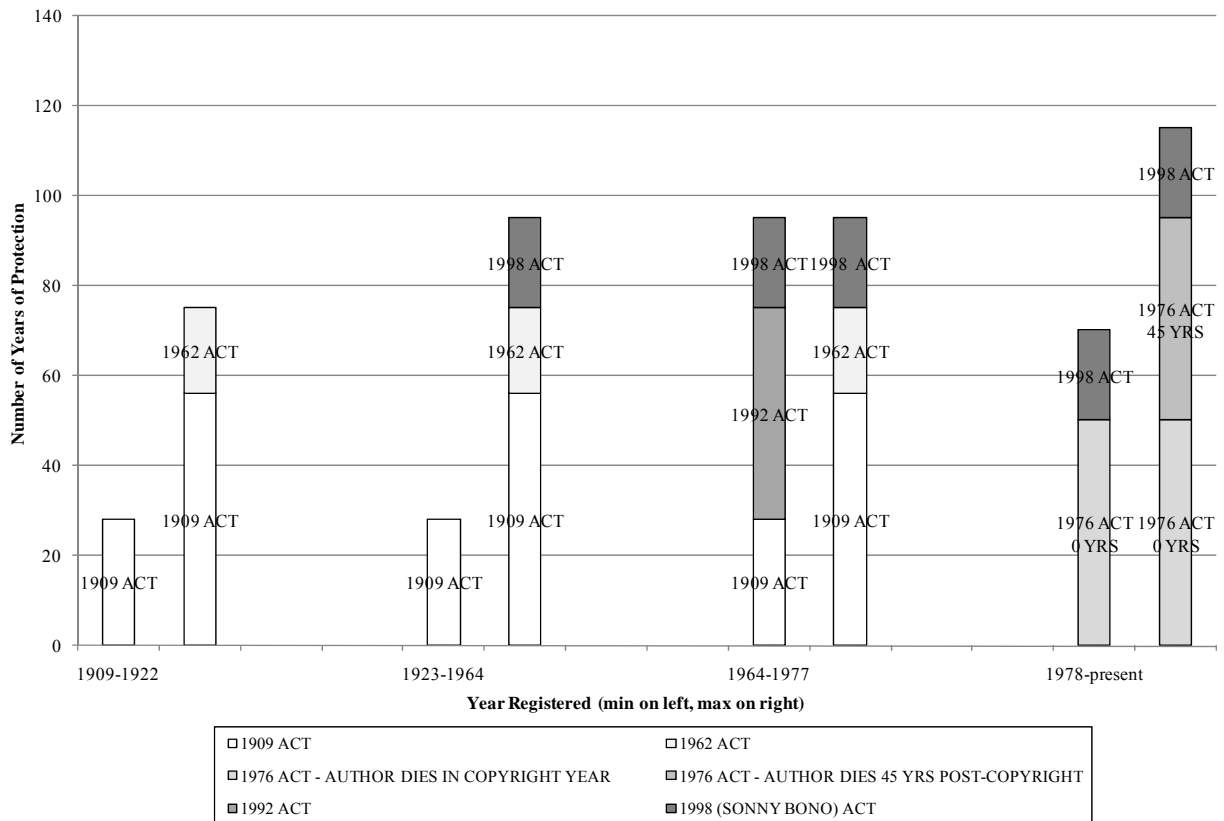
This attention was certainly warranted, moreover, because the CTEA's consequences will persist beyond the first-order effect of its lengthening of the term of protection by 20 years for works copyrighted after January 1, 1923. Works copyrighted by individuals since 1978 were granted a term limited to the author's "life plus 70" rather than the pre-existing "life plus 50." Works made by or for corporations (referred to legally as "works made for hire") were granted 95 years. These extensions were applied retroactively to works in copyright at the time of the Act, implying that some works copyrighted before 1978 were thus shielded for a total of 95 years, regardless of how they were produced. The rescue of Disney's copyright on Mickey Mouse was the legislative achievement that attracted popular notice; ironically, Disney itself has based many of its animated films on books that were in the public domain, including *Snow White and the Seven Dwarfs*, *Cinderella*, *Pinocchio*, *The Hunchback of Notre Dame*, *Alice in Wonderland*, and *The Jungle Book*, released exactly one year after Kipling's copyrights expired (Sprigman, 2002).

Figure 1 displays in summary form the effects of U.S. legislative history on the upper and lower limits of the term of protection on works registered in successive intervals since 1909. For works published before 1964, the lower bound is the initial copyright term, and the upper bound is the initial term plus the renewal term. The 1962 Act retroactively provided seventeen more years of protection for all renewed copyrights registered between 1909 and 1964, and the Sonny Bono act retroactively provided an additional twenty years of protection for books in their renewal period published between 1923 and 1964. The upper bound remains the same for books registered between 1964 and 1977 as it was in the previous period (95 years), but the lower bound increases because the 1992 Act abolished the renewal requirement, providing equal protection to all books.

¹⁵ Sonny Bono Copyright Term Extension Act of 1998: P.L. 105-298, 112 Stat. 2827, amending U.S.C. §17, ch. 3.

¹⁶ For the legislative background of the Sonny Bono CTEA (1998), and a reflective treatment of the Supreme Court challenge by the counsel for the plaintiff (Eldred) see Lessig (2004, ch. 13-14), a work that has been influential in directing more public attention to the cumulative effects of the legislative changes in U.S. copyright law since 1790. *Eldred v. Ashcroft*, 537 U.S. 186 (2003) was a case heard before the Supreme Court of the United States, challenging the constitutionality of the 1998 CTEA. Following oral arguments heard on October 9, 2002, the court held (on January 15, 2003) the CTEA constitutional by a 7-2 decision.

FIGURE 1: MINIMUM AND MAXIMUM PROTECTION AVAILABLE FOR COPYRIGHTS REGISTERED, 1909-PRESENT



For the post-1978 era, it is necessary to use a heuristic devise to exhibit the range of potential upper and lower term limits due to the Act of 1976, which ended grants of fixed terms for the initial copyright and its (optional) renewal and replaced them with variable terms defined by adding a specified number of years to the life of the author. Purely for this illustrative purpose, Figure 1 displays the effects of copyright legislation on post-1977 registered works under two alternative (arbitrary) assumptions: in the right-most member of the pair of vertical bars it is assumed that the author survived 45 years beyond the year of the publication's copyright year, providing a proxy for the "maximum protection length". The left-most member of the pair assumes that the author of the work did not survive beyond the year of its publication – indeed, this is the minimum length accorded under the new regime.

4. COPYRIGHT LAW AND THE PUBLIC DOMAIN – A COUNTERFACTUAL STRATEGY FOR QUANTIFYING THE EFFECTS OF THE LEGISLATIVE HISTORY

Producing a less arbitrary, statistically appropriate assessment of the changing duration of the U.S. copyright term for the post-1977 period is not a trivial matter. Doing so would necessitate obtaining the convolution of two empirical distributions: the age distribution of authors at the date(s) of their publication(s), and the distribution of the authors' ages of death. In the case of books and pamphlets by individual authors, upon which we will concentrate, sample distributions of both kinds may be constructed from existing catalogues of published books, and

successive cohort life tables. This would require modifying the latter for the differential survival rates of (male and female) book authors, on the basis of sample data compiled from biographical entries of American writers. The task can be greatly simplified by settling for approximations of the expected duration of life beyond the date of book publication, calculated for successive authorship cohorts under the (questionable) assumption that age of authorship and age of death are statistically independent.

In working out the approach just described, it became apparent that this challenging empirical task could be almost completely finessed by posing a related but slightly different question about the quantitative impact of copyright legislation. Rather than providing calculations that require knowing when a copyrighted work published in year T could be expected to enter the public domain, let us turn the matter around and ask the following: "How many books published at T will remain *out* of the public domain until year $T+N$?" Because of the specific chronology of changes made in the U.S. copyright statutes, it turns out that answers to that question may be found without any information about the ages to which authors are expected to survive following publication of their own works. Indeed, for works first entering U.S. copyright in each year before 2003, it is possible to make quite precise counts of the number of books that remain out of the public domain – in every year until 2027.¹⁷

An additional advantage afforded by the approach just described is that it allows us to handle a set of questions that have remained beneath the surface of the foregoing discussion. To this point, the exposition of the research strategy has quietly avoided any suggestion that the research described would yield a comprehensive assessment of the consequences of the U.S. statutes that have assigned (limited) legal monopoly privileges to holders of copyright in books – let alone the system of copyright as a whole. Yet one cannot simply ignore the argument that there are potential positive social effects as well as private benefits to be gained from the award of copyright protection. The social benefits are supposed to come because a copyright is a potentially valuable, transferable private economic asset (sometimes viewed as an option, because its future worth is uncertain) that may induce prospective recipients to invest resources in producing "original works of authorship fixed in any tangible medium of expression" (as the matter is phrased in 17 USC §102(b)).¹⁸ Therefore, the effects of such additional incentives as have been afforded by the changes in the copyright statutes upon the flow of new works have to

¹⁷ Qualification of "precise" in this statement is necessary because during the period 1909-1926 the registration statistics provided for the comprehensive category of "books" by the Annual Reports of the Register of Copyrights included "contributions to newspapers and periodicals," whereas the latter were removed from the "books" series reported for 1927-1959. Therefore, adjustments must be made (by lowering the former set of registration figures) in order to arrive at a series whose scope is consistent over the whole of the 1902-1977 period. A second ambiguity-raising complication arises from the Copyright Office, in 1978, having changed all the detailed classifications in which registrations were reported. This introduced the category "monographs" in place of "books", which in 1984 was combined with "machine readable texts." A decade later, the latter sub-class was relabeled "computer related works".

¹⁸ The formulation in the text deliberately eschews the rhetorical constructions frequently employed by copyright lawyers, and which have been embraced in conventional introductions to the economics of copyright. The basic analysis conceptualizes the institution as a necessary and reasonably efficient (if not always harmonious) bargain struck between "author" and "audience". The latter are willing to pay something more (in royalty mark-ups on the physical costs of embodying content in a convenient physical medium) more for their reading pleasures, so that the former will have a greater incentive to provide more and better content to be read. "Producers" and "users" occasionally take the places of "authors" and "audience" in these simplified stories. See Paul Goldstein (2003) for an illustration of this expositional genre by a leading legal scholar, whose popular account periodically recognizes that copyright's highway has been largely laid down in the interests of the publishing industry, not the authors.

be recognized in assessing the numbers of books (or other works) that are kept out of the public domain when the term of copyright protection is extended. Such recognition can be accorded in numerous ways, but the simplest treatment of the problem is to notice that if no attempt is made to estimate the effects upon the number of book registrations of increasing the private economic value of copyright by extension of the term of protection, the figures obtained for the additional numbers of books published after 1923 that are being held out of the public domain will be "lower-bound" estimates. For the purposes of demonstrating that the magnitude of those effects is bigger than is generally recognized, however, a conservative – lower-bound – estimation method is just what one might want, and is the one we employ in our analysis.

5. ANALYSIS

5.1. HOW MANY BOOKS HAVE BEEN KEPT OUT OF THE PUBLIC DOMAIN?

This section presents the results of a series of calculations which estimate the number of books kept out of the public domain by twentieth century U.S. copyright legislation. Using copyright registration and renewal data from various Annual Reports of the Librarian of Congress and Annual Reports of the Register of Copyrights, we commence this exercise by determining the number of books registered after 1902 (when our data begins) that are in the public domain. We then ask the question, "How many books would have been in the public domain if each law were not passed?" That is, we calculate a counterfactual in which each legislative act (and all the subsequent laws) does not exist, providing us with an estimate of each law's marginal effect on the public domain.

We illustrate the method by beginning with calculations that assess the marginal impact of the Sonny Bono Act. Had this act not been passed, copyrights passing their 75th year of age (including Steamboat Willie!) would have fallen in the public domain between 1998 and the present. The counterfactual thus includes these works as being in the public domain. Likewise, to estimate the effect of the 1992 Act, we assume a counterfactual world where the 1992 Act and Bono Acts do not exist. In this case, copyrights that would not have been renewed between 1991 and 2005 would have fallen into the public domain. Because no data is available for book renewals post-1977, this counterfactual requires a bit of estimation (and even if the data were available, the 1992 Act eliminated the renewal requirement anyway). We report here the lower bound estimates, which we derive by assuming an increasing renewal rate in the post-1977 period.¹⁹

Since copyright registration has been voluntary since 1978 (due to the conditions of the 1976 Act), our data underestimate the true number of post-1977 copyrights. Yet, incentive to register remained significant in this period, especially for larger publishers, as registration remained a prerequisite to the initiation of an infringement suit. Regardless, this problem only arises in our counterfactual analysis – and as desired, it entails that we provide a *lower* bound of the number of books kept out of the public domain. It is possible, of course, that copyright term extension may have another, non-quantified effect – namely, the increased potential rents may induce more authors to write more books. However, as Landes and Posner have pointed out, the commercially valuable portion of the life of most book copyrights is likely to end well before the

¹⁹ To determine the lower bound, we conduct an OLS regression on the renewal rate for 1947-77, and place the fitted values for 1978-2027 as the estimated renewals. The fitted value of the equation is $\text{Renewal Rate} = \beta_0 + \beta_1 \text{Year} + \varepsilon$, where Year is marked 1 (1947) through 31 (1977), providing estimates $\beta_0 = 0.013224$, $\beta_1 = 0.05696$.

initial 28-year period of protection that was available until 1976 (Landes and Posner, 2003); moreover, before renewal was made automatic, copyrights were renewed beyond their initial term for only a minor fraction (under 0.15) of the books published in the U.S.²⁰ The evidence that renewals, and longer renewal durations are considered worthwhile for the fringe of exceptional works, does not carry an implication that this option is sufficiently valuable to exert a significant quantitative effect on the supply of new titles – although it might significantly alter differential incentives to publish works of certain kinds. Inasmuch as our analysis would not be able to detect the latter compositional shift effects, we are justified in proceeding by dismissing them (as quantitatively negligible) for our present purposes.

Similar counterfactuals are calculated for the 1976 and 1962 Acts (the formulas for all calculations can be found in the Appendix), and the results are displayed in Figure 2. Since there is little "action" in this figure before 1990, we replicate it post-1990 in Figure 3. It is apparent in these figures that the 1962 and Sonny Bono Acts had little marginal effect on the number of books available in the public domain. However, the 1976 and 1992 Acts have had an immense impact, with the effect of the former increasing by the year. The common denominator between these two acts is their abolishment of the renewal requirement. In the time period under question, the book copyright renewal rate wavered between 3% and 20% (except for the outlier year 1973, see Figure 4), indicating that authors (or publishing houses) did not consider the renewal fee and time cost associated with obtaining a renewal worthwhile for most works. Since the 1976 and 1992 Acts abolished the renewal requirement (the former for books copyrighted post-1977 and the latter for books registered between 1964 and 1977), between 80% and 97% of books that would have fallen into the public domain under previous regimes receive an additional term of protection.

²⁰ Landes and Posner (2003) undertake to estimate the effect of changes in the expected duration of renewals on the volume of new copyrights, but their conceptual approach and econometric procedures are unsatisfactory. This matter is the subject of separate (forthcoming) paper, where we re-run their model using book copyright data and obtain results indicating no significant effects of changes in statutory terms. That is, the length of the renewal period has no statistically significant effect on the numbers of books copyrighted in any of the years following the change.

FIGURE 2: CUMULATIVE BOOKS IN THE PUBLIC DOMAIN, INCREMENTING FOR EFFECT OF EACH LAW, BOOKS REGISTERED 1902-PRESENT

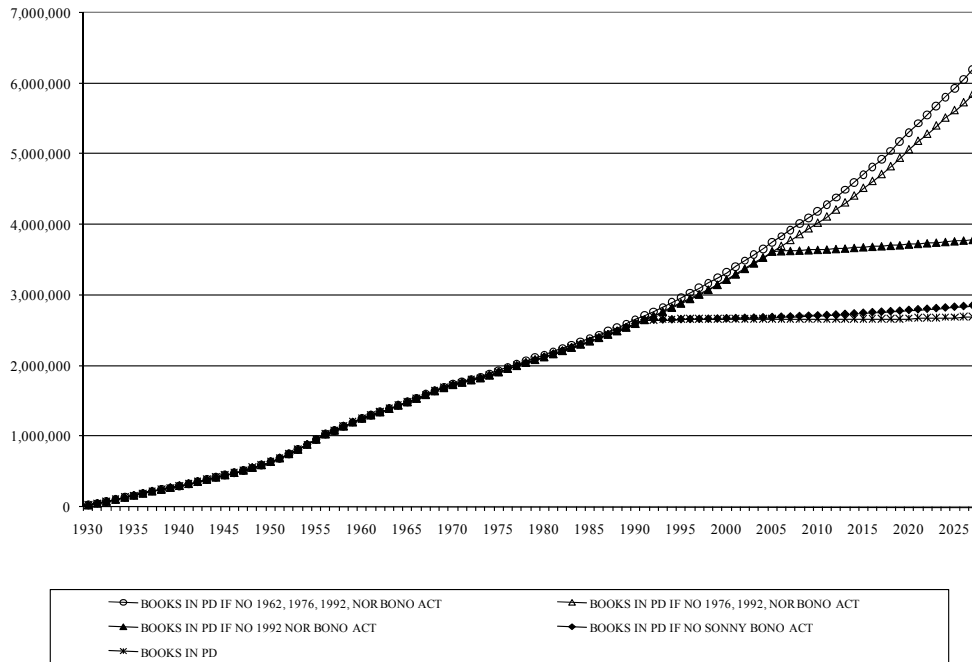


FIGURE 3: CUMULATIVE BOOKS IN PUBLIC DOMAIN (1990-2027), INCREMENTING FOR EFFECT OF EACH LAW, BOOKS REGISTERED 1902-PRESENT

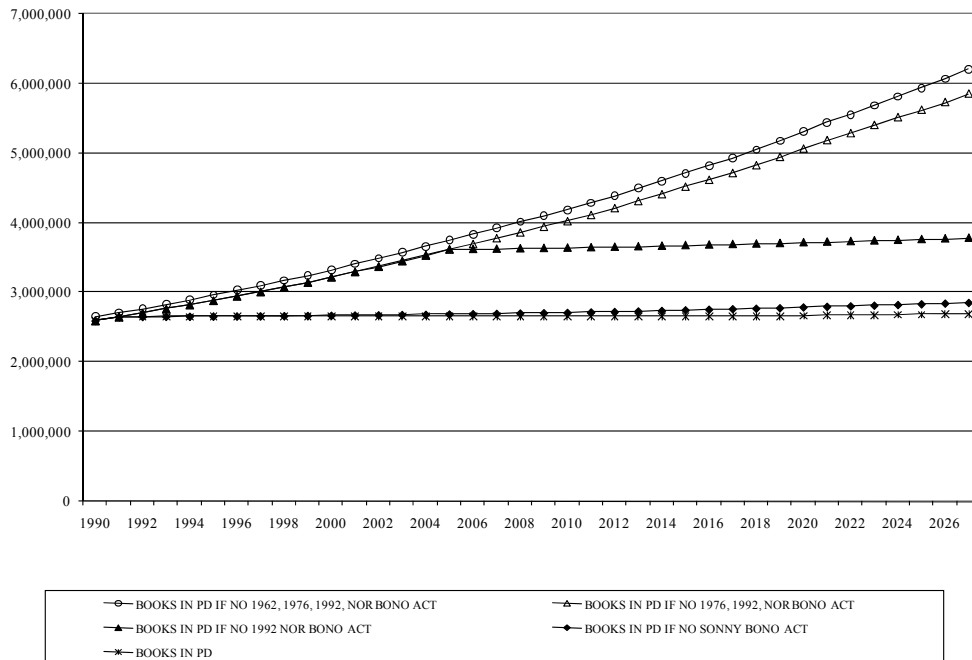
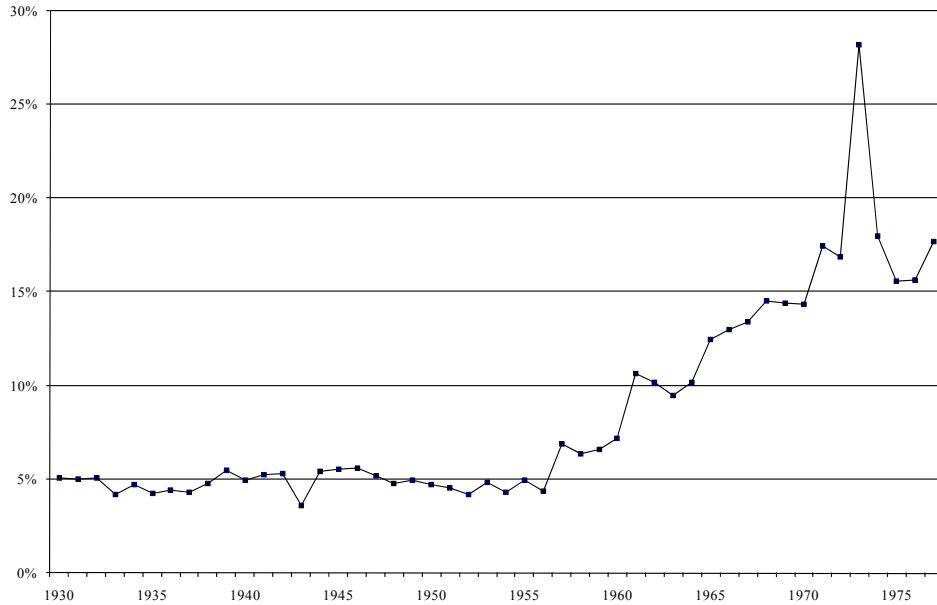
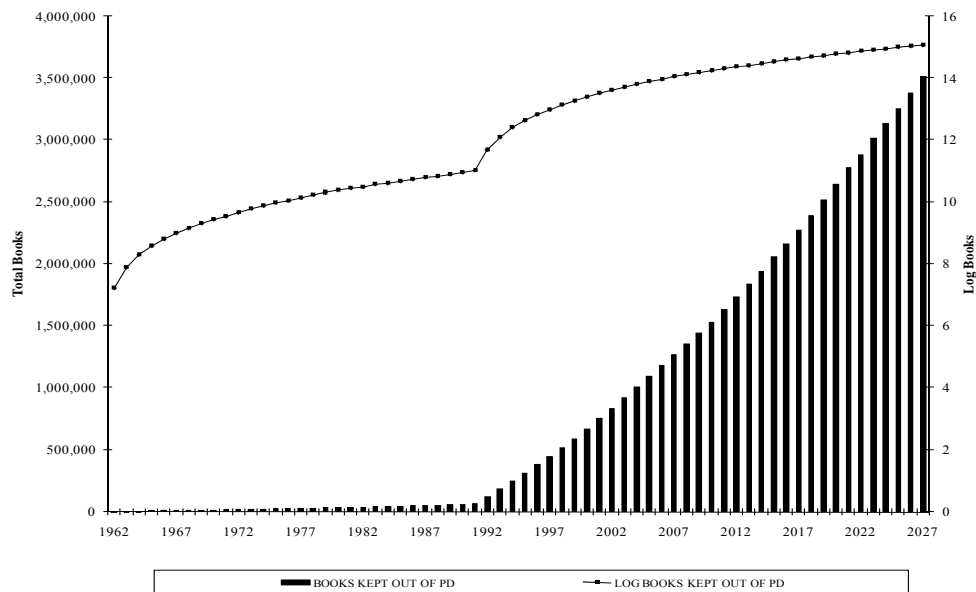


FIGURE 4: BOOK COPYRIGHT RENEWAL RATE, 1930-1977



Figures 5 and 6 present the results of these calculations in a different way: instead of estimating the counterfactual volume of books that would have been in the public domain, we ask the question "How many books were kept *out* of the public domain due to these changes in U.S. copyright law?" The logic underlying each calculation is similar to before, and again, the

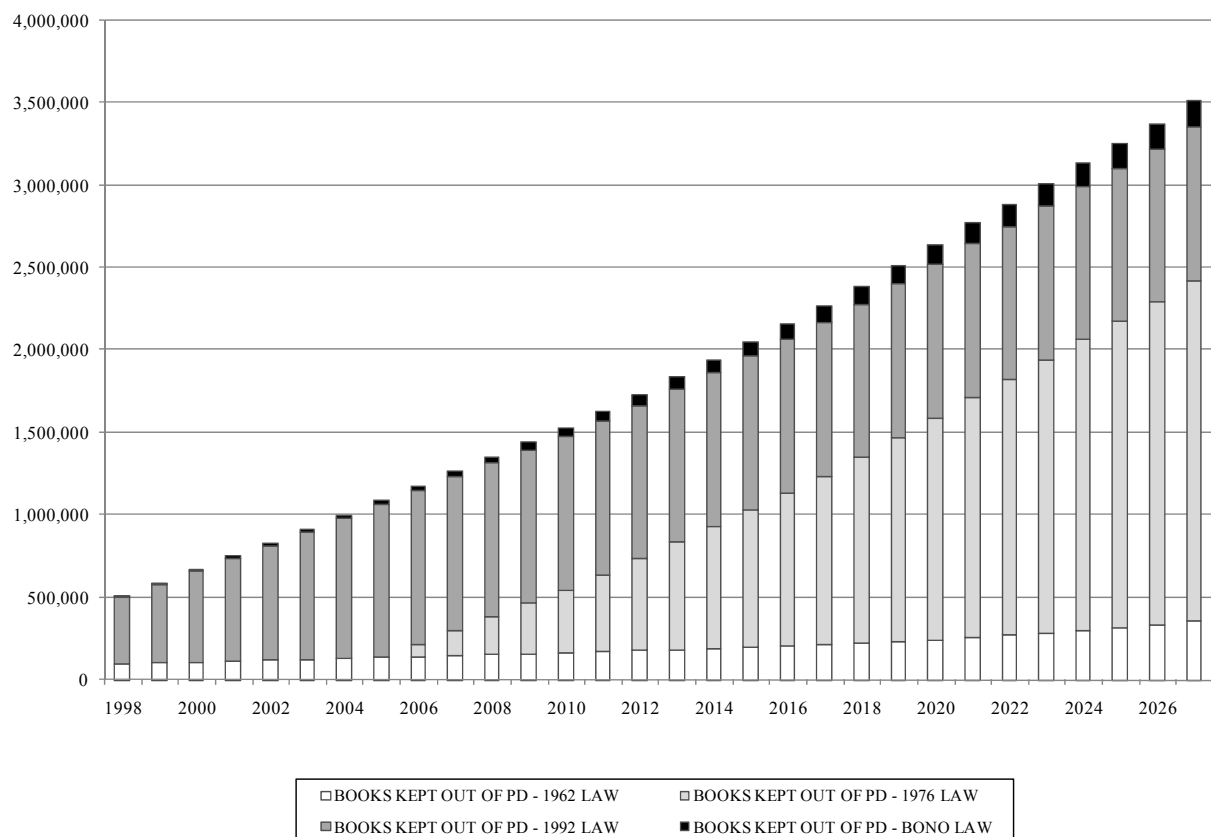
FIGURE 5: CUMULATIVE BOOKS AND LOG OF BOOKS *NOT* ENTERING PUBLIC DOMAIN BECAUSE OF 1962, 1976, 1992, AND BONO ACTS, BOOKS REGISTERED 1902-PRESENT



1976 and 1992 Acts stand out as the primary culprits. Indeed, Figure 5 reveals that the aggregate number of books kept out of the public domain remained relatively small before the 1992 Act, after which the (cumulative) number of titles began rising dramatically. This is also evidenced by the logarithmic plot in the same figure; the changing slope of that curve exhibits the discontinuity, or “shock” to the rate of increase in the volume of works being withheld from the public domain as a consequence of the retroactive removal of the renewal requirement as dictated by the 1992 legislation.

Figure 6 breaks down the aggregate results displayed in Figure 5, providing estimates of the effect of each law on the cumulative books kept out of the public domain. As expected, the 1992 Act immediately had a large effect after its enactment, but eventually is superseded by the 1976 Act (in 2017). The consequences of the 1992 Act were much greater than was appreciated at the time – the measure was presented as a technical adjustment that “grand-fathered” the removal of the renewal application formality for works copyrighted between 1964 and 1977. In this quiet way, an additional 47 years of protection were extended to a large number of books which otherwise have been allowed to fall out of copyright.

FIGURE 6: CUMULATIVE BOOKS *NOT* ENTERING PUBLIC DOMAIN BY LAW (1998-2027), BOOKS REGISTERED 1902-PRESENT



While the question of whether a statutory grant of *retroactive* extension of protection to existing works was consistent with the constitutional basis for Congressional legislation in this area would become a hot issue during debates about the Sonny Bono Copyright Term Extension

bill enacted in 1998, the question remained largely unnoticed in 1992. The effects of the 1992 Act were merely compounded by the 1998 CTEA – the books that it kept in copyright which otherwise would have been returned to the public domain will now not begin entering the public domain for another half-century, until sometime between 2059 and 2072.

The combined quantitative effects of the 1976 and 1992 Acts (as revealed by Figure 6) are staggering – by 2018, these two Acts will have kept over two million books with relatively little commercial value out of the public domain. Looking back at the estimates in Figure 2 and Figure 3, it may be seen that in the absence of this pair of legislative changes, the number of copyrighted titles that would have been returned to the public domain by 2027 in the absence of these laws would have been over 110% greater than the number that can be anticipated if renewal rates conform to the levels projected in these calculations on the basis of historical experience.

These findings support the emphasis that was placed upon the significance of the 1976 and 1992 pieces of legislation by Stanford's Lawrence Lessig:

This change [the 1976 and 1992 Acts] meant that American law no longer had an automatic way to assure that works that were no longer exploited passed into the public domain. And indeed, after these changes, it is unclear whether it is even possible to put works into the public domain. The public domain is orphaned by these changes in copyright law. Despite the requirement that terms be "limited", we have no evidence that anything will limit them (Lessig, 2004).

The "Sonny Bono" Copyright Term Extension Act of 1998 subsequently attracted wide public attention and newspaper coverage, in large part because of the longer-term implications of the extension of protection to digital music, images and video products, and Lessig's own role in arguing the Supreme Court suit brought by Eric Eldred to obtain an injunction against CTEA's enforcement. But that can be seen from our calculations to have been a battle fought over principles, rather than about practical contemporary consequences. The intent of the Constitutional powers accorded to the Congress for the promotion of "the Progress of Science and useful Arts" had already been palpably undermined by legislative measures that were passed with little public attention and, in the case of the 1992 Act, with virtually no opposition whatsoever.

5.2. HOW MANY BOOKS (UNAMBIGUOUSLY) HAVE BEEN KEPT OUT OF THE PUBLIC DOMAIN?

Our above estimates highlight the fact that many books registered between 1923 and 1964 were not renewed and are thus in the public domain. However, the transaction costs associated with searching for the copyright status and original owner of such books are often prohibitively expensive, forcing the organizations involved with the OCA (as well as Google) to play it safe and only scan books clearly out of copyright.²¹ Indeed, from the viewpoint of the individual would-be readers who sought access to works published in a particular year, the uncertainty of the copyright status of any particular title belonging to that publication vintage, and the potentially substantial costs entailed in determining the identity of the copyright that had been

²¹ For a broader discussion on the role that transaction costs play in limiting the amount of books available to be scanned, see Varian (2006) and Lessig (2005).

renewed, might mean that those volumes whose status remained uncertain were effectively not accessible from the public domain – whether or not they had been scanned for online access.

These considerations make it instructive to estimate the number of books that have *clearly* been kept out of the public domain by 20th century U.S. copyright legislation. To derive this estimate, we determine the number of books kept out of the public domain assuming that the renewal status of *any* book is indeterminable. Here, our results are different from those in section 5.1., since each book’s renewal status has no effect on the calculation. Thus, the 1992 law does not affect the number of books clearly kept out of the public domain, and there will be no effect stemming from the 1976 law until 2053 (when pre-1978 copyrights would have fallen into the public domain sans CTEA). The primary culprits are thus the 1962 Act and the Bono Act, both of which lengthened the maximum *possible* duration by extending the renewal period. We display our estimate of the effect of each law on the number of books clearly kept out of the public domain in Figure 7. We break this result down in Figure 8, which reveals that the CTEA and the 1962 Act – both of which extended duration retroactively – have prevented over 1.5 million books from currently residing in the public domain, and will keep over 2.3 million books out in 2027.

FIGURE 7: CUMULATIVE BOOKS CLEARLY ENTERING THE PUBLIC DOMAIN, INCREMENTING FOR EFFECT OF EACH LAW, BOOKS REGISTERED 1902-PRESENT

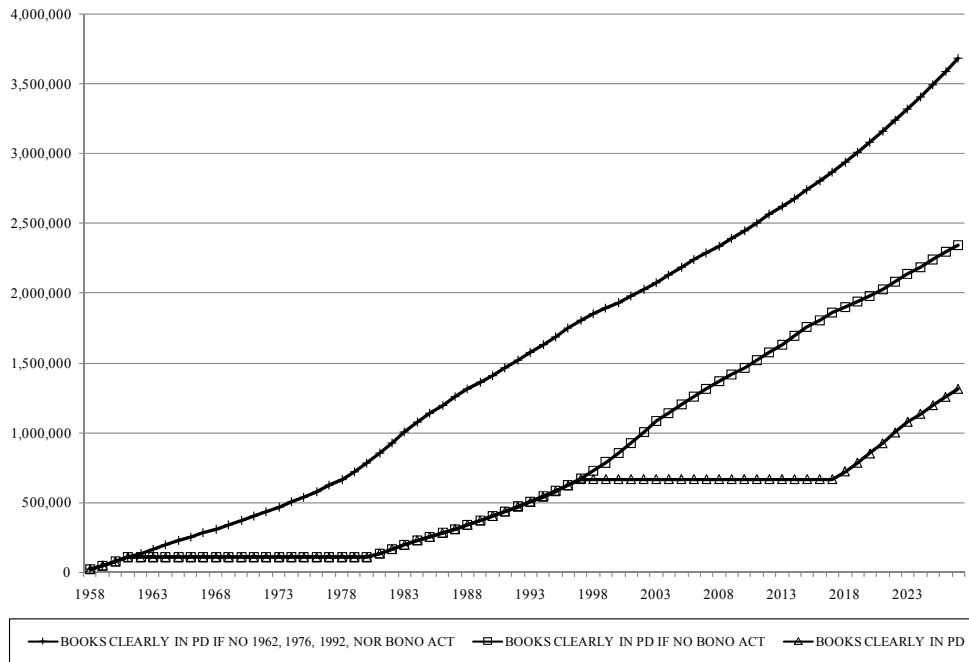
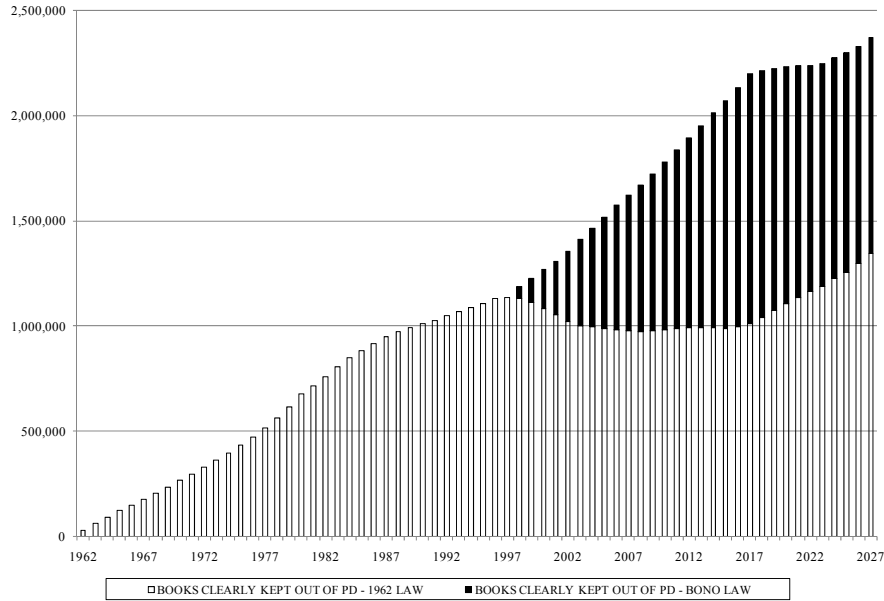


FIGURE 8: CUMULATIVE BOOKS CLEARLY *NOT* ENTERING PUBLIC DOMAIN BY LAW (1962-2027), BOOKS REGISTERED 1902-PRESENT



A related question which can be solved with our data is “How many books originally copyrighted in year T have been kept out of the public domain due to twentieth century copyright legislation?” By answering this question, we gain insight into *which* books have been kept out of the public domain and thus will not be available for download on Google and OCA websites.

To this end, we have placed each book in our dataset into its “vintage cohort” – the range of years in which it was initially copyrighted. For each vintage cohort, we have estimated how many books were in the public domain and how many books were *clearly* in the public domain

FIGURE 9: DEFINITE AND ESTIMATED NUMBER OF BOOKS IN PUBLIC DOMAIN IN 2005, BY 5-YEAR PUBLICATION VINTAGE

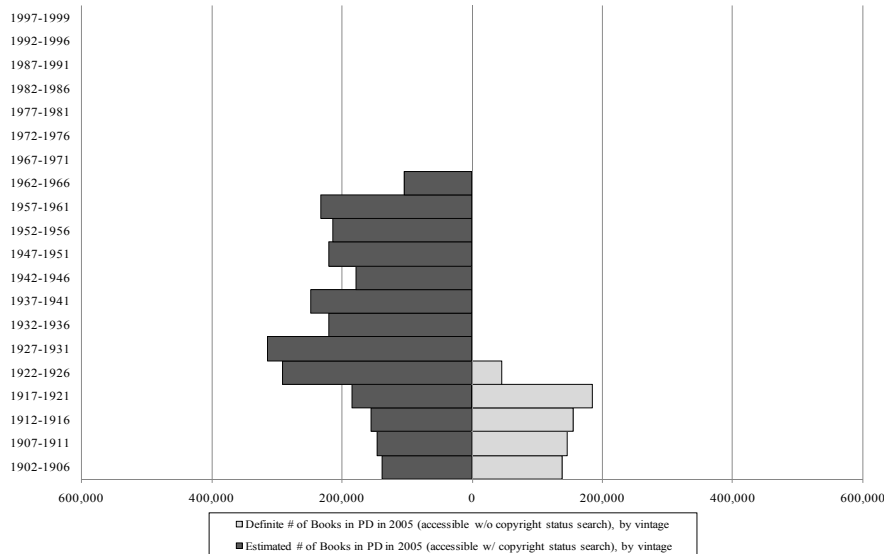


FIGURE 10: DEFINITE AND ESTIMATED NUMBER OF BOOKS IN PD IN 2005 (IF NO 1962, 1976, 1992, NOR BONO ACTS), BY 5-YEAR PUBLICATION VINTAGE

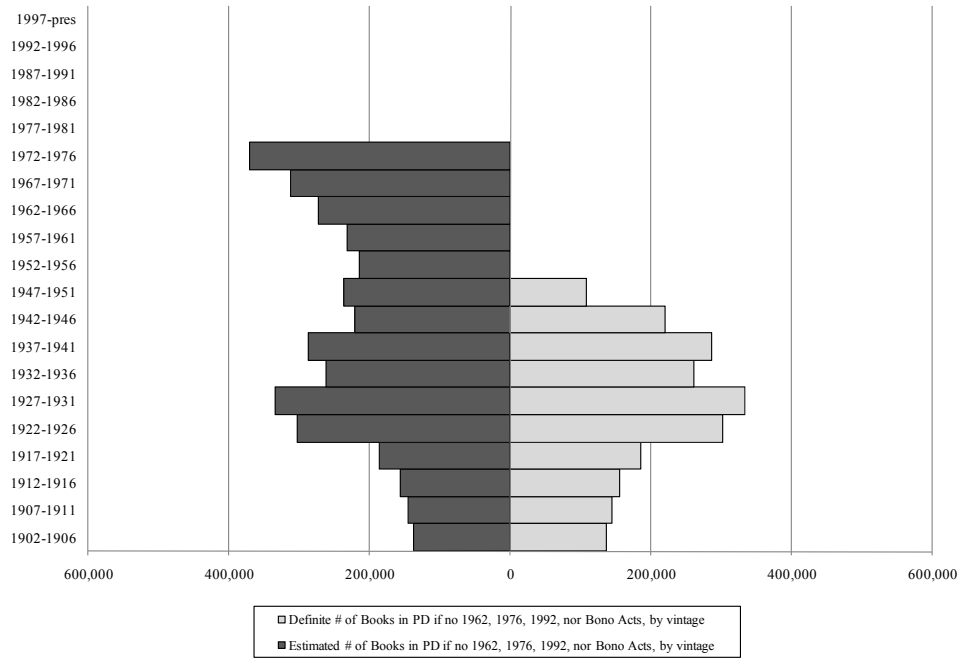
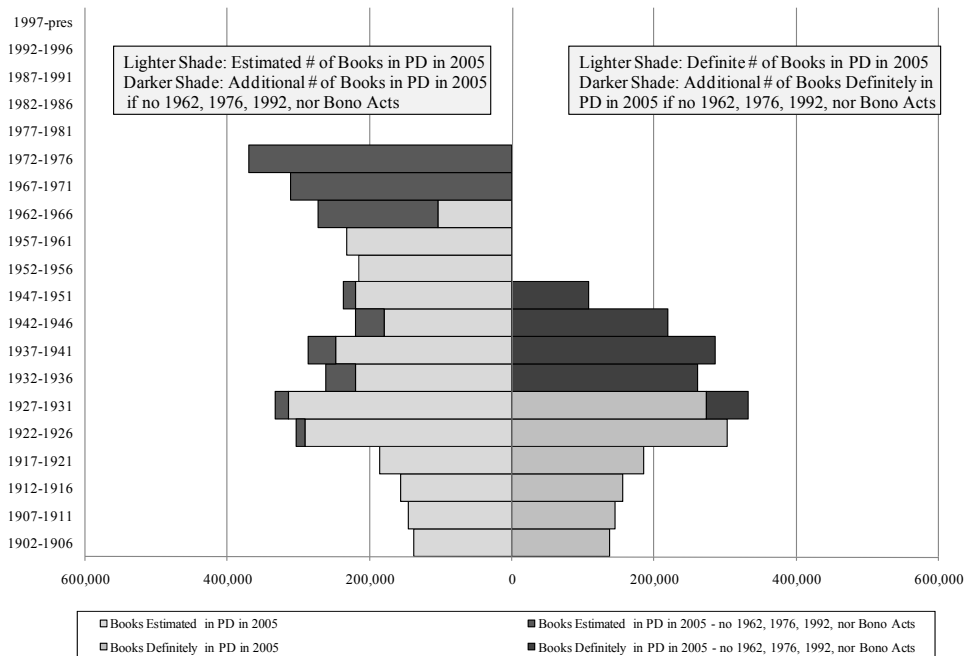


FIGURE 11: NUMBER OF BOOKS IN PD IN 2005 ACTUAL AND COUNTERFACTUAL (IF NO 1962, 1976, 1992, NOR BONO ACTS), DEFINITE AND ESTIMATED COUNTS



in 2005 and will be in 2025, and calculated the same estimation for the most significant counterfactual case in which the 1962, 1976, 1992, and CTEA never occurred. Figures 9 and 10 show the 2005 estimates as “population pyramids”, where the population in each vintage cohort is the number of books in the public domain (or the number clearly in the public domain). In Figure 11 these two estimates have been placed against each other. This figure illustrates in a straight-forward manner the magnitude of the 20th century legislation’s affect on the public domain.

Figures 12, 13, and 14 show the same estimates for the year 2025. The latter results are much starker than those for 2005: in 2025, there are eight (five-year) vintage cohorts in which *all* books would clearly have been in the public domain by 2025, in the counterfactual world, but are

FIGURE 12: DEFINITE AND ESTIMATED NUMBER OF BOOKS IN PD IN 2025, BY VINTAGE

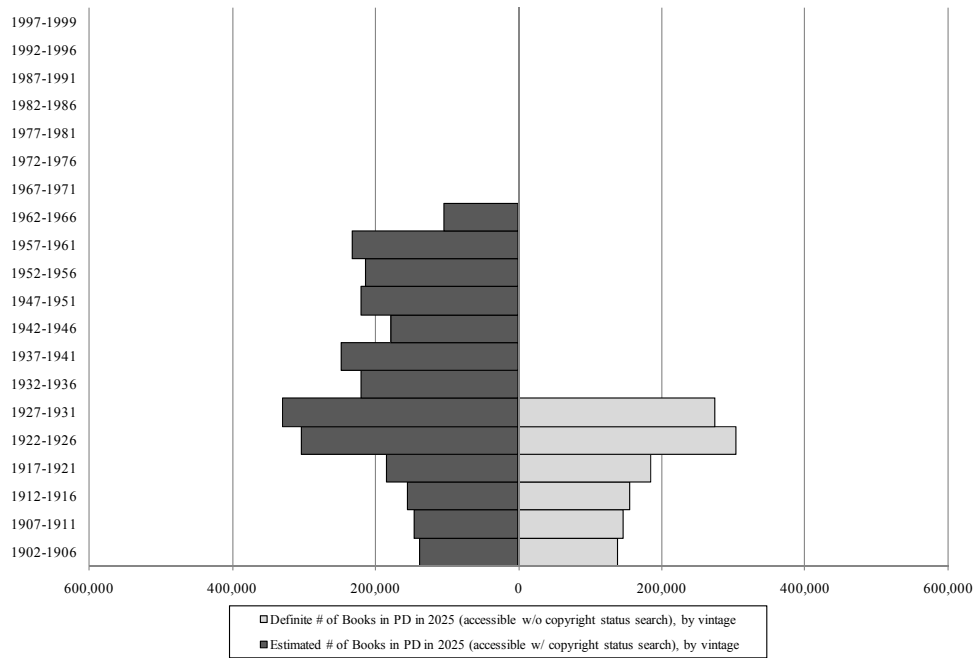


FIGURE 13: DEFINITE AND ESTIMATED NUMBER OF BOOKS IN PD IN 2025 FOR THE COUNTERFACTUAL CASE (NO 1962, 1976, 1992, AND 1998 ACTS), BY VINTAGE

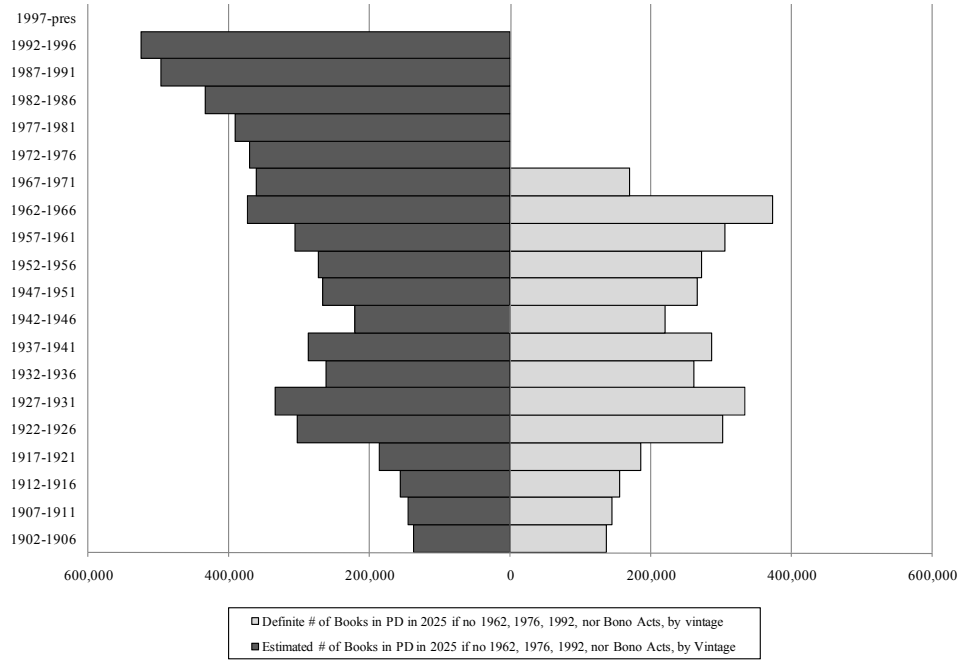
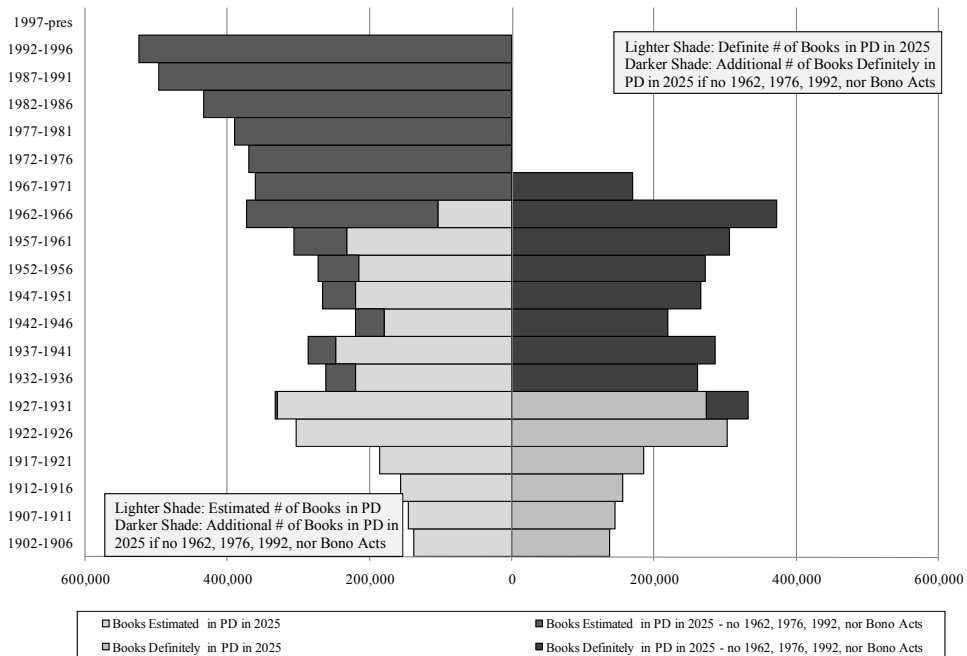


FIGURE 14: NUMBER OF BOOKS IN PD IN 2025, ACTUAL VS COUNTERFACTUAL (NO 1962, 1976, 1992, NOR BONO ACTS) COMPONENTS OF THE DEFINITE AND ESTIMATED COUNTS, BY VINTAGE



5.3. THE CULTURAL SIGNIFICANCE OF DIFFERENTIAL ACCESS TO “PUBLICATION VINTAGES”

The measurement of the “publication cohort effects” of U.S. legislative history on access to books (via the postponing of the re-entry of copyrighted works into the public domain) provides a simple way of indicating the future time-path of the differential “filtering out” of more modern contributions to the mass of cultural and scientific material that otherwise might gain free circulation via the Internet. Taken in conjunction with available compilations and samplings from the chronology of copyright registrations, the resulting numerical estimates offer some guide to the volume and content distribution of the published works in specific fields in which reliable technical and historical information, popular tastes, and socially acceptable modes of expression were changing – and consequently, to the qualitative effects of full Internet access remaining restricted, only to have the restraints suddenly removed. This aspect of the foregoing results should be of considerable interest both for research into the determinants of the “quality” of Web-accessed information and for studies of the formation of popular culture.

The latter observations follows quite directly from the economic proposition that resource allocation decisions in a market system are determined at the margin, in comparisons of incremental costs and incremental material gains among choice alternatives. The logic of this applies quite generally, and so is pertinent to the situation in the publishing industries and in the markets for information search and retrieval. There it is well recognized that the differentially lower costs of accessing and reproducing works which are no longer copyright protected is likely to induce their more frequent reproduction, and their greater exposure at sites accessible to the consuming/using public. The workings of the easier-to-use forms of automated search engines in the modern Internet environment would tend to reinforce this generic process, further amplifying both the widespread familiarity and more ubiquitous citation of the differentially accessible works, and thereby deepening and extending their impress upon individual tastes and collective cultural values.

A simple but striking illustration of the potential cultural impact of an interaction between new communication technology and copyright law is offered by the history of the “accidental social construction” of an American film classic. The copyright on Frank Capra's 1946 film *It's a Wonderful Life* was not renewed upon the expiration of its initial 28 year term, seemingly due to an oversight (the cost of the mandatory renewal registration being quite small at the time). That event subsequently has been regarded as a “tragic” accident, at least by some spokespeople for the intellectual property interests of the motion picture industry. But, only after its “fall” into the public domain did this particular film, largely ignored when it was first released and barely remembered – save by the most dedicated fans of Jimmy Stewart – commence its rapid ascent in the late 1970's to a perennial place in popular television-programming for the Christmas holiday season.

A parallel, but somewhat more intricate passage in the cultural history of the English reading public may be remarked upon here, indicating the broader scope of the issues upon which this straightforwardly quantitative research project will touch. William St. Clair recently has made a persuasive case for the idea that enduring literary tastes may have not only a materialist basis, but one that is quite serendipitous, in being shaped by quite transient conjunctures of events affecting the economics of the book trade (St. Clair, 2004:ch. 20-23). In the course of developing this thesis, St. Clair (2004) documents the persisting and remarkably strong impact of the poets and novelists of English Romantic period upon the reading public of the Victorian age, and shows that the literary canon that prevailed in 1900 owed much to the

particular circumstances that arose in the business of printing and publishing in Britain at a much earlier point in the 19th century. The application of stereotype printing technology in particular ushered in the profitable mass reprinting of inexpensive titles that could be kept "in print" for an unprecedented length of time – beginning with the cheap Bibles of the 1820's. By 1837, when Victoria came to the throne, the works of the remarkable preceding generation of poets and novelists – Scott, Byron, Coleridge, Keats, Shelley, Campbell, Southey, and Wordsworth – presented themselves for similar treatment. Many were dead, or had withdrawn from writing for publication, but their work had appeared during the transient interval of short copyright protection that was ushered in by the judicial implementation (in *Beckett v. Donaldson*, 1774) of the statutory copyright prescribed by the Act of Anne 8 (1709).

In this way the literature of the Romantic period serendipitously emerged from copyright to reach a greatly enlarged readership in innumerable cheap editions within only a generation of their having been written, whereas after 1841 in Britain, the span of copyright protection was lengthened to two, and then to three generations. It is intriguing, therefore, to speculate on what analogous effects may follow from the bulges and indentations that appear (Figure 13) in the “vintage population pyramid” of works that will be definitively in the public domain by 2025, and to consider the counterfactual situations that would have obtained at that date had late 20th century copyright legislation in the U.S. taken a different course.

Indeed, speculating on this effect on the readership of some important American authors from the period in question may be enlightening. Consider, for example, the protection given to Sinclair Lewis' *Babbitt* (published in 1922), which entered the public domain in 1997, compared with some of his other prominent works, such as *Arrowsmith* (published in 1925) and *Elmer Gantry* (published in 1927), which, being published only a few years after *Babbitt*, will not (unambiguously) enter the public domain until 2020 and 2022, respectively. All three works received 19 years extra duration from the 1962 Act, but only the latter two received an extra 20 years of protection from the CTEA. One can speculate that over the next decade, *Babbitt* will maintain its place as an American classic, while the later works of Lewis – especially the less heralded ones – will lose readership to earlier works of other authors. Similar examples abound – the later works of F. Scott Fitzgerald, such as *Tender is the Night* (published in 1934) have been given 20 more years of protection than some of his earlier works, potentially preventing these works from entering the echelon of *The Great Gatsby* in American fiction lore (though there is nothing preventing the Fitzgerald estate from placing his works in the public domain). Likewise, T.S. Eliot's *The Waste Land* (published in 1922) is likely to maintain its status as one of the finest pieces of American poetry – it entered the public domain in 1997 – while the legacy of his later works is more dependent on English teachers and professors continuing to assign his copyright-protected works.

The “popular canonization” of literary works may be particularly sensitive, as St. Clair's (2004) work has suggested, to the shifts in differential availability created by such “accidents” of timing in the intersection of the chronologies of publications and the copyright statutes, and royalty free Internet access may well compound this effect. But in other domains of publishing such as historical writings, and the sciences, where there is a strong cumulative character to the progress of knowledge, the prospects for analogous quirks of timing to influence popular learning and culture seem more circumscribed. Indeed, there the growth of “open access” to Wikipedia and other online compendia that are being continuously updated makes it unlikely that high-schoolers' and college students' essays – however good or bad they will be in other respects – will not be marred by excessive recourse to books that can be conveniently and

cheaply accessed without a trip to the library, even though, or because, they are seriously out-of-date.

6. CONCLUDING DISCUSSION: THE QUANTITATIVE FINDINGS AND THEIR BROADER SIGNIFICANCE

The contrast between our findings and the earnest hopes of Google’s visionary founders – and the expectations expressed by Stanford’s Librarian Michael Keller that, within two decades from 2005 “most of the world’s knowledge will be digitized and available, one hopes for free reading on the Internet” – could hardly be starker. Here we have an emblematic instance of the ironic situation created by two powerful and opposing historical trends, a conflict that has grown more obtrusive during the past two decades. The technical possibilities of accessing enormous resources of cultural and scientific information have been vastly expanded by spectacular advances in digital information and computer-mediated telecommunication technologies. But, at the same time, and not entirely due to independent developments, the evolving course of copyright legislation in the U.S. has greatly lengthened the duration over which access to protected content can be impeded by the terms that copyright owners may exact when licensing its reproduction and use.

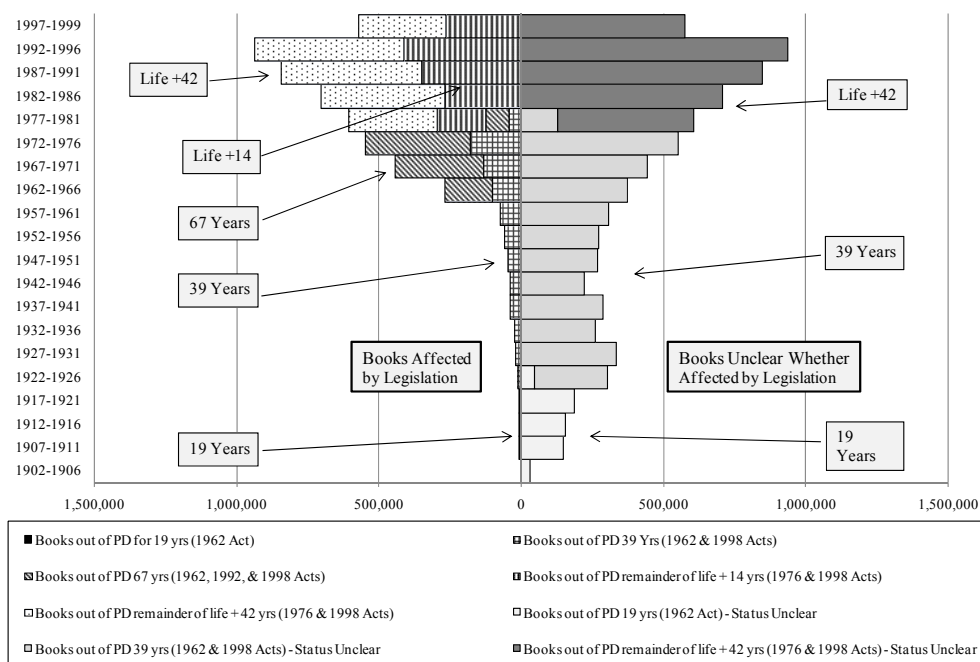
The quantitative estimates presented also shed light on the differentially deleterious effects of the specific statutory changes during the twentieth century that have worked to lengthen U.S. copyright term duration. We find that by 2027, changes in copyright laws over the last half-century will have prevented over 3.5 million books that would otherwise have entered the public domain from doing so. Of the four major laws in question, the two most responsible for this phenomenon are the 1976 and 1992 Acts, both of which eliminated the copyright renewal requirement. These findings provide empirical support for the claims of copyright law specialists such as Lessig and Goldstein, who argue that the renewal requirement which Congress discarded provides an important safeguard for an ample and growing public domain. On the other hand, the 1962 and 1998 Acts, by extending the renewal period, have increased the total amount of books not *clearly* in the public domain. By 2027, these two laws will have prevented over 2.3 million books from clearly being in the public domain – and thus available for digitalization by Google and the OCA without the possibility of copyright infringement.

Our analysis also raises a number of questions that are worth noticing even though they cannot be addressed properly within the scope of the present paper. First, there is the matter of the light that an examination of the outcomes casts upon the way that the process of legislation affecting copyright is viewed. Indeed, although the more recent controversy over the CTEA’s retroactive extension of copyright term garnered some media attention, it is clear that this feature of the 1998 Act set no new precedent. Congress had been enacting similarly retroactive measures since 1962, but this seems to have passed without notice; or at least without comment from the prominent economists who contributed an *amicus curiae* brief in the *Eldred v. Ashcroft* case. The latter sided with the counsel for the plaintiff, Lawrence Lessig (2004) in viewing the retroactive feature of the CTEA as a legislative departure that was impossible to justify on the “incentives” interpretation of the Constitution framers’ intent. Moreover, to the best of our knowledge, none of the CTEA’s critics remarked on the fact that its provisions actually compounded the effects of previous retroactive grants of protection to copyright owners: books registered between 1923 and 1941, for example, which had been “saved” once from falling into the public domain by the 1962 Copyright Act were granted 20 additional years of protection under the CTEA. This is only one instance of the more general tendency to consider the consequences of legislative alternations in

the copyright laws in isolation, rather than to evaluate the ways in which each proposed statute would interact with the provisions of prior statutes.

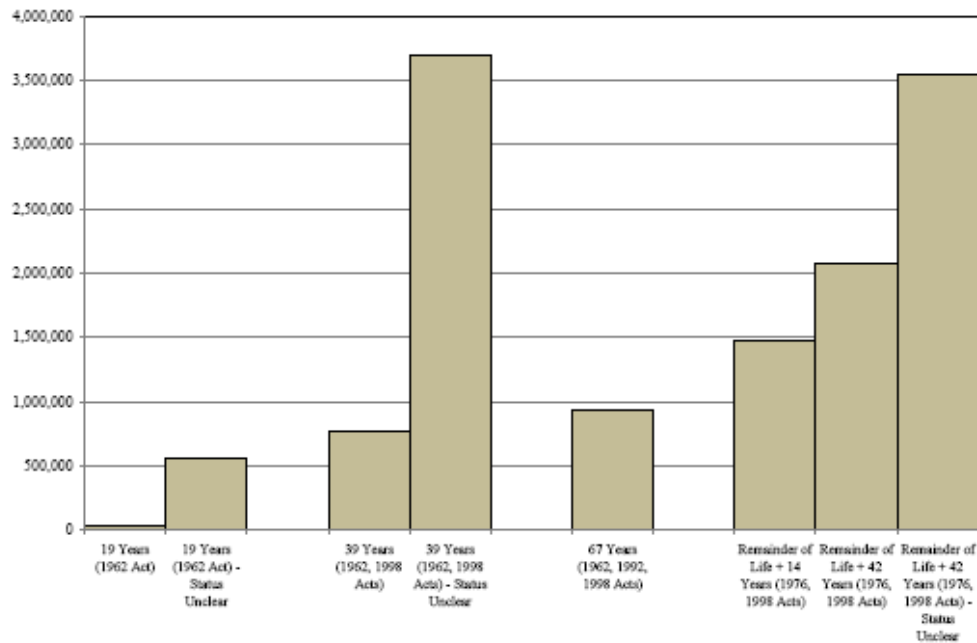
To sharpen this point and assess its quantitative significance, we can examine “legislative interaction effects” by undertaking counterfactual calculations that compare the impact of the 1998 legislative change conditional on the 1962 Act being in force, with its effects in the absence of the 1962 Act (and its immediate legislative sequelae). Figures 15 and 16 present the results of employing the cohort vintage approach to make such a calculation. Allowing for the difficulty of determining copyright status (and the consequent aversion to scanning them for online distribution), these figures indicated that potentially as many as 3.6 million books that had been registered between 1923 and 1977 were given an initial 19 additional years of “de facto protection” by the 1962 Act, and were thus eligible under the 1998 Act for a further 20 years of grace before being returned to the public domain. Almost one million such titles can be said definitively have been held out of the public domain in this way for an extra 39 years.

FIGURE 15: NUMBERS OF BOOKS KEPT OUT OF PUBLIC DOMAIN BY 20TH C. LEGISLATION FOR SUCCESSIVE PUBLICATION-COHORTS, SHOWING THE EFFECTS OF THE OPERANT ACTS ON THE DURATIONS OF COPYRIGHT PROTECTION EACH 5-YEAR “VINTAGE” RECEIVED



Economists say that "sunk costs are sunk," and so we might eschew dwelling on the results of actions that were taken in the past, finding there to be little point in critiquing past policy measures if we are not going to propose ways to remedy them by some new legislation. Nevertheless, a point to be made in that regard is that by becoming more aware of the potentialities for currently contemplated legal enactments to interact in surprising ways with pre-existing statutes, the designers of legislative “reforms” should be more able to mitigate the

FIGURE 16: TOTAL NUMBER OF BOOKS REGISTERED DURING 1902-1999 WHOSE COPYRIGHT LIVES WERE EXTENDED BY LEGISLATION, GROUPED BY ADDITIONAL NUMBER OF YEARS HELD OUT OF THE PUBLIC DOMAIN UNDER THE INDICATED STATUTES



unintended and undesired consequences of interactions between the new and the old bodies of law. It would therefore seem both feasible and desirable to go further than awareness, by setting some precautionary conventions: for example, a rule of practice in Congressional mark-ups of copyright legislation might be established that would require protected works to be automatically excluded from benefiting from “legislative grandfathering” on more occasion.

Of course, such a customary rule would not be likely to stand against pressures to grant longer retroactive gifts to copyright owners, and this thought raises the question of whether the history of successive twentieth century modifications in the U.S. statutes is to be interpreted as the product of disconnected episodes of Congressional inattention to the implications of proposals that sought now to achieve administrative simplifications for the Copyright Office, now to bring U.S. statutes into alignment with international copyright conventions, and occasionally to satisfy some iconic commercial interest – like the preservation of ownership control over the uses of Mickey Mouse. Alternatively, the twentieth century trend toward longer and longer durations of copyright protection might be read, not as the happenstance outcome of independent legislative occurrences, but as the result of a systematic process of legislative accommodation to business lobbying motivated by a powerful trend in information technology that has been working episodically but cumulatively to raise the economic benefits of longer spans of copyright monopoly. Such a view would cast the recent spate of legislation as just the most latest in a succession of institutional accommodations to “punctuated” changes in the techniques of printing, reprographics and distribution of “content” that can be identified from Gutenberg onwards, changes that have both enlarged the industrial domain within which copyright protection assumed economic importance and (since the later part of the 19th century)

dramatically increasing the value of extended copyright duration to publishers of books, recorded music and images.²²

The point suggested here turns on the question that the economic implications of modifying copyright statutes might have been viewed by legislators and public commentators within a traditional framework of industrial interests that paid little regard to contemporary advances in the technology of computer-mediated telecommunications that were visible, but whose implications were not articulated in political circles before they became evident in the “Internet tidal wave.” Whatever the construction that may be placed upon the “causes” of the course of copyright legislation (a matter into which we cannot enter on this occasion), the quantitative assessment carried out here has established the point that the consequences are far from negligible. The volume of material whose return to the public domain has been legislatively postponed for many years really is quite staggering, especially so if one imagines that it issued from the concatenation of unrelated impulses to effect institutional “reforms” – a blind sequence in which at each step the implied future curtailment of public domain conditions of access remained unseen, or was dismissed as too speculative to deserve consideration.

Should we judge the seriousness of the legislative history’s effects anachronistically, viewing them in hindsight framed by the Internet’s emergence? Would the magnitude of the material withheld from the public domain have mattered so much in a world without the digital information revolution? Most of the books under consideration – those with little to no commercial value after 28 years – are primarily of interest only to researchers, and the books in question would still have been available in academic libraries and thus accessible to scholars regardless of copyright status. If legislator’s had such considerations in mind and thought that the injury to the general reading public from keeping such works out of the public domain a little longer would not have been significant, it would be difficult to rationalize the benefit to the publishers of offering the incentives of longer protection to the mass of works with such limited commercial lives. But the situation has been transformed by the revolutionary advances in digital information technology, and while recognition of the difficulty of foretelling those changes and their implications from the vantage points of 1962, or 1976, or even 1992 should temper the severity of critical judgments about the legislators who modified the copyright statutes at those points in the past, it is not inappropriate to reckon what now can be seen to be the consequences of those actions.

Today, the greater capacity for the dissemination of knowledge, for cultural creativity and for scientific research carried out by means of the enhanced facilities of computer-mediated telecommunication networks, has greatly raised the marginal social losses that are attributable to the restrictions that those adjustments in the copyright law have placed upon the domain of information search and exploitation. But that is not the only consequences that should be weighed. The implications of the particular ways in which the public domain has been curtailed are harder to discern at this point, but these too deserve notice. Under the current regime, the Web-browsing public will be given virtually free and convenient online access to an accumulated mass of publications that can be and will be scanned for distribution in digitalized

²² For more, see David (2004a). The key differences between the two views that the text poses of the legislative process is really a question of whether those long-term tendencies and their implications for the effects of copyright on different economic interests (consumers, researchers, authors, and publishers) were understood by industry lobbyists but never grasped by legislators, or whether the trade-off issues were symmetrically understood but the legislative stance systematically placed accommodating the needs of the publishing industries above all other considerations. Obviously, these issues of political economy cannot be pursued here.

form because they can be presumed to be no longer in copyright, thereby avoiding the more time-consuming and costly procedures of having to search for and obtain copyright permissions from material that might be in the public domain, but might turn out not to be, and, worse, not to have an identifiable copyright owner.

The obvious irony is that material that is more out-of-date will be most readily accessible in the new environment, whereas under the previous technological regime it would have been locked away in dusty library stacks. While there is no classification of "content obsolescence" that permits precise assessment of the magnitude of this problem, and the effects of differential access costs on scholarship and popular culture have been seen to be potentially quite complicated and difficult to foresee, it is entirely feasible to settle the question of how far into the future the differential accessibility afforded to such works will persist, and the magnitude and time-distribution of the problem.²³ Indeed, it would seem evident that taking seriously an "options" approach to modifications of copyright protections should be incumbent upon the legislators and the judiciary in interpreting how the powers assigned to Congress under Article 1 of the Constitution are to be used. In other words, recognition of the continuing rapidity with which digital information technologies are advancing – providing new and more powerful search facilities in both the domains of consumer enjoyment of cultural products and the scientific/technical inquiry – should systematically weigh *the likely future opportunities* to promote the progress of science and the useful arts that would be effectively diminished by the proposed modifications.

Our findings also have some bearing on the issue of whether the research libraries involved ought to take it upon themselves to select which items among their holdings should be given priority in the scanning process, deferring others until sufficiently "less obsolescent" alternatives become available. This concern was raised by one of the authors in a brief memo to Paul N. Courant, President of the University of Michigan, and to the Stanford University Librarian, Michael Keller:

[T]he extension of network access increases the scope for network externalities. Among the latter some undoubted are good, but others less so. In particular, through programmed mimetic action, network externalities create positive feedback that can greatly reduce the diversity of accessed information. The effect of Google and other search engines in amplifying concentrations of "hits" on the more popular URLs, is to reduce the relative "search costs" of locating a subset of documents that need not be most informative – even when the user has entered the most appropriate search terms. Since we know that students (at high schools as well as universities) increasingly conduct their research without physically entering a library, and without accessing an on-line catalogue, this is worrisome ... In raising the foregoing points, I am not suggesting that Michigan and Stanford (let alone Oxford) should be held responsible for the emergent properties of the Google implementation of the virtual library concept. But if these great institutions and their librarians ... are not thinking about how to meliorate the

²³ Moreover, as Kelly (2006) points out, the advent of properties such as tags and links allows for digitalized books to be interconnected with all areas of prior human thought in ways unimaginable a mere twenty years ago. These properties provide positive value to millions of works that had long been discarded by their publishers as valueless – the readership of such marginal works (no matter how esoteric) certainly increases when they are digitalized and available via Internet search.

effect of replacing catalogues formed by many decades of scholarly expertise with one that is dynamically re-created by the continuous sampling of inexperienced readers' search behaviors, who will? (David, 2004b)

Gaining a more precise understanding of the impacts of intellectual property rights upon the advancement of and access to the diverse forms of knowledge shared by human cultures should be seen as the larger purpose that the research reported here is meant to serve. Apart from the intrinsic interest of that big and complicated question, deepening our understanding of the variety of particular instances in which it is encountered is the most feasible approach to constructing an empirical basis for rational and pragmatic discussion of appropriate social policy responses to the problems of mutual adaptation of technological and institutional infrastructures. The statutory copyright regime, along with other institutional structures that impinge upon the production and distribution of cultural and scientific information, is being reshaped, but not necessarily in ways that will enable society to benefit most fully from the technical capacities afforded by enhanced telecommunication network infrastructures and networked digital information applications.

APPENDIX: FORMULAS FOR COPYRIGHTS ENTERING THE PUBLIC DOMAIN

In this section, we estimate the number of books in the public domain as well as four counterfactuals for copyrights registered between 1902 and 1999.

TC_i = Total Copyrights Entering Public Domain in year i

RG_i = Total Registrations in year i

RN_i = Total Renewals in year i

1) Public Domain as it Exists

1930-1957: $TC_i = RG_{i-28} - RN_i$

1958-1961: $TC_i = RG_{i-28} - RN_i + RN_{i-28}$

1962-1980: $TC_i = RG_{i-28} - RN_i$

1981-1991: $TC_i = RG_{i-28} - RN_i + RN_{i-47}$

1992-1997: $TC_i = RN_{i-47}$

1998-2017: $TC_i = 0$

2018-2027: $TC_i = RN_{i-67}$

2) No Bono Act

1930-1957: $TC_i = RG_{i-28} - RN_i$

1958-1961: $TC_i = RG_{i-28} - RN_i + RN_{i-28}$

1962-1980: $TC_i = RG_{i-28} - RN_i$

1981-1991: $TC_i = RG_{i-28} - RN_i + RN_{i-47}$

1992-2027: $TC_i = RN_{i-47}$

3) No 1992 nor Bono Acts

1930-1957: $TC_i = RG_{i-28} - RN_i$

1958-1961: $TC_i = RG_{i-28} - RN_i + RN_{i-28}$

1962-1980: $TC_i = RG_{i-28} - RN_i$

1981-2005: $TC_i = RG_{i-28} - RN_i + RN_{i-47}$

2006-2027: $TC_i = RN_{i-47}$

4) No 1976, 1992, nor Bono Acts

1930-1957: $TC_i = RG_{i-28} - RN_i$

1958-1961: $TC_i = RG_{i-28} - RN_i + RN_{i-28}$

1962-1980: $TC_i = RG_{i-28} - RN_i$

1981-2027: $TC_i = RG_{i-28} - RN_i + RN_{i-47}$

5) No 1962, 1976, 1992, nor Bono Acts

1930-1957: $TC_i = RG_{i-28} - RN_i$

1958-2027: $TC_i = RG_{i-28} - RN_i + RN_{i-28}$

BIBLIOGRAPHY

- Authors Guild. 2006. "The Battle over Books: Authors and Publishers Take on the Google Print Library Project," (The Celeste Bartos Forum, held at The New York Public Library, November 17, 2005. Edited for the Authors Guild Symposium) *Authors Guild Bulletin*, Winter.
- Band, Jonathan. 2006. "The Google Library Project: Both Sides of the Story," 1 *Plagiarism: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification* 1-17.
- David, Paul A. 2004a. "The End of Copyright History?," 1 *Review of Economic Research on Copyright Issues* 5-10.
- David, Paul A. 2004b. E-mail correspondence, sent to Paul N. Courant and Michael Keller, December 18, 2004.
- Edwards, Elizabeth. 2005. "The Google deal (down on the Farm)," *Confessions of a Mad Librarian – Digital Issues Archive Blog*, January 07, 2005. Available at: http://edwards.orcas.net/blog/archives/2005_01.html.
- Goldstein, Paul. 2003. *Copyright's Highway: From Gutenberg to the Celestial Jukebox, Revised Edition*. Stanford, CA: Stanford University Press.
- Hafner, Katie. 2005a. "In Challenge to Google, Yahoo Will Scan Books," *New York Times*, October 3, 2005. Available at: <http://www.nytimes.com/2005/10/03/business/03yahoo.html>.
- Hafner, Katie. 2005b. "Microsoft to Offer Online Book-Content Searches," *New York Times*, October 26, 2005. Available at: <http://www.nytimes.com/2005/10/26/technology/26book.html>.
- Hafner, Katie. 2007. "Libraries Shun Deals to Place Books on Web," *New York Times*, October 22, 2007. Available at: <http://query.nytimes.com/gst/fullpage.html?res=9A0DE2D9123AF931A15753C1A9619C8B63>.
- Hanratty, Elizabeth. 2005. "Google Library: Beyond Fair Use?," 10 *Duke Law & Technology Review*.
- Kelly, Kevin. 2006. "Scan This Book!," *New York Times*, May 14, 2006. Available at: <http://www.nytimes.com/2006/05/14/magazine/14publishing.html>.
- Landes, William M. and Richard A. Posner. 2003. "Indefinitely Renewable Copyright," 70 *University of Chicago Law Review* 471-518.
- Lessig, Lawrence. 2004. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin Press.

Lessig, Lawrence. 2005. "Let a Thousand Googles Bloom; Copyright Reform is Vital to the Spread of Culture and Information," *L.A. Times*, January 12, 2005.

Markoff, John and Edward Wyatt. 2004. "Google Is Adding Major Libraries to Its Database," *New York Times*, December 14, 2004. Available at: <http://donswaim.com/nytimes.google.html>.

Sprigman, Christopher. 2002. "The Mouse that Ate the Public Domain: Disney, The Copyright Term Extension Act, and *Eldred v. Ashcroft*," *FindLaw's Legal Commentary*, March 5, 2002. Available at: http://writ.news.findlaw.com/commentary/20020305_sprigman.html.

Sprigman, Christopher. 2004. "Reform(aliz)ing Copyright," *57 Stanford Law Review* 485-568.

St. Clair, William. 2004. *The Reading Nation in the Romantic Period*. Cambridge, UK: Cambridge University Press.

Subcommittee on Patents, Copyrights, and Trademarks of the Committee on the Judiciary. 1991. *Hearing on S. 756, A Bill to Amend Title 17, United States Code, The Copyright Renewal Provisions, and for Other Purposes*. Washington, DC: U.S. GPO.

Toobin, Jeffrey. 2007. "Google's Moon Shot: The Quest for the Universal Library," *The New Yorker*, February 5, 2007. Available at: http://www.newyorker.com/reporting/2007/02/05/070205fa_fact_toobin.

Varian, Hal R. 2006. "The Google Library Project," Technical Report, UC-Berkeley.

Wyatt, Edward. 2005a. "Google Library Database is Delayed," *New York Times*, August 13, 2005. Available at <http://query.nytimes.com/gst/fullpage.html?res=9A0CE5DC153EF930A2575BC0A9639C8B63>.

Wyatt, Edward. 2005b. "Writers Sue Google, Accusing It of Copyright Violation," *New York Times*, September 21, 2005. Available at <http://www.nytimes.com/2005/09/21/technology/21book.html>.