THE BATTLE OVER DIGITAL RIGHTS MANAGEMENT:

A MULTI-METHOD STUDY OF THE POLITICS OF

COPYRIGHT MANAGEMENT TECHNOLOGIES

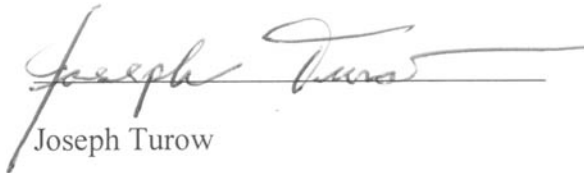Bill D. Herman

A DISSERTATION

In

Communication

Presented to the Faculties of the University of Pennsylvania in Partial

Fulfillment of the Requirements for the Degree of the Doctor of Philosophy

2009

Michael X. Delli Carpini

Supervisor of the Dissertation

Joseph Turow

Graduate Group Chairperson

A full copy of this work is available online from the author's Social Science Research

Network author page. See:

http://ssrn.com/author=545408

This work is also available on the author's personal website. See:

http://www.billyherman.com

DEDICATION

This dissertation is dedicated to my wife, partner, and best friend,

Christina M. Collins

ACKNOWLEDGEMENTS

Thanks as well to the Collins family and to my wife's extended family. Tina, your

undying love and support mean the world to me.

I could not have done this without all of you, and I hope I can continue to make

each of you proud.

ABSTRACT

THE BATTLE OVER DIGITAL RIGHTS MANAGEMENT:

A MULTI-METHOD STUDY OF THE POLITICS OF

COPYRIGHT MANAGEMENT TECHNOLOGIES

Bill D. Herman

Michael X. Delli Carpini

Digital rights management (DRM) refers to various technological systems by which copyright holders seek to exert control over the use and circulation of their works. This dissertation explores the policy debate over copyright law as a potential vehicle for regulating DRM technologies. It examines this debate in three separate time periods, between 1989 and 2006, as it took place in Congress, in *The New York Times* and *Washington Post*, and online. It answers the question: Which policy actors communicate most regularly in which media about DRM and copyright law, and how has this changed over time?

Methods used include quantitative content analysis of documents from all three media, qualitative historical policy analysis, and web graph analysis tools that quantify and map the hyperlinks between websites. This work builds upon and extends the methodology of using web graphs as a tool for identifying the most central actors within a topical cluster of websites.

Results illustrate the birth and growth of a fairly unified multi-sector strong fair use coalition. Voices of opposition to the regulation of DRM via copyright have moved from profound underrepresentation to approximate parity in congressional access, successfully moved press coverage in a more favorable direction, and dominated the online debate. Policy outcomes reflect this shift; while the strong copyright coalition successfully pushed through two major laws expanding copyright in the 1990's, by the mid-2000's, the strong fair use coalition had fought them to a draw, stopping proposed expansions of copyright and winning key congressional allies for a proposal to reduce DRM regulations.

This dissertation's results suggest the substantial power of online issue advocacy. In particular, the web benefits policy coalitions that have a disadvantage in financial capital but a comparatively large base of support. Coalitions still need regular interpersonal communication with policymakers, but online coalition building and advocacy appear to be of substantial help, legitimizing and amplifying the message of under-resourced coalitions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

In June 2007, world-famous legal scholar Lawrence Lessig surprised the world by announcing that he would no longer focus his scholarship and public advocacy on copyright, the primary area of law on which he had built his reputation (Lessig, 2007). In a recent lecture that this author was fortunate to attend (Lessig, 2009; Lessig, Fairey, & Johnson, 2009; Stemmler, 2009), he described his trajectory away from focusing on the substance of copyright reform. Copyright law[1] kept surprising him. Time and time again, he would think that a given policy choice posed by copyright law was a "no brainer," and yet Congress kept getting them wrong. He thought that only a Congress filled with "idiots" could be so mistaken so often.

Then, Lessig came to believe that the problem was not with Congressional idiocy but with the policymaking process—that the problem was not unique to copyright law or curable by better policymaker understanding of the issue. He came to the conclusion that the political system is profoundly corrupted by the influence of money—not that politicians are generally guilty of lining their own pockets, but that the need to raise staggering amounts of campaign cash is a corrupting process that leads to laws that favor the most economically concentrated interests at the expense of the broader public good. Before we can have better copyright law—or better health care policy, environmental policy, and so on—he believes we need a better system for determining whose voices are

---

[1] For the reader who is unfamiliar with the contours of copyright in general, I have included a primer on the basics. Please see Appendix A.

heard by policymakers. In short, he started by focusing on the subject of copyright law but has transitioned to a focus on policy process.

