

BESTSELLING MUSICAL COMPOSITIONS (1913-32) AND THEIR USE IN CINEMA (1968-2007)

PAUL J. HEALD

ABSTRACT. Some economists assert that as valuable works transition from copyrighted status and fall into the public domain they will be underexploited and their value dissipated. Others insist instead that without an owner to control their use, valuable public domain works will be overexploited or otherwise debased. This study of the most valuable musical compositions from 1913-32 demonstrates that neither hypothesis is true as it applies to the exploitation of songs in movies from 1968-2007. When compositions fall into the public domain, they are just as likely to be exploited in movies, suggesting no under-exploitation. And the rate of exploitation of these public domain songs is no greater than that of copyrighted songs, indicating no congestion externality. The absence of market failure is likely due to producer and consumer self-regulation.

1. INTRODUCTION

A growing group of commentators assert that the public may suffer when valuable copyrighted works fall into the public domain. One concern is under-exploitation, the possibility that a work without an owner will not be adequately distributed or otherwise made available to the public. According to Landes and Posner (2003), “[A]n absence of copyright protection for intangible works may lead to inefficiencies because . . . of impaired incentives to invest in maintaining and exploiting these

Thanks to Liz Wheeler for her work as my research assistant, to Doug Baird, Barton Beebe, Justin Hughes, Jay Kesan, Stan Liebowitz, Kirk Willis, and to participants at workshops at the University of Chicago, Case Western University, Chicago-Kent Law School, Bournemouth University, Birkbeck College, the Conference on Empirical Legal Studies, and the World Intellectual Property Organization, Geneva. Thanks also to Kun Xu and Professor Jaxk Reeves of the Statistics Department of the University of Georgia who prepared the statistical analysis in the Appendix. Another version of this paper will appear in volume 60 of the Case Western Reserve University Law Review. This version appears with its permission.

works.” Congress,¹ the courts,² and the Copyright Office³ all cited this concern in support of recent copyright term extension legislation.⁴ As to popular novels, at least, worries of under-exploitation appear to be unfounded. A recent empirical study (see Heald, 2008) of bestselling fiction from 1913-32 demonstrates that from 1988-2001, famous public domain novels were as available as their copyrighted counterparts.

A different, and until now empirically untested, claim asserts that a popular work falling into the public domain may be overexploited, “overgrazed” to use the terminology found in the tragedy-of-the-commons literature (e.g. Hardin, 1968). Landes and Posner (2003, p. 487) assert that the value of “a novel or a movie or a comic book character or a piece of music or a painting” could be depleted in much the same way as “unlimited drilling from a common pool of oil or gas would deplete the pool prematurely.” Others suggest that the value of ownerless works could be dissipated through excessive or inappropriate uses (see, for example, Liebowitz and Margolis 2004, Fennell 2004, Green 2003, Kozinski 1994, and Hughes 1999). In response, Mark Lemley (Lemley 2004) has argued that “this justification for intellectual property depends on proof that there is in fact a tragedy of the commons for information.” Since proponents of the under- and over-exploitation theories have done little testing of their hypotheses, the present study fills a significant gap in the literature.

Lemley (2004, pp. 129-31) identifies both the under-exploitation and the over-exploitation arguments as “ex post” justifications for protecting works in that they provide a rationale for extending protection without reference to “ex ante” incentives to create. The ex post justifications outlined above stand in the forefront of the world-wide debate over whether copyright terms for existing works should be retroactively extended.⁵ Because the standard incentive-to-create rationale cannot

¹Congress found in 1998 that retroactive extension of protection to existing works nonetheless “would provide copyright owners generally with the incentive to restore older works and further disseminate them to the public.” H.R. Rep. No. 105-452, p. 4 (1998).

²*Eldred v. Ashcroft*, 537 U.S. 186, 207 (2002) (Congress “rationally credited projections that longer terms would encourage copyright holders to invest in the restoration and public distribution of their works”); *Eldred v. Reno*, 239 F.2d 372, 379 (D.C. Cir. 2001), *aff’d as Eldred v. Ashcroft*, 537 U.S. 186 (2003) (arguing that works falling into the public domain will be less available to the public).

³See Copyright Term, Film Labeling, and Film Preservation Legislation: Hearings on HR 989, HR 1248 and HR 1734 before the Subcommittee on Courts and Intellectual Property of the House Committee on the Judiciary, 104th Cong. 1st Sess 50, 161, 171, 188 (statement of Marybeth Peters, Register of Copyrights) (arguing that publishers will not risk investing in a work that they do not own and therefore term extension is needed to assure availability of works).

⁴See 17 U.S.C. § 302(a) (1994 & Supp. IV 1998).

⁵See, for example, Gowers Review of Intellectual Property at http://www.hm-treasury.gov.uk/independent_reviews/gowers_review_intellectual_property/gowersreview_index.cfm This study, commissioned by the British Treasury department, rejects ex post justifications for extending copyright protection for existing works. Japan is also considering retroactive

justify extending the term of protection for a work that already exists, ex post justifications are also likely to drive the debate over further extensions in the U.S. when the present 20-year extension runs out in 2018. Claims of inefficient exploitation of public domain works have already been relied upon heavily by the successful apologists for the 1998 Copyright Term Extension Act.

Neither the over- nor under-exploitation theories have gone unchallenged. Lemley (2003, p. 138) scoffs at under-exploitation worries, stating that the claim “that control by a single firm is necessary to induce efficient production [is] theoretically unsound” and wondering why there is “some greater need to subsidize [by granting exclusive rights] the making of more copies of *Ulysses* than the making of more paper clips.” Amicus briefs, including one signed by five Nobel Laureate economists (Akerlof et al. 2002), rejected the under-exploitation argument when it was made in *Eldred v. Ashcroft*, and my own empirical work (Heald, 2008) concludes that popular books falling into the public domain are not underexploited in comparison to their copyrighted counterparts. The over-exploitation theory has also come under attack (see Bradford, 2005). Richard Epstein is a doubter, suggesting that “[a]nyone is hard pressed to believe that Shakespeare’s star has been dimmed by the calamities committed in his name . . .” (see Epstein, 2005). So too Lemley (2003) and Dennis Karjala (2006), both of whom deploy market-based economic arguments to allay fears of a congestion externality caused by overuse of copyrighted works. They conclude that “a belief that the original creator (or his transferee) can best manage the work in the public interest runs strongly contrary to our long-standing and fundamental reliance on free markets to allocate resources to the production and distribution of goods.” (Karjala, 2006, p. 1079).

Although the theoretical arguments on both sides are interesting, commentators have so far assumed (but not necessarily believed) that works falling into the public domain will be exploited at a different rate than their copyrighted counterparts. Exploitation rates are, of course, observable and ripe for empirical analysis. In Part I of the article, I explain the methodology of my study of popular musical compositions from 1913-32 as they appear in movies from 1968-2007. The study tracks songs from 1913-22 as they fall into the public domain and compares changes in exploitation rates with songs from 1923-32 that are still protected by copyright.⁶

term extension. See <http://williampatry.blogspot.com/2007/11/japan-looks-at-copyright-term-extension.html>. The author recently presented his findings on the exploitation of bestselling public domain fiction in Tokyo. See <http://www.21coe-win-cls.org/english/info/index.html>

⁶The United States did not move to a life of the author regime until the 1980’s. For the works dealt with in this paper (works published from 1908-22), the effective term of protection is 75 years.

Studying musical compositions has several advantages over my prior study of best-selling books. First, tracking the appearance of compositions in movies provides data on the exploitation of derivative works. Musical compositions usually appear in movies as works realized by someone other than the copyright owner. In a movie we hear a recording of the composition, a derivative work under the Copyright Act. Since those worried about over-exploitation inevitably warn against unauthorized derivative works as their most serious potential concern, the study provides especially relevant data. Second, relying on the appearance of musical compositions in movies provides an alternative, and possibly superior, measure of availability than the counting of book editions and book publishers in my prior study.⁷ Therefore, the present study's finding of no under-exploitation is not merely duplicative.

In Part II, the results of the study are reported: The rate of exploitation of public domain songs is not statistically significantly different than the rate for copyrighted songs, indicating that in this context worries of both over- and under-exploitation are misplaced. Part III joins the theoretical debate and suggests why self-regulation by both producers and consumers of copyrighted works explains the lack of market failure. Two novel tests are offered to predict unusual cases when over- or under-exploitation might be legitimate concerns.

2. METHODOLOGY

Previous studies confirm that most copyrighted works do not hold their value over time. Landes and Posner (2003, p. 473) note that “fewer than 11 percent of the copyrights registered between 1883 and 1964 were renewed at the end of their 28-year term, even though the cost of renewal was small.” They point out that of 10,027 books published in the U.S. in 1930, only 1.7 percent remained in print in 2001. An amicus brief in *Eldred v. Ashcroft* put the figure for books published in 1930 even lower, at 1.3 percent.⁸ Even those worried about what happens when works fall into the public domain agree that there is little reason to extend copyright protection to works with no current value (Landes and Posner, 2003, p. 474). In fact, extending copyright for those works would entail significant tracing and transaction costs and would almost certainly be inefficient. Given this consensus, the present study identified the 1294 most popular musical compositions

⁷The study measured availability of public domain books listed in Books In Print and tracked the number of editions and publishers. These figures were used as proxy for more direct measures, such as sales figures (which are usually proprietary) or a nationwide survey of availability in book stores. See Heald (2008, p. 1040).

⁸See Brief of Amici Curiae The Internet Archive, Prelinger Archives, and Project Gutenberg Literary Archive Foundation in Support of Petitioners, *Eldred v. Ashcroft*, No. 01-618, n. 10 (S Ct filed May 2002). Available on Westlaw at 2002 WL 1059714. The brief reports that 180 books out of 13,470 published in 1930 were “currently available for sale.”

from 1913-32 and focuses on the 74 most enduringly valuable of those compositions as they appeared in movies from 1968-2007. The years 1968-2007 were chosen because the compositions from 1913-22 began to fall into the public domain in 1988, the mid-point in that timeline.

Compositions from 1913-1932 were chosen because the works published from 1913-22 are all in the public domain and properly renewed works published from 1923-1932 are all still protected by copyright as a result of the 1998 Copyright Term Extension Act,⁹ allowing for a basically symmetrical comparison of ten years' worth of works from each group. Until extension, the effective copyright term for these works was 75 years, so works from 1913 fell into the public domain in 1988, works from 1914 in fell into the public domain in 1989, and so on until the 1998 legislation ended the flow of works into the public domain (see Cohen, 2006).

Studying a group of works from approximately the same era provides the opportunity to compare what happened to works from 1913-22 after they fell into the public domain and to compare rates of exploitation with those works from 1923-32 that remained protected. The initial data set included 601 of the most popular compositions from 1913-22 and 693 of the most popular compositions from 1923-32, as listed in the most accepted compilation of popular historical musical compositions (see Mattfield, 1962). All of these songs were then tracked in the Internet Movie Database (www.imdb.com) movie soundtrack database, which contains comprehensive information on almost 380,000 movies.¹⁰ Since the present debate revolves around only those works that have substantial present value, the primary statistical analysis was performed on the 74 musical compositions that appeared in at least four movies from 1968-2007 (although the findings hold for the entire population of compositions).¹¹ Since current sales data or licensing information of historic compositions is mostly proprietary and unavailable, appearance in movies serves as a proxy for present popularity. Movie producers invest significant resources into choosing music for their soundtracks. Their goal is to please audiences. Observing their choices provides an objective and neutral indication of what historic music presently has value to consumers.

The full list of 1294 songs can be obtained from the author; the subset of 74 appears in the Appendix. A substantial majority of the compositions (44 out of 70) were published in the six-year period from 1926-31, indicating the significance of the golden age of Tin Pan Alley (see Furia, 1990 and Jason, 1988), an extraordinary time period which marked the publication of many enduringly familiar works like

⁹The extension only applied to works that had been properly renewed in their 28 year after publication under the 1909 Act.

¹⁰See http://www.imdb.com/database_statistics (listing 379,871 titles) (last visited on September 5, 2008).

¹¹See the Appendix for a full statistical analysis.

Bye Bye Blackbird, *Blue Skies* [*Smiling at Me*], *My Blue Heaven*, *Let's Do It* [*Let's Fall in Love*], *Let's Misbehave*, *When You're Smiling* [*The Whole World Smiles with You*], *Bolero*, *Happy Days Are Here Again*, *Singin' in the Rain*, *Stardust*, *Embraceable You*, *Georgia on My Mind*, *Get Happy*, *I've Got Rythym*, *Just a Gigolo*, and *Mood Indigo*.¹² During this time, Cole Porter, the Gershwin Brothers, Harold Arlen, Hoagy Carmichael, Duke Ellington, and many others were at the prime of their famous composing careers. Since only 15 of the compositions dated from the 1913-22 time period, four qualifying songs from 1909-12 augment that portion of the data.¹³

The public domain songs were tracked during the period they were protected by copyright law and then after they fell into the public domain, 75 years after publication. For example, *Danny Boy*¹⁴ was first published in 1913 and entered the public domain in 1988. So, its use in movies from 1968 through 1987 (twenty years) when it was protected by copyright was tracked separately from its use in movies from 1988 through 2007 (twenty years) when it was in the public domain. Compositions from 1914 were therefore tracked from 1968-1988 (twenty-one years) and then from 1989-2007 (nineteen years), and so on.

In order to make the graphic comparison seen in Figure 1, each year's worth of compositions from the public domain song set were matched with the corresponding year a decade later in the copyrighted song set. Compositions from 1913 were paired with 1923, 1914 were paired with 1924, and so on. For example, three songs from 1913 appeared in a total of four movies from 1968-1987 (a rate of 4/60), before the songs fell into the public domain. Those same three songs appeared in 20 movies from 1988-2007 (a rate of 20/60).¹⁵ Therefore, the single song in the data set of copyrighted songs from 1923 was also measured in the same time frame, counting its use in movies from 1968-1987 (denominated "period one") and then from 1988-2007 (denominated "period two"). The song *Bugle Call Rag* appeared in no movies

¹²Specifically, the artists, songs and dates concerned are the following: Mort Dixon and Ray Henderson, *Bye Bye Blackbird* (1926); Irving Berlin, *Blue Skies* (1927); George Whiting and Walter Donaldson, *My Blue Heaven* (1927); Cole Porter, *Let's Do It* (1928); Cole Porter, *Let's Misbehave* (1928); Mark Fisher, Joe Goodwin, and Larry Shay, *When You're Smiling* (1928); Maurice Ravel, *Bolero* (1929); Jack Yellen and Milton Ager, *Happy Days Are Here Again* (1929); Arthur Reed and Nacio H. Brown, *Singin' in the Rain* (1929); Mitchell Parrish and Hoagy Carmichael, *Stardust* (1929); Ira Gershwin and George Gershwin, *Embraceable You* (1930); Stuart Gorrell and Hoagy Carmichael, *Georgia on My Mind* (1930); Ted Kohler and Harold Arlen, *Get Happy* (1930); Ira Gershwin and George Gershwin, *I've Got Rythym* (1930); Irving Caesar and Leonello Casucci, *Just a Gigolo* (1930); and Duke Ellington, Irving Mills, and Albany Bigard, *Mood Indigo* (1931).

¹³There are several that appeared in at least four movies from 1968-2007. They are: *By the Light of the Silvery Moon* (1909); *Let Me Call You Sweetheart* (1910); *Alexander's Ragtime Band* (1911); and *It's a Long Way to Tipperary* (1912).

¹⁴Fredrick Weatherly, *Danny Boy* (1915). See also http://en.wikipedia.org/wiki/Danny_Boy.

¹⁵The rate is 4/60 and 20/60 rather than 4/20 and 20/20 because each of the three songs was measured during a twenty-year time period, a total of sixty measurable song years (three songs × twenty years = sixty song years).

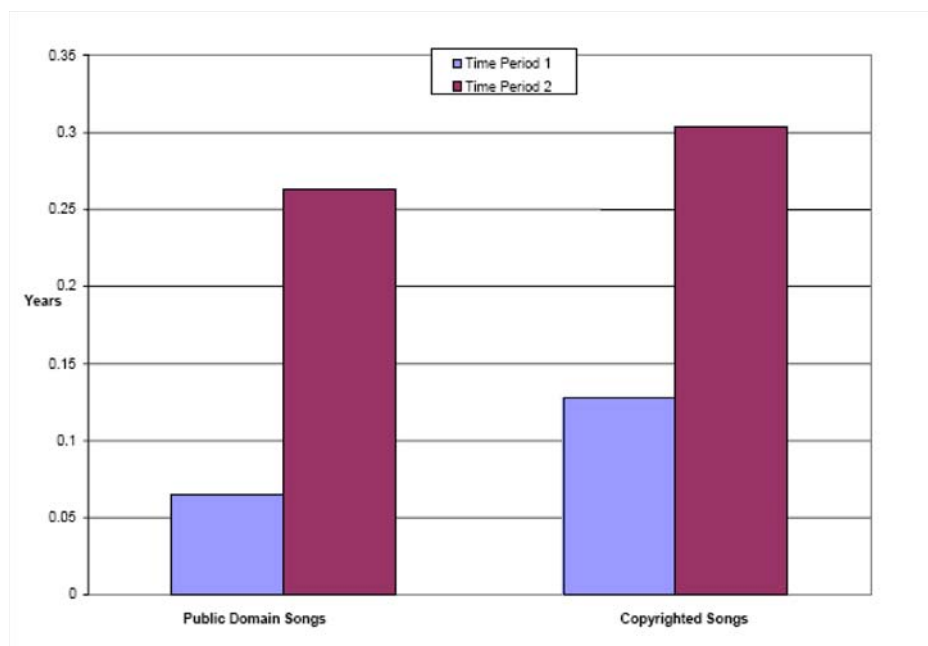


FIGURE 1. Average yearly use of a song in a movie

from 1968-87 (a rate of 0/20) and in four movies from 1988-2007 (rate of 4/20). For songs from 1914 and 1923, the relevant time periods for measuring uses in movies was 1968-1988 (period one) and 1989-2007 (period two); for songs from 1915 and 1925, from 1968-89 (period one) and 1990-2007 (period two), and so on.

The aggregate number of uses in movies of the 1913-22 songs during the period they were still under copyright was compared to the aggregate number of uses of the 1923-32 songs in time period one. Then, the aggregate number of uses in movies of the 1913-22 songs after they fell into the public domain was compared with the aggregate number of uses of the 1923-32 songs in time period two. This comparison allows for a straightforward explanation of the formal statistical regressions presented in the Appendix which employ a more robust, but less narratively engaging, methodology.

3. DATA ANALYSIS

The goal of the analysis was to answer two questions. First, when compositions from 1913-22 fell into the public domain were they exploited at a significantly different rate than while they were still protected by copyright? Second, if the rate of exploitation of public domain works increased after they fell into the public domain, did the change indicate signs of over-exploitation in comparison to the rate of exploitation of copyright works?

3.1. No Evidence of Under-Exploitation. Before the compositions from 1913-22 fell into the public domain, they appeared in movies on average at a rate of once every 15.3 years. After they fell into the public domain, they appeared in movies on average at a rate of once every 3.8 years. At first glance, this rate change appears to show a significant increase in exploitation, but the rate change must be compared to the rate of uses of copyrighted songs during the same time period. After all, all songs from this general era, regardless of their legal status, may be appearing more frequently in recent movies. This, in fact, appears to be the case. During the same comparative time periods, the rate at which copyrighted songs from 1923-32 appear in movies increased from once every 7.8 years in time period one to once every 3.3 years in time period two. Figure 1 shows the comparative increase in terms of average yearly use of a song in a movie, an increase for public domain songs from .065 uses per year to .263 uses per year and an increase for copyrighted songs of from .128 uses per year to .304.

Since the songs from 1913-22 fell into the public domain, they have been used on average four times more frequently in movies. The songs from 1923-32 also appear more frequently in movies over the same time period. The change, however, is more modest, an increase of a little less than two and one-half times as frequently. The formal statistical regressions in the Appendix, not surprisingly, demonstrate that the transition from protected work to unprotected work did not render public domain compositions under-exploited in relation to works that remained protected by copyright. Public domain songs from this era do not become orphans that are unavailable for public consumption.

This result is generally consistent with my prior study of bestselling fiction from the same period (Heald, 2008). That research compared the 166 bestselling novels from 1913-22 with the 167 bestselling novels from 1923-32 and found that from 1988-2001, novels in the public domain were in print at a rate insignificantly different from novels still under copyright. After 2001, however, the public domain novels were in print at a significantly higher rate, with significantly more editions per novel. In 2006, the in-print rate for the public domain novels was 98% as compared to 74% for the copyrighted novels. A comparison of the sub-sets of the twenty most enduringly popular novels generated similar results.

Although the music composition data show no evidence of under-exploitation, the study does not prove a positive public domain effect on availability, like that demonstrated for public domain books after 2001. A superficial comparison of the rate changes for music exploitation looks significant ($4\times$ as compared to $2.5\times$), but the logistic regressions performed in the Appendix expose the confounding effect of time as a variable and show that the comparative rates of exploitation of public domain and copyrighted music are not significantly different.

Why is there a positive public domain effect with books, but not with musical compositions as they appear in film? One difference may be that the study of best-selling fiction measured the availability of copies of an original work. The costs of scanning a book into a computer, printing it, and selling it are quite low; many Dover versions of bestselling classics sell for less than four dollars.¹⁶ If one chooses to publish a copyrighted book instead of a public domain book, the additional licensing cost will have a significant effect on the overall cost of production. On the other hand, the proportional cost savings of choosing a public domain song for a movie is likely to be much lower. Because a musical composition, whether it is protected by copyright or not, can only appear in a movie as a derivative work, the director of the film must either hire musicians or singers (or both) in order to realize a version of the composition, or she must obtain a license to use an existing recording of the composition. Creating the derivative work from “scratch” will likely entail significant costs, and the alternative of using an existing recording will likely entail the payment of a significant licensing fee to the owner of the recording. These costs will be incurred *even if* the underlying musical composition is in the public domain.¹⁷

Using a musical composition in a movie, therefore, is likely to be significantly more expensive than copying a book because it entails the creation of a new derivative work or the purchasing of a license to use one created by someone else. A film director can save some money by telling her musical director to choose only public domain compositions for the score, but the savings will be proportionally smaller than those enjoyed by the book publisher. Because of the marginal savings of choosing public domain music, it is not surprising that compositions are not exploited at a rate significantly exceeding that of protected music.¹⁸

3.2. No Evidence of Over-Exploitation. Two sorts of over-exploitation arguments have been offered by those who worry about what happens to works when they fall into the public domain. First, works may simply be overused and worn out, like a song we have heard so frequently we do not want to hear it again. Second, inappropriate uses, even if infrequent, may “recode” the original meaning of a work (see Hughes, 1999), debase it or otherwise make it less valuable to consumers. The examples most frequently given involve uses of copyrighted fictional characters in new pornographic works.¹⁹

¹⁶See amazon.com (advanced book search under “publisher/Dover” and “subject/literature and fiction”).

¹⁷Sound recordings of public domain compositions are independently protectable under 17 U.S.C. 106(7).

¹⁸They are exploited at a higher rate, but the difference is not significant. See the Appendix.

¹⁹See, for example, Liebowitz and Margolis 2004, pp. 5-6, where the authors speculate about porno tales involving Dr. Seuss’s character the Grinch.

3.2.1. *No Evidence of Worn-Out Songs.* As noted earlier each song in the public domain data set on average appears in a movie once every 3.8 years; each song in the copyrighted data set on average appears in a movie once every 3.3 years. The Appendix shows that these rates are statistically the same. This result makes it very difficult to argue that these songs need owners in order to prevent them from being worn out and devalued. If copyright owners are willing to license their compositions at a higher rate than public domain compositions are used, then the evidence against over-exploitation seems conclusive.

Even the most intense periods of usage of the public domain songs, *Danny Boy* (1913), with nine movie appearances between 1993 and 2001 and *After You've Gone* (1918), with nine movie appearances between 1996 and 2006, do not outstrip the periods of most intense usage for compositions protected by copyright. For example, in the 1930's, *Sweet Georgia Brown* (1925) appeared in 15 movies, *Am I Blue?* (1929) in 17 movies, and *Happy Days Are Here Again* (1929) in 34 movies. More recently, the Irving Berlin classic *Blues Skies* (1927) appeared in 10 movies from 1994-2004; *Stardust* (1929) appeared in 10 movies in the 1990's; and *Dream a Little Dream of Me* (1931) appeared in 10 movies from 1995-2005. Copyright owners seem to be willing to license their compositions at rates equal to or exceeding that of the most intensely used public domain compositions. When a song falls into the public domain, the data provide no evidence that it will be overexploited and worn out by moviemakers.

3.2.2. *Debased Works?* Even if a song is not subject to overly frequent use, some worry that a handful of "inappropriate" uses might debase the value of the original work, rendering it less desirable for consumption. If public domain songs have been subjected to damaging uses, therefore, one would expect them to be used less frequently in movies thereafter. After all, a rational film director would not want to alienate her audiences with a composition that had been previously debased. Evidence of debasement should show up in decreasing demand for public domain music over time as compared to copyrighted music from the same era. The data as a whole show no evidence of this, but the number of movie uses in any particular year is too small to measure accurately whether any particular public domain song has been damaged, damage that might be masked by its inclusion in the larger set of songs.

Evidence from my previous study of bestselling fiction, however, provides some interesting evidence on individual works. At Year 75 after publication, the twenty most enduring popular works from 1913-22 were in print at an average of 4.7 editions per title. At Year 80 after publication, the average is 9 editions per title, and at year 85 it rises to 13.4 editions per title. In the year 2006, an average of 26.6

editions per title are in print. The data demonstrate no evidence that pervasive inappropriate uses have reduced the attractiveness of the works for production and delivery to the public. The story is the same when one looks at the individual titles. Eighteen of the twenty titles were in print in more editions in Year 80 after publication than in Year 75.²⁰ All twenty experienced an increase from Year 80 after publication to Year 85, and all twenty experienced an increase in the number of available editions from Year 85 after publication to the year 2006. Moreover, the steepness of the upward sloping curve of editions exceeds that of copyrighted works from the same era over the same periods (see Heald, 2008, Figure 3). This is not to assert, of course, that there have been no shocking uses of either the songs or the books studied. As discussed below, producer and consumer self-regulation may explain why works are likely safe from even pornographic uses.

4. THE EFFICIENT EXPLOITATION DEBATE

Given the lack of empirical support, the persistence of claims that value is dissipated when works fall into the public domain seems curious. In this final section, I explore the paradigmatic examples of inefficient exploitation that have been offered and suggest a test to identify when problems might occur. Previous skeptics have argued that even if value is dissipated, we should not worry when it results from the natural interaction of market forces (Karjala, 2006).²¹ I explore below why value may be unlikely to be dissipated at all when works fall into the public domain.

4.1. Under-Exploitation. In my previous work, I identified three conditions that might justify extending copyright protection to an existing work in order to prevent its under-exploitation: 1) The cost of making the initial copy of a work available to the public is high; 2) the cost to free riders of making subsequent copies is low; and 3) the newly available work does not incorporate independently protectable material. The test had its genesis in arguments over whether old public domain

²⁰The exceptions are *Pollyana* (1913), by Eleanor Porter, which was published in 5 editions in year 75 after publication and only 4 in year 80, and *Scaramouche* (1921), by Raphael Sabatini, which was published in 5 editions in year 75 after publication and only in 3 editions in year 80. By 2006, *Pollyana* was available in 30 different editions and *Scaramouche* in 18.

²¹Karjala (2006, p. 1072) criticizes Landes and Posner, arguing that “A change in the demand curve for a work, however, while showing a change in how much society values a particular work relative to whatever else is available, says nothing about the total value to society of all the goods and services available.” Karjala notes that if the public’s taste for buggies shifts to cars then “[b]uggies are indeed less valuable, but society has incurred no economic loss.” Also, Lemley (2003, pp. 135-6) notes that competition changes consumption patterns with durable goods and should also with creative goods formerly protected by copyright – “Our normal supposition is that the invisible hand of the market will work by permitting different companies to compete with each other [to produce a good the public wants].” Compare this to Heald (2008, p. 1054) – “If we trust the market to produce the optimal amount of tangible goods like string, bubble gum, and diet soda without entrusting central control of those products to a single authority, why should we treat intangible public goods like *My Antonia*, the color yellow, or the word “coffee” any differently?”.

films needed owners in order to ensure their preservation and distribution (Lemley 2003, p. 134 and footnote 3). If an old film requires a significant expenditure to repair and yet could easily be copied and distributed without authorization once it is in digital form, the owner of the physical copy of the film may lack an adequate financial incentive to restore the film. The above test builds on this seemingly sensible intuition about a narrow category of works that might require owners to ensure their availability. Given the reality surrounding aging films, which may be more efficiently husbanded by non-owners,²² we should add a fourth proviso: 4) owners are in fact more willing than non-owners to preserve and distribute. This new fourth condition finds support in a recent study undertaken by the Library of Congress that shows non-owners have been making historic sound recordings available in digital form at a higher rate than their owners. For example, according to Brooks (2004), copyright owners have made only 14% of popular recording from 1890-1964 available on CD, while non-owners have made 22% of them available to the public on CD.

Where the four conditions are met, perhaps we should be worried, but it seems clear that they are generally not met with respect to the vast majority of books, music, films or computer programs and other works that are cheap and easy to produce. In general, the copyright term seems adequate if it is long enough to stimulate the creation of the work in the first instance. Extra extension, like that found in the Copyright Term Extension Act, is probably not justified except in a tiny fraction of cases. In the absence of these four conditions, we should not expect to see problems with under-exploitation when a work falls into the public domain.

Applying the test to musical compositions as they appear in movies helps explain why we see no under-exploitation with these works. As noted above, a musical composition as it appears in a movie is a derivative work that may be quite costly for the music director to use and thereby make available in a new form to the public. Unlike with the making a copy of a book, the first condition arguing in favor of ownership may often be met. Condition two is also probably met: if the movie is in a digital format, it will be quite easy to copy. Condition three, however, is not met, and songs in movies provide a nice example of the salience of that condition. A musical composition as it abides in a soundtrack is surrounded by independently protected work, like the script, the cinematography, and the sound recording itself, whose copyright is owned by its producer. The musical composition

²²See Lemley (2003, p. 137 and footnote 27). Lemley cites Mulligan and Schultz (2002); "According to the Internet Movie Database, 36,386 motion picture titles were released from 1927 to 1946. Of those, only 2,480 are currently available on videotape; only 871 are available on DVD; only 114 are available on Pay-Per-View/TV; and only thirteen are available in theaters." Lemley notes, "By contrast, just one archive – the Prelinger Archive – has over 27,000 public domain films and has put more than 1,100 online. See Rick Prelinger, Prelinger Archives, online at <http://www.prelinger.com>."

per se, the sheet music, cannot be easily extracted without offending the rights of copyright owners of neighboring works. The realization of the old public domain work within a new protected format means that the filmmaker has few real worries about competitors free riding its labor. In other words, the public domain status of the underlying musical composition should not pose a threat to its continued exploitation, precisely what the data analyzed above shows.

4.2. Overexploitation: Worn Out Works and Inappropriate Uses. Trademark law provides a nice example of how both sorts of overexploitation fears discussed in Part II become operationalized in law. One of the primary bases, for example, for the enactment of the Federal Trademark Anti-Dilution Act was the fear that unauthorized uses of a trademark would blur its ability to identify the source of its owner's goods or services. Even if a new "KODAK Café" or "EXXON Telephone" were of impeccable quality, Congress feared that a proliferation of uses would render marks like KODAK or EXXON less able to call to mind their original owners. Overuse might literally wear out the marks. I am currently collecting data on whether such unauthorized uses actually occurred prior to anti-dilution protection, but there is little doubt that the "wearing out" theory motivated Congress to pass the law in 1988 (Magliocca, 2001).

On the other hand, traditional trademark infringement provides a good example of how inappropriate uses can directly alter, as opposed to just wear out, the meaning of a symbol.²³ In fact, accountants routinely testify about the amount of pecuniary damage done to the value of a trademark when consumers are confused by an infringer (see Smith, 1997). If a garment maker sells shirts under the trademark "EXCELSIOR" and establishes a reputation for a high quality product, a subsequent user of the trademark on inferior goods will not only lower the trademark's value to the garment maker, but also make the word "EXCELSIOR" less usable to the public. Before the infringement, "EXCELSIOR" meant high quality shirts; afterwards it does not. If consumers are successfully confused by an infringer, then the public has been robbed of a valuable mnemonic device. The mark is debased.

Given the data presented in Part II, we need to ask why these two concerns do not seem to have the same traction in the context of copyrighted works.

4.2.1. *Worn Out Songs? Worn Out Anything?* As noted in Part II, public domain songs appear in movies at a slightly lower absolute rate per than their copyrighted counterparts. At least in the context of musical compositions in movies, there appears to be no chance that public domain songs are wearing out at a higher rate than their copyrighted counterparts. But what about songs as they are heard on

²³The cause of action for dilution via tarnishment of a mark's image is designed to protect a mark from altered meanings. See 35 U.S.C. § 1125(c).

the radio or in television advertising? Is it possible that public domain songs are being worn out via overexposure in non-movie media?

Landes and Posner (2003), and Liebowitz and Margolis (2004) recognize that congestion externalities usually are not thought to be a problem with works, like those typically protected by copyright law, which have the characteristics of non-rivalrousness and inexhaustibility. They understand that a song can be sung by one or two or one thousand people at the same time (demonstrating non-rivalrousness), over and over again, day after day, without wearing the song out (demonstrating inexhaustibility). Since the marginal cost imposed by each additional user is zero, limiting access would result in a deadweight loss. In fact, if one defines the value of a good in terms of its continued usability, then overuse is theoretically impossible with pure public goods. Landes and Posner, and Liebowitz and Margolis, however, argue that the relevant measure of value is market value, not usability, and therefore posit that certain sorts of marginal additional uses of a public good may impose positive costs. For example, if dozens of advertisers all chose the same song to market their products on television, the public might tire of the tune, and demand for it would drop, reducing its market value. We might, they speculate, see a musical version of the tragedy of the commons.

With songs, this eventuality seems unlikely. First, the vast majority of media airplay occurs through the broadcaster's acquisition of an ASCAP license. The standard license in no way restricts the number of times a song can be broadcast over any period of time.²⁴ In other words, copyright owners, acting through their primary agent, the American Society of Composers and Authors and Publishers, seem utterly uninterested in limiting the airplay of their compositions. Broadcasters, not copyright owners, determine how frequently the public should hear a song. Presumably, broadcasters voluntarily choose not to overplay a song for fear of alienating the public or reducing the value of a good they would like to offer in the future. Overplaying a musical composition, whether it is copyrighted or in the public domain, is bad business, a fact that copyright owners seem to recognize by not restraining broadcasters. In the broadcasting context, public domain songs seem no more likely to be worn out, therefore, than copyrighted songs. It seems specious, at least as to broadcasting, to argue that each song needs an owner to limit its use.

That leaves "background" music used in advertising, in films, and on television which is not licensed through ASCAP.²⁵ My data cast doubt on overuse of public

²⁴See http://www.ascap.com/licensing/radio/RMLC_License.pdf (standard license agreement for radio stations).

²⁵Licenses must be negotiated directly with the copyright owner. See <http://www.harryfox.com/public/hfaPurpose.jsp> where it states that the Harry Fox Agency does not "issue licenses for the use of music in advertising, movies, and TV programs (aka

domain music in movies, and over-exploitation seems unlikely in other contexts also. With a virtually infinite commons of music to choose from, advertisers are unlikely to risk alienating the public by choosing the same theme music as too many of their peers. Decades of watching television and listening to radio support this economic intuition.²⁶ The traditional tragedy of the commons analogy may be inadequate to capture the market for something like music in advertising.

To illustrate the tragedy of the commons, economists tell the story of a common field subject to overgrazing because no one owns it and therefore no one has the proper incentive to maximize its value. And, of course, empirical evidence shows an increase in agricultural production in England when common fields were enclosed (Ellickson, 1993). An advertising jingle presents a significantly different situation. Unlike the farmer who has limited options as to where to graze his cattle, the advertiser has thousands of songs to choose from. A farmer with a thousand choices of equally cheap and desirable fields on which to graze his cattle would rationally choose not to overgraze any particular one. It would be pointless and might cost him in the future. Overgrazing in the presence of numerous choices of fresh fields might even impose a reputational cost. So too with advertisers choosing music to sell their products. Advertisers have no reason to overgraze when musical options are plentiful, and, more importantly, when the costs associated with annoying the public are too high. Overuse of promotional music, as with broadcast music, would be a bad marketing decision that is unlikely to need regulation.

Outside of the context of background music, the role of consumer choice may also help explain any absence of overused works. Consider books, which unlike trademarks and sometimes songs, require an element of consumer choice in their consumption. One can imagine the public getting tired of encountering a ubiquitous song or getting tricked by a misused trademark, but it's difficult to see how the multiplicity of editions of a book could make the public sick of the story. *My Antonia* (1918), by Willa Cather, is available in at least 50 different editions by at least 50 different publishers in many formats (cheap paperback, trade paper, hard cover, large print, curricular unit, ebook, audio tape and audio cd) at prices as low as \$2 and as high as \$108;²⁷ yet, no consumer has to unwillingly encounter the story or its characters. If a consumer encounters the same song in the advertising for fifty products, he or she may get tired of hearing it. The song could not be avoided without turning off the television, switching off the radio, and avoiding places which broadcast ads, but the consumer of books will never be forced to consume even a

synchronization licensing or 'synch') . . . [t]o obtain a synch license, print right, or sample clearance, you need to contact the music publisher directly."

²⁶Two pieces that have come to annoy me in commercials, Gershwin's *Rhapsody in Blue* and the famous choral section from Orff's *Carmina Burana*, are still under copyright.

²⁷See www.booksinprint.com.

single one of the fifty editions of *My Antonia*. It is difficult to see a work ever wearing out in a situation when the public only encounters it when it chooses to. Consumer choice/avoidance can be an effective form of non-governmental regulation preventing a work from wearing out.

In order to state general conditions where concerns of overexploitation might be justified, one must consider the likely private regulation by both producers of works and consumers of them. Consistent with the findings in this study, we should expect to find congestion in markets for intangible goods potentially protected by copyright only when three conditions exist: 1) substitutes for the good are not cheap and plentiful; 2) additional subsequent uses of the good entail no significant reputational or other costs to the producer (e.g. by alienating consumers); and 3) consumption of the good by consumers cannot easily be avoided by them (e.g. some advertising uses).

4.2.2. *Debased Songs? Debased Anything?* The data analyzed in Part II suggest that public domain musical compositions appear in movies with about the same frequency as one would predict that similar copyrighted compositions would appear. This result suggests they have not been debased in some way by inappropriate uses that render them no longer fit for public consumption. My earlier study of fiction is even more strongly suggestive of a lack of this sort of congestion. Yet, worry over inappropriate uses debasing works persists.

As noted above, virtually every commentator who takes the possibility of debasement seriously points to unauthorized uses of fictional characters as his or her prime example, rather than the making of unauthorized copies of books or songs. The entire debate seems to turn on the effect of having unauthorized porn movies starring Mickey Mouse (see Landes and Posner, 2003) or Superman (see Green, 2003). No commentators worried about unauthorized pornography seem aware of the vast amount of unauthorized “inappropriate” works that have already been produced. A quick search of the Internet Adult Film Database (www.iafd.com) reveals six pornographic movies with “Cinderella” in the title, including *Cinderella in Chains* and its two sequels, three with Snow White in the title, and a whopping 19 featuring Santa Claus. Searches on the same database of “Apollo” and “Zeus” turn up numerous examples of gay cinematic achievement. Unauthorized porn fan fiction also abounds, starring such characters as Harry Potter, Captain Kirk and Mr. Spock, and Starsky and Hutch.²⁸ Is there a serious argument that Cinderella, Santa, mythical Greek Gods, Harry Potter, and Star Trek characters are worth less now than before these works were produced?

²⁸See “Slash fiction” at http://en.wikipedia.org/wiki/Slash_fiction.

Probably not. Consumer and producer self-regulation likely combine to nullify the potential negative effects of unauthorized uses of fictional characters. Consumers who would be offended by a porno Mickey will not purchase a movie or read the fan fiction setting forth his daring new exploits. Those who deliberately seek out the new Mickey will do so because the porn version enhances Mickey's value to them, rather than detracts from it. Movies, books, and images that must be deliberately sought out by consumers are unlikely to affect negatively the value of the fictional characters portrayed therein.

This observation suggests that the most serious problem might be posed by goods, like t-shirts, which cannot be avoided by the public when the wearer strolls down the street. This danger is probably lessened by the natural reluctance of producers and distributors to sell offensive material. The GAP is unlikely to start selling a t-shirt portraying Mickey and Goofy in bed together. In other words, producer self-regulation, like consumer self-regulation, diminishes the likelihood that serious damage will be done to an iconic character. The internet, however, provides a venue where the reputation costs of selling offensive items like t-shirts may be low enough to sustain a market. If the GAP will not sell the offensive t-shirt, then someone on-line might. An internet purchase might end up being displayed on the chest of someone walking down the street. We could potentially encounter an image portraying Mickey and Goofy in compromising circumstances, despite our best efforts to avoid it.

The number of pedestrians wearing offensive gear, however, is likely to be quite low. There are reputational costs to the wearer that will deter all but a handful of people from displaying such goods in public. And more importantly, Disney will employ its lawyers to prevent the unauthorized sale of its trademarked images.²⁹ Trademark law provides strong protection against unauthorized uses of franchised fictional characters. Not all characters function as trademarks, however, so the potential for an offensive Cinderella or Snow White t-shirt remains a possibility, although the author has never encountered one.

To generalize conditions from the discussion above, debasement of a work not protected by copyright would seem unlikely when: 1) Consumers must deliberately seek out and consume the good; 2) Presenting the good to the consumer entails no reputational or other costs to the producer (e.g. by alienating consumers); 3) Public consumption entails no reputational costs to the consumer; and 4) Consumption is lawful (e.g. it entails no violation of trademark law, obscenity law or libel). These

²⁹See <http://tess2.uspto.gov/bin/showfield?f=doc&state=df4mjh.2.22> (trademark registration for image of Mickey Mouse); <http://tess2.uspto.gov/bin/showfield?f=doc&state=df4mjh.5.3> (Disney trademark registration for Goofy).

four conditions should be met so infrequently that the burden of proving over-exploitation should be squarely placed on those who claim it is a serious problem worthy of government intervention in the market.

5. CONCLUSION

The study of the most popular musical compositions published from 1913-32 as they appear in movies from 1968-2007 suggests that the film market for public domain music functions as efficiently as the market for copyrighted music without any special governmental intervention, such as retroactive copyright term extension. This confirms similar research conducted on the exploitation of bestselling fiction from the same era. These studies cannot prove that copyright protection beyond that necessary to stimulate the creation of a work in the first instance is never necessary, but they suggest that the over- and under-exploitation hypotheses are over-stated. Surely the time has come to place the burden of proof on those who predict valuable works in the public domain will suffer from serious market failure. Legislation should be based on sound empirical evidence.

In the absence of concrete evidence, we are left with predicting the behavior of rational actors, which indicates that self-regulation by producers and consumers of public domain goods will discipline the market. Their likely behavior suggests four conditions necessary for under-exploitation and four conditions necessary for over-exploitation. These conditions suggest that any legislative response should be very specifically targeted to a very narrow set of works. Blanket term extension to all sorts of works in all sorts of contexts, with its significant attendant costs, cannot be justified by a handful of very narrow, and unproven, hypothetical assumptions.

REFERENCES

- Ackerloff, G.A. et al.** (2002), Amicus Brief in *Eldred v. Ashcroft*, available on <http://cyber.law.harvard.edu/openlaw/eldredvashcroft/supct/amici/economists.pdf>
- Bradford, L.** (2005), "Parody and Perception: Using Cognitive Research to Expand the Fair Use Exception in Copyright Law", *Boston College Law Review*, 46; 706-70.
- Brooks, T.** (2005), Survey of Reissues of U.S. Sound Recordings, available on <http://www.clir.org/pubs/reports/pub133/pub133.pdf>
- Cohen, J. et al.** (2006), *Copyright in a Global Information Economy*, Boston, MA: Aspen.
- Ellickson, R.** (1993), "Property in Land", *Yale Law Journal*, 102; 1315-1400.
- Epstein, R.** (2005), "Liberty v. Property? Cracks in the Foundations of Copyright Law", *San Diego Law Review*, 42; 1-28.
- Fennell, L.** (2004), "Common Interest Tragedies", *Northwestern Law Review*, 98; 907-90.
- Furia, P.** (1990), *The Poets of Tin Pan Alley: A History of America's Greatest Lyricists*, New York: Oxford University Press.

- Green, M.** (2004), "Copyrighting Facts", *Indiana Law Journal*, 78; 919-64.
- Hardin, G.** (1968), "The Tragedy of the Commons", *Science*, 162(3859); 1243-8.
- Heald, P.** (2008), "Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Best Sellers", *Minnesota Law Review*, 92; 1031-63.
- Hughes, J.** (1999), "Recoding, Intellectual Property and Overlooked Audience Interests", *Texas Law Review*, 77; 923-1010.
- Jasen, D.A.** (1988), *Tin Pan Alley: The Composers, The Songs, The Performer and Their Times*, New York: Oxford University Press.
- Karjala, D.** (2006), "Congestion Externalities and Extended Copyright Protection", *Georgetown Law Journal*, 94; 1065-86.
- Kozinski, A.** (1994), "Mickey and Me", *University of Miami Entertainment and Sports Law Review*, 11; 465-70.
- Landes, W. and R. Posner** (2003), "Indefinitely Renewable Copyright", *University of Chicago Law Review*, 70; 471-518.
- Lemley, M.** (2004), "Ex Ante Versus Ex Post Justifications for Intellectual Property", *University of Chicago Law Review*, 71; 129-49.
- Liebowitz, S. and S. Margolis** (2005), "Seventeen Famous Economists Weigh in on Copyright: The Role of Theory, Empirics, and Network Effects", *Harvard Journal of Law and Technology*, 18; 435-57.
- Magliocca, G.** (2001), "One and Inseparable: Dilution and Infringement in Trademark Law", *Minnesota Law Review*, 85; 949-1035.
- Mattfield, J.** (1962), *Variety Music Cavalcade: 1620-1961. A Chronology of Vocal and Instrumental Music Popular in the United States*, New Jersey: Prentice-Hall.
- Mulligan, D. and J. Schultz** (2002), "Neglecting the National Memory: How Copyright Term Extensions Compromise the Development of Digital Archives", *Journal of Appellate Practice and Process*, 4; 451-83.
- Smith, G.V.** (1997), *Trademark Valuation*, New York: Wiley.

APPENDIX: The data set of the popular songs³⁰

The subset of the 74 songs with $k \geq 4$ used in the analysis is the following:

Year	Title (Composer(s))
1909	<i>By the Light of the Silvery Moon</i> (Edward Madden; Gus Edwards)
1910	<i>Let Me Call You Sweetheart</i> (Beth Whitson; Leo Friedman)
1911	<i>Alexander's Ragtime Band</i> (Irving Berlin)
1912	<i>It's a Long Way to Tipperary</i> (Jack Judge; Harry Williams)
1913	<i>El Choclo</i> (A.G. Villoldo; G.J.S.W.)

³⁰The statistical analysis in this appendix was compiled by Professor Jaxk Reeves and Kun Xu (Statistics Department, University of Georgia).

- Danny Boy* (Frederick E. Weatherly)
You Made Me Love You—I Didn't Want to Do It (Joe McCarthy; James V. Monaco)
- 1914 *St. Louis Blues* (William Christopher Handy)
- 1915 *Pack Up Your Troubles in Your Old Kitbag and Smile, Smile, Smile*
 (George Asaf; Felix Powell)
- 1916 *Colonel Bogey* (Kenneth J. Alfred (pseud. of Major F.J. Ricketts))
I Ain't Got Nobody (Roger Graham; Spencer Williams & Dave Peyton)
Poor Butterfly: The Big Show (John L. Golden; Raymond Hubbell)
- 1917 *Over There* (George Michael Cohan)
- 1918 *After You've Gone* (Henry Creamer & Turner Layton)
- 1920 *Avalon* (Al Jolson & Vincent Rose)
Look for the Silver Lining: Good Morning, Dearie (Bud De Sylva; Jerome Kern)
Whispering (Malvin Schonberger; John Schonberger)
- 1921 *The Sheik of Araby: Make it Snappy* (Harry B. Smith & Francis Wheeler; Ted Snyder)
- 1922 *Hot Lips* (Henry Busse, Henry Lange & Lou Davis)
- 1923 *Bugle Call Rag* (Jack Pettis, Billy Meyers & Elmer Schoebel)
- 1924 *The Man I Love: Strike Up the Band* (Ira Gershwin; George Gershwin)
Tea for Two: No, No, Nanette (Irving Caesar; Vincent Youmans)
- 1925 *Manhattan: Garrick Gaieties* (Lorenz Hart; Richard Rodgers)
Rhapsody in Blue (George Gershwin)
Show Me the Way to Go Home (Irving King)
Sweet Georgia Brown (Ben Bernie, Maceo Pinkard & Kenneth Casey)
Yes Sir, That's My Baby (Gus Kahn; Walter Donaldson)
- 1926 *Are You Lonesome Tonight?* (Roy Turk & Lou Handman)
Bye Bye Blackbird (Mort Dixon; Ray Henderson)
La Cumparsita (G.H. Matos Rodriguez; Vincenzo Billi)
Someone to Watch Over Me: Oh, Kay! (Ira Gershwin; George Gershwin)
- 1927 *The Best Things in Life Are Free: Good News* (Bud G. De Sylva, Lew Brown & Ray Henderson)
Blue Skies (Irving Berlin)
My Blue Heaven (George Whiting; Walter Donaldson)
- 1928 *I Can't Give You Anything But Love* (Dorothy Fields; Jimmy McHugh)
I Wanna Be Loved By You: Good Boy (Bert Kalmar; Herbert Stothart & Harry Ruby)
If I Had You (Ted Shapiro, Jimmy Campbell & Reginald Connelly)

- Let's Do It: Paris* (Cole Porter)
- Let's Misbehave: Paris* (Cole Porter)
- Makin' Whoopee!* (Gus Kahn; Walter Donaldson)
- Sweet Lorraine* (Mitchell Parish; Cliff Burwell)
- When You're Smiling—the Whole World Smiles with You* (Mark Fisher, Joe Goodwin & Larry Shay)
- 1929 *Ain't Misbehavin': Hot Chocolates* (Andy Razaf; Thomas Waller & Harry Brooks)
- Am I Blue?* (Grand Clarke; Harry Akst)
- Bolero* (Maurice Ravel)
- Happy Days Are Here Again* (Jack Yellen; Milton Ager)
- Honeysuckle Rose: Load of Coal* (Andy Razaf; Thomas Waller)
- Singin' in the Rain* (Arthur Freed; Nacio Herb Brown)
- Star Dust* (Mitchell Parish; Hoagy Carmichael)
- You Do Something to Me: Fifty Million Frenchmen* (Cole Porter)
- 1930 *Beyond the Blue Horizon* (Leo Robin; Richard Whiting & W. Franke Harling)
- Body and Soul: Three's a Crowd* (Edward Heyman, Robert Sour & Frank Eyton; John W. Green)
- Embraceable You: Girl Crazy* (Ira Gershwin; George Gershwin)
- Exactly Like You* (Dorothy Fields; Jimmy McHugh)
- Georgia On My Mind* (Stuart Gorrell; Hoagy Carmichael)
- Get Happy* (Ted Koehler; Harold Arlen)
- I Got Rhythm: Girl Crazy* (Ira Gershwin; George Gershwin)
- Just a Gigolo* (Irving Caesar; Leonello Casucci)
- Love for Sale: The New Yorkers* (Cole Porter)
- My Ideal* (Leo Robin; Richard Whiting & Newell Chase)
- On the Sunny Side of the Street* (Dorothy Fields; Jimmy McHugh)
- Sleepy Lagoon* (Jack Lawrence; Eric Coates)
- Three Little Words* (Bert Kalmar; Harry Ruby)
- You Brought a New Kind of Love to Me* (Sammy Fain, Irving Kahal & Pierre Norman)
- 1931 *Dancing in the Dark: The Band Wagon* (Howard Dietz; Arthur Schwartz)
- Dream a Little Dream of Me* (Gus Kahn; W. Schwandt & F. Andree)
- I Found a Million Dollar Baby—In a Five and Ten Cent Store: Billy Rose's Crazy Quilt* (Billy Rose & Mort Dixon; Harry Warren)
- Life is Just a Bowl of Cherries* (Lew Brown & Ray Henderson)
- Minnie, the Moocher—The Ho De 'Ho Song* (Cab Calloway & Irving Mills)

Mood Indigo (Duke Ellington, Irving Mills & Albany Bigard)
Out of Nowhere (Edward Heyman; John W. Green)
 1932 *It Don't Mean a Thing* (Irving Mills; Duke Ellington)
Night and Day (Cole Porter)
You're Getting to Be a Habit with Me (Al Dubin; Harry Warren)

Description

This set of data consists of 74 songs, composed in 1909-1932, which appeared at least 4 times in films during 1968-2007. The most popular songs *Star Dust* and *La Cumparsita* both appeared in film 17 times in our study period. Nineteen of these songs were published between 1909 and 1922. These 19 songs are all currently in the public domain, but were not necessarily in the public domain during the entire 40-year period of this investigation (1968-2007). The other 55 songs were published between 1923 and 1932 are not yet in the public domain. This data set of 74 songs, where $k \geq 4$, is used for most of the analysis, but similarly analysis using thresholds if $k \geq 3$, $k \geq 2$, and $k \geq 1$ are also included. As far as the total appearance of the songs in our list for a certain year in our sample period of 1968-2007 is concerned, this ranges from a low of 2 in 1971 to a high of 41 in 1998.

Data manipulation

As stated in the introduction, the first analysis of the popular songs concerns “availability” of songs during 1968-2007. Each song was measured at every year from 1968-2007, a total of 40 time points. A variable called *AFPUB* was then created, representing the number of years after the original publication of the song, so that the values for *AFPUB* are 59,60,...,98 respectively. The modified data set thus has $74 \times 40 = 2960$ observations. This modified data set is called the song-year version of the popular songs. Three other variables, $YR = PUBYR + AFPUB$, *MOV* (indicator of the appearance of the song; 1=appears in that year; 0=does not appear in that year), and *PD* (indicator of the copyright; 1= in the public domain; 0=not in public domain), were also created from the original data set of $N = 74$ songs and carried over to the new data set of 2960 song-year events.

Statistical Analysis and Results

Before presenting the analysis results, it is necessary to briefly describe the tools and methodology used in the following analyses. Each of the 4 analyses took the same general path. First, the data were explored by numerical and graphical summaries. Then, more sophisticated analyses followed. Since the response variable in this problem is dichotomous, a logistic regression was applied.

Exploratory Data Analysis

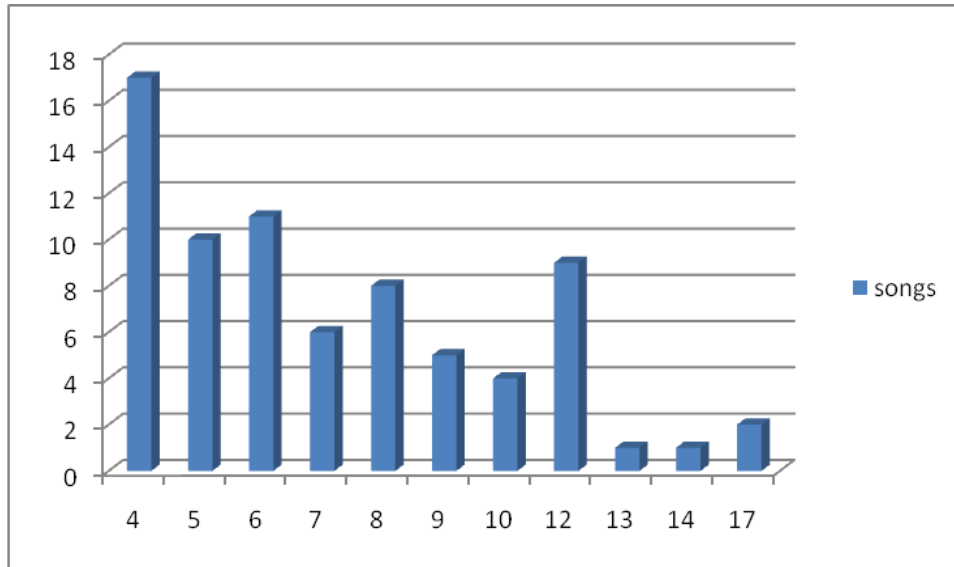


FIGURE 2. Popular Songs by Number of Appearance ($N = 74$)

1. *Preliminary Analysis.* The appearance time of each popular song varies from 4-17 and the total number of appearances is 537. In Figure 2, the histogram shows the frequency of songs appearances. Because no song appears exactly 15 or 16 times in the data set, these two columns don't show in the chart. The average appearance for each song is about 7 times.

Through data manipulation, the appearance of a single song in particular year becomes a dichotomous variable *MOV* (0 if the song didn't appear and 1 if it appeared in that year's movie). Because there were 64 occasions where the some song appeared in more than one film during some year the total number of events in the dichotomous data set was reduced from 537 to 473 unique events. Of the 2960 observations, only 312 are in the public domain and the rest are copyrighted. The percentage of these two groups are shown in Figure 3 below. The copyrighted observations are the majority with percentage 89.46%.

Furthermore, consider the total appearances in one year (shown in figure 4 below). By focusing on the total appearances, illustrated by the upper-most line, one can see an increase after year 1984 when the songs published in 1909 came to the public domain. The total appearance also shows a sharp decrease after 1998 when the songs in our study stopped entering the public domain. At the same time, appearances of copyrighted songs, illustrated by middle line (during the year 1968-1987, the two lines overlap), shows a steady increase through the whole time

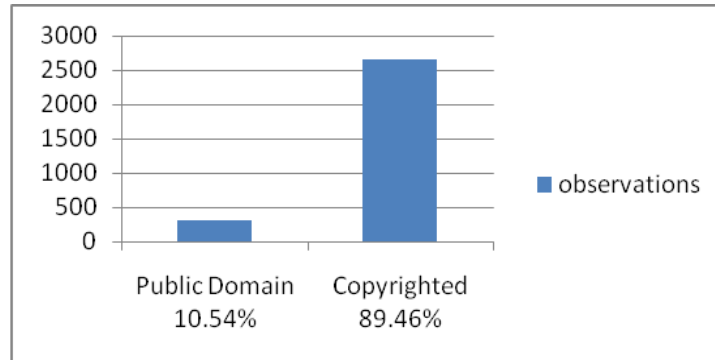


FIGURE 3. Observations by Copyright Status

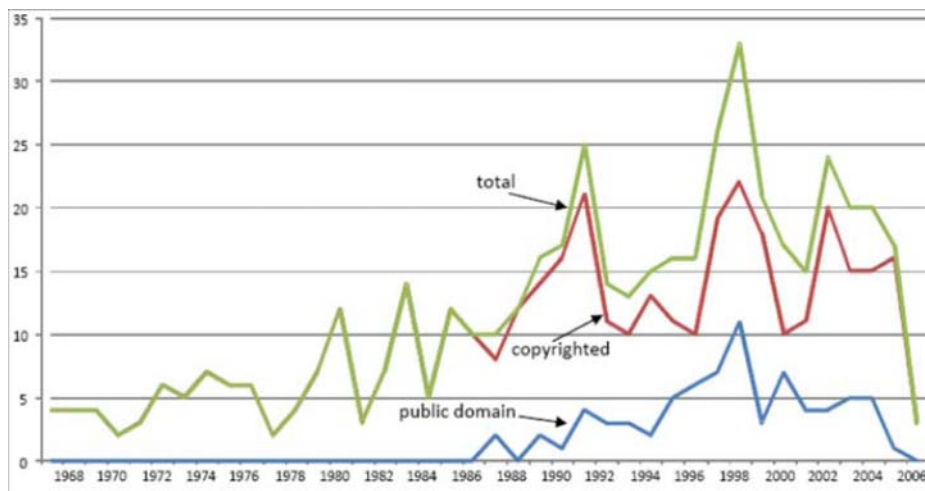
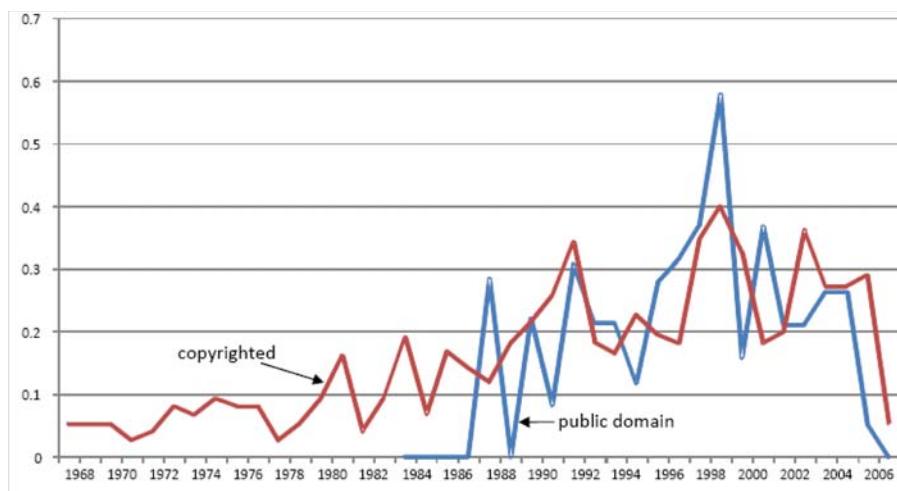


FIGURE 4. Song appearances by year and status

period. Contributions from the songs in public domain give a linear increase in appearance time after 1984.

On the other hand, the total number of observations for songs in the public-domain and copyrighted are not equal. As shown in Figure 3, the proportion of appearances may be more appropriate to illustrate the copyright effect. Divide the number of appearances in any year by the total number of that set for both public-domain and copyrighted song observations. As shown in the Figure 5, we can see a slight difference between the copyright statuses. From all the above, we can propose a null hypothesis that there is no significant difference in occurrence probability between the public domain songs and copyrighted songs, also an alternative

FIGURE 5. Appearance probability by year and status ($k \geq 4$)

hypothesis is that the songs in public domain are more likely to be used in film. To decide which hypothesis is more probable we must perform further analysis.

According to Figures 4 and 5, we also notice that the value in year 2007 has an abnormally sharp decrease. We also do the same preliminary analysis on the popular songs that appear more than 1, 2 and 3 times, those graphs show abnormally sharp decrease in year 2007 as well. It is reasonable to consider the year 2007 as an outlier in this study, (which may be caused by incomplete data), so we don't include observations in 2007 in our further analysis. In year 2007, no public domain song appeared in the film and the copyrighted songs appeared only 3 times. After deleting this year for all 74 songs in 2007, we have a total of 470 appearances including 75 public domain songs and 395 copyrighted songs. The total observation number for all years combined decreases to $74 \times 39 = 2886$.

2. *Popular Songs Analysis 1 (Availability by Song-Year)*. The results of the song-year analysis of the popular songs are presented in this section. The frequency table of availability (*MOV* in rows) versus copyright status (*PD* in columns) is shown below in Table 1:

	Public domain	Copyrighted	Total
Appear	75 24.04%	395 15.35%	470 16.29%
Not appear	237 75.96%	2179 84.65%	2416 83.71%
Total	312	2574	2886

Table 1. *MOV* and *PD* Frequency

The results from Table 1 show that over the period of analysis, 15.35% of copyrighted songs appeared versus 24.04% of the public domain. Assuming each determination of availability is independent from others (which is not quite true here), the frequencies shown above imply that there exists an association between the rows and columns. But is the association statistically significant? The chi-square test for independence of rows and columns is the following.

$$\begin{aligned}
 \chi^2 &= \sum \frac{(O - E)^2}{E} \\
 &= \frac{(2179 - 2154.81)^2}{2154.81} + \frac{(237 - 261.19)^2}{261.19} + \frac{(395 - 419.19)^2}{419.19} + \frac{(75 - 50.81)^2}{50.81} \\
 &= 0.2715 + 2.2403 + 1.3959 + 11.5166 \\
 &= 15.4242
 \end{aligned}$$

The p -value from the chi-square test indicates severe dependency between copyright status and appearance of songs in the movie. The Fisher exact test for positive association (upper-tail test for large sample) is the following (Where c is the sum of the first column, i.e. 312; r is the sum of the first row, i.e. 470; N is the grand sum, i.e. 2886, and $T_2 = 75$).

$$\begin{aligned}
 Z &= \frac{T_2 - \frac{rc}{N}}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}} \\
 &= \frac{75 - \frac{470 \times 312}{2886}}{\sqrt{\frac{470 \times 312 \times (2886 - 470) \times (2886 - 312)}{2886^2 \times (2886 - 1)}}} \\
 &= 3.9265 \\
 P(Z > 3.9265) &< 0.0001
 \end{aligned}$$

The p -value from the Fisher exact test shows that songs in the public domain were used by movie makers at a significantly higher rate than those which were copyrighted. The above result is based on the assumption that all observations are independent from the others. It was used to determine if there exists an association to warrant further analyses. Since a strong dependency exists between copyright status and works' appearance, we proceed with further analysis. Of course, the results above are exaggerated to some extent because each song appeared, on average, about 6 times in the above analysis, and the availability status for a particular song is surely positively correlated over time. However, even under the most severe assumption (that observations for a particular song are completely correlated, so that the sample size is exaggerated by a factor of 6), the value obtained (15.42) would still lead to a very strong evidence of a public domain effect.

3. *Results for Other Thresholds.* The results presented above and analyzed in the bulk of this paper concern the dataset when restricted to the $N = 74$ songs which had appeared in at least 4 films during the 39 years between 1968-2006. This restriction was made so as to include the songs which were clearly ‘popular’ over the period. On the other hand, this is a rather restrictive requirement, since it includes only 74 of the 1294 popular songs released from 1909-1932, with only 19 of these being current public domain songs. If the threshold for inclusion were lowered from $k \geq 4$ to $k \geq 3$, $k \geq 2$, or $k \geq 1$, many more songs could be included, but the reliability of results might decrease. The Table below contains summaries of the data which would occur if one used other inclusion thresholds. The remainder of this appendix will concentrate on the $k \geq 4$ case described in the first row of the table, and discussed heretofore, but results for the other 3 data sets will be presented at the end of the appendix.

$k \geq 4$ ($N = 74$, $PD = 19$, $CP = 55$)			
	<i>PD</i> events	<i>CP</i> events	all events ($Ev = 537$)
<i>UEv</i>	395	75	470
<i>ESY</i>	2574	312	2886
<i>P</i>	0.1535	0.2404	0.1629
$k \geq 3$ ($N = 99$, $PD = 23$, $CP = 76$)			
	<i>PD</i> events	<i>CP</i> events	all events ($Ev = 612$)
<i>UEv</i>	76	464	540
<i>ESY</i>	341	3520	3861
<i>P</i>	0.2229	0.1318	0.1399
$k \geq 2$ ($N = 146$, $PD = 40$, $CP = 106$)			
	<i>PD</i> events	<i>CP</i> events	all events ($Ev = 706$)
<i>UEv</i>	91	542	633
<i>ESY</i>	552	5142	5694
<i>P</i>	0.1649	0.1054	0.1112
$k \geq 1$ ($N = 259$, $PD = 79$, $CP = 180$)			
	<i>PD</i> events	<i>CP</i> events	all events ($Ev = 819$)
<i>UEv</i>	113	633	746
<i>ESY</i>	1058	9043	10101
<i>P</i>	0.1068	0.0700	0.0739

Table 2. Summary of Data Sets Based on Inclusion Threshold k

For each threshold level, Table 2 categorizes the N songs that meet the threshold requirement into those that have entered the public domain (*PD*) and those that are still copyright protected (*CP*). It should be remembered, course, that the *PD* songs were not *PD* for the entire period of observation. The ‘all events’ column

counts the total number of times that a song is used in a film in the 39-year period from 1968-2006. The true number of events (Ev) is reduced slightly to include only unique events, UEv , since we allow a song to be counted at most once in a given year. Eligible song-years (ESY) is given by $ESY = N \times 39$, since each song is eligible to be in a film for each of the 39 years. The last row of this section measures the proportion of songs used in films, $P = UEv/ESY$. Finally, “ PD events” and “ CP events”, simply subdivide all song-years and associated events into those which occurred under “ PD ” and those which occurred under “ CP ” conditions. For all 4 threshold conditions, it can be noted that P is higher under the PD conditions than under the CP conditions. One can easily perform Chi-squared tests (as was done above for the $k \geq 4$ dataset) to show that the differences are significant. One objection to these tests could be that they do not account for time effects – the PD group has a higher proportion of its songs eligible during the latter years of the observation period than does the CP group. So, if there is an increase in utilization rate over time due to factors unrelated to copyright status, the Chi-squared tests could overstate the importance of the copyright status effect. To investigate this, in the next section of this report, logistic regression models which can control for both copyright status and time (year) are introduced.

Logistic Regression

In analysis 1 (song-year level) of the popular songs, the response variable (MOV) is dichotomous (0 if the song didn’t appear, and 1 if it did appear in that year’s movie). Logistic regression is appropriate for modeling this type of response variable.

Using copyright status (PD) alone to model availability (MOV) might omit other significant factors affecting songs appearance in films. Other variables which could be included in the model are $PUBYR$, $AFPUB$, and YR . All four variables (PD , $PUBYR$, $AFPUB$, and YR) are possible explanatory variables for $CPUB$. Since copyright status is the explanatory variable of primary interest, it was the first variable included in the model. Care needs to be taken when choosing additional variables to include in the model to avoid confounding effects since some of these variables are function of others. For example, copyright status (PD) depends solely on publication year ($PUBYR$) and age of the work ($AFPUB$), and the calendar year of the measurement (YR) is the sum of publication year ($PUBYR$) and age of the work ($AFPUB$). According to our data, the year 1984 is a key observation point, because the songs in our study start to fall in public-domain in that year. We make a new variable $PY84$, defined as $PY84 = YR - 1984$. Since period is another effect of interest and $PY84$ was not highly correlated with PD , it was included in the model (Figure 4 shows a increase in total appearance after year

1984). Including either *PUBYR* or *AFPUB* in this model (along with *PD* and *PY84*) will cause some confounding, so we did not attempt this.

Of course, just because appearance is more likely for *PD* than *CP* events, this doesn't prove that *PD* is significantly higher than *CP*. The main confounder is year, since there were many more *PD* eligible during later years, and there seems to be a strong year effect. To investigate this, we considered a 7-level hierarchy of linear models:

$\ln(P/Q) = B_0$ (model 0)
$\ln(P/Q) = B_0 + B_1 \times PD$ (model 1; <i>PD</i> only)
$\ln(P/Q) = B_0 + B_2 \times PY84$ (model 1L; linear in <i>PY84</i>)
$\ln(P/Q) = B_0 + a(grp)$ (model 1G; grouped year)
$\ln(P/Q) = B_0 + B_1 \times PD + B_2 \times PY84$ (model 2L; linear, additive)
$\ln(P/Q) = B_0 + B_1 \times PD + a(grp)$ (model 2G; grouped, additive)
$\ln(P/Q) = B_0 + B_1 \times PD + B_2 \times PY84 + B_3 \times PD \times PY84$ (model 3L; linear, interaction)

In fact, Models 1L and 1G are similar in all cases, since the trend is close to linear.³¹ The real question is whether the B_1 coefficient in model 2L (or 2G) is significantly different from zero, or whether it can be thrown out, bringing us back to Model 1L (or 1G). It turns out that in every case, the answer is “not significant”; there is no effect of *PD/CP* on appearance, once one controls for the year effect. The fit for selected models for $k \geq 4$ is shown in Table 3 below.

model	B_0	B_1	B_2	B_3	$-2 \ln L$	AIC	SBC
0	-1.6371				2565	2567	2573
1	-1.7077	0.5571			2551	2555	2567
1L	-1.9642		0.0598		2405	2409	2421*
1G	-1.7840		grp 5		2398	2408*	2437
2L	-1.9602	-0.0554	0.0603		2405	2411	2429
2G	-1.7708	-0.0530	grp 5		2398	2410	2446
3L	-1.9704	0.4635	0.0618	-0.0365	2403	2411	2435

Table 3. Summary of 7 Hierarchical models for the $k \geq 4$ Dataset

Based on the AIC or BIC we can either pick the model with continuous year effect or the grouped year effect as our final model. Both of the models have the same interpretation of the data, which is that the probability of the songs appearing in the film increase over time, but there is no effect due to *PD/CP*.

³¹The grouped method uses 5 blocks of 8 years, but similar results were found with 10 blocks of 4.

We also do the same analysis on other threshold and the result shows the same trend on the data set. The crucial results come from the analyses of Model 2L for each data set. In each case, the p -value for the PD effect ($B1$ in the model) is not statistically significantly different from zero, as shown in Table 4 below. Thus, after accounting for the increase in appearance rates over time, there is no evidence of a positive or negative effect on appearance probability due to being in public domain or not. This holds for all 4 data sets.

data set	$B1$ estimate	$SE(B1)$	z -stat	2-tailed p -value
$k \geq 4$	-0.0534	0.1504	-0.355	0.7241
$k \geq 3$	0.0328	0.1505	0.218	0.8277
$k \geq 2$	-0.0854	0.1304	-0.655	0.5120
$k \geq 1$	-0.1171	0.1145	-1.022	0.3067

Table 4. Parameter Estimates of $B1$ for 2L Models

Conclusions

A naïve analysis of the data (the chi-squared & fisher's test) demonstrated a clear difference in availability of songs between copyrighted and public domain works, with public domain works being significantly more appearance in film. A serious objection to this analysis is that it controlled for neither period effects nor for popularity of songs considered. After the logistic regression analysis to control for time-period effects, we find that the copyright status plays no significant role in affecting the probability of a song's appearance in a film.

VISITING PROFESSOR OF LAW, UNIVERSITY OF CHICAGO; ALLEN POST PROFESSOR OF LAW, UNIVERSITY OF GEORGIA; VISITING PROFESSOR, CENTRE FOR INTELLECTUAL PROPERTY POLICY AND MANAGEMENT, BOURNEMOUTH UNIVERSITY.