Much like Lessig, this dissertation is the result of a process that saw its author's interests migrate from an intense interest in the issue of copyright to an interest in the policy process. Starting with a critique of the romantic theory of musical authorship born of my hobby as an electronic music DJ (B. D. Herman, 2002, 2006c), I quickly discovered just some of what was already a wealthy and growing literature tying this critique to a concern about the ways in which copyright limits artistic creativity (see, e.g., Jackson, 1995; McLeod, 2001; Negativland, 1995; Vaidhyanathan, 2001; Woodmansee & Jaszi, 1994). This led to a focus on Title I of the Digital Millennium Copyright Act (1998), the most important copyright law in decades (Nimmer, 2000). As discussed in much fuller detail below, the law makes it illegal to circumvent most kinds of digital rights management (DRM), the technologies deployed by copyright holders to limit what users can do with copyrighted works.

Along with Oscar Gandy (B. D. Herman & Gandy Jr., 2006), I have already expressed the opinion that the DMCA is a bad law skewed toward copyright holders' wishes, and the vast majority of commentary echoes this belief. Yet this law faced little opposition in Congress, and attempts to reform it have been turned back. Those who seek to expand copyright and retain past expansions—the strong copyright coalition—have long been the dominant coalition, driving copyright policymaking. The DMCA stands as a testament to this power.

In the process of researching the DMCA, I realized that to more fully understand DRM regulation, I needed a deeper understanding of the policy process behind copyright law and the law in general.[2] This dissertation is my most substantial venture yet in that direction, and while those who seek to reform copyright still have less political capital than those who seek to expand it, this work does offer hope to would-be reformers. Lessig may be right that systemic reform is necessary to effect copyright reform, but this research finds reason to suspect that his highly improbable crusade for systemic political reform may not be the only route to a smaller role for copyright. In particular, it suggests that new media technologies such as the web give a substantial political boost to those groups calling for moderated copyright law—what I call the strong fair use coalition— and this newfound communication power makes dreams of reform substantially more realistic.

If there is a candidate for a formerly obscure issue on which public sentiment might be brought to bear in trumping special interests, it may be copyright law, and internet advocacy is an elemental part of that possibility. This dissertation suggests that the internet provides a new and exciting tool for reshaping the policy advocacy process. The strong fair use coalition has been particularly effective at leveraging this tool to their advantage. While the time period under study does not include their ultimate triumph at the bargaining table—as of this writing, what I describe as the strong fair use coalition still has not won a major legislative victory—it does include the beginning of their time as a genuine force at that table, and this says something new and exciting about the policy

---

[2] Professor Gandy realized this well before I did.

process in general. By studying the birth and growth of the strong fair use coalition, I hope to say something about how under-resourced coalitions in general can leverage what strategic advantages they have, using a combination of offline and online advocacy.

More specifically, this project is a study of the debate over DRM policy as that debate happened over four major policy proposals across three time periods. These proposals and the time periods studied include:

- Audio Home Recording Act, or AHRA (1989-1992)

- Digital Millennium Copyright Act, or DMCA (1995-1998)

- Proposals to reform the DMCA (2003-2006)

- Proposals to impose a DRM system called the "broadcast flag" (2003-2006)

I include a qualitative description of each of these policy debates, but the heart of the research is a quantitative study of the debates as they played out in Congress, in major national newspapers, and—in the case of the two most recent debates—online.

The remainder of this introduction gives a preview of the rest of the dissertation. First, I discuss the project's theoretical framework. Second, I apply that framework to the copyright debate and give an overview of the research questions and hypotheses that follow. Third, I summarize the methodological strategy I employ in answering these questions. Fourth, I give a brief overview of the four major policy debates that are included in the study. Fifth, I describe some of the more noteworthy quantitative results in my study of congressional hearings, *The New York Times* and *Washington Post*, and

the online copyright debate. I conclude with some brief thoughts about the study as a whole.

Theoretical Framework

There is a wealth of literature on policy systems and a bounty on political communication, but there is room for far more intermingling between these areas if one seeks to understand how strategic political communication helps shape policy outcomes. Chapter Two explores some of the literature in both areas and then seeks to combine the two. I begin with theories of policy stability and change, then consider theories of mediated communication, and finally seek to fuse and update these theories to account for recent developments in media technology—especially the widespread adoption of the internet.

Theories of Policy Stability and Change

On most issues, most of the time, US federal, state, and local law and policy are remarkably constant. Yet there are generally some changes afoot at all levels of government, and almost every policy area eventually does see substantial change. Policy change happens in fits and starts—brief windows of substantial change punctuate seemingly interminable periods of equilibrium. I draw on several theoretical frameworks to provide some light on this cycle of stability punctuated by substantial change. They include the theory of the iron triangle (Gais, Patterson, & Walker, 1984), the Advocacy Coalition Framework, or ACF (Sabatier & Jenkins-Smith, 1999), and the general punctuated equilibrium (Jones & Baumgartner, 2005). I explain each in detail in Chapter

Two. Then, I integrate them into a theory of policy change that can be summarized as follows.

Most policy debates happen within subsystems, composed of groups of individuals and organizations that regularly try to influence policy on a given topic. Each subsystem generally features one to four policy coalitions, each made up of like-minded individuals and groups who work together toward a commonly held set of policy-specific values and goals. Coalition members may include business interests, nonprofit groups, interested individuals, policymakers, scholars, and journalists. In most cases, if a subsystem features at least two competing coalitions, the status quo at any given time will most closely represent the will of the coalition that most recently succeeded in enacting a major overhaul of the policy in dispute. This most recently victorious group is the governing or dominant coalition, and they generally fend off attempts by one or more challenger coalitions to effect a new overhaul of policy.

In addition to the political power they exhibited in pushing through the most recent policy overhaul, the dominant coalition also enjoys several structural advantages in their efforts to defend the status quo. Humans have cognitive limitations, including a narrow bottleneck of attention and the tendency to economize thinking by reproducing earlier decisions and relying on previously established rules. The division of labor within policymaking bodies further entrenches the status quo. Specific decisions are generally delegated to members of specialized bodies. These policymakers come to learn a great deal about their specialties, their opinions harden, and they often gravitate into a coalition.

The structural characteristics of human cognitive processing and policymaking bodies do much to account for the pattern of punctuated equilibriums. Policymakers use a familiar rule to decide an issue—often simply the decision to delegate to other policymakers—and they then pay attention to other matters until some substantial disturbance to a policy subsystem forces them to pay attention again. Once forced to reexamine their earlier decision, policymakers tend "to overreact with 'alarmed discovery'" (Jones & Baumgartner, 2005, p. 52). Humans and the Senate floor can only pay attention to one thing at a time; once they do, however, they often behave as if nothing else exists. Thus, moments in which the entire policy system pays attention to an issue are generally necessary parts of a policy window—an opportunity for challenger coalitions to enact major reform.

Challenger coalitions thus have an incentive to seek the attention of the broader policy system and the voting public in an effort to expand the scope of conflict. They will try to force policymakers outside the given subdivision—e.g., outside the appropriate congressional committee—to pay attention to a new issue, to adopt a new paradigm, and to overrule their colleagues with a longer history of studying the matter. Each case is a difficult one to make. Reform-minded coalitions also have an incentive to appeal to public opinion, as a mobilized constituency can often force members, whether on the "right" committee or not, to revisit an issue. This appeal to public opinion is generally conducted, at least in part, via mediated communication; thus, it is also fruitful to consider theories of mediated communication and seek to fit these theories into the policymaking process.

Theories of Mediated Communication

Countering decades of research findings to the contrary, recent research into the effects of mediated political communication has generally expanded the estimated degree of influence on audiences (Iyengar & Simon, 2000). Chapter 2 reviews some of these effects, a non-comprehensive list of such effects that nonetheless makes a case for why mediated political messages are important. Audiences learn about the state of the world through mediated messages. Media help shape audiences' opinions about which issues deserve attention—which issues belong on the public agenda, and which among them are more important. Mediated messages frame the important questions, helping to limit the range of possibilities into a cognizable number of options. There is even good evidence that media messages persuade their audiences—challenging the long-held belief among media researchers that media are not particularly effective at telling us what to think, only what to think about.

Those who have a stake in media content certainly behave as if mediated messages affect their audiences. In the form of public relations and other messages, such as policy reports, interested groups and individuals provide favorable information to the media and to policymakers, subsidizing the high costs of acquiring information. Particularly in the case of media organizations, providing press-ready copy or broadcast-ready video or audio is an effective information subsidy, creating in cost-conscious media companies a dependence on the kinds of well-funded sources who can provide these subsidies (Davis, 2002; Gandy Jr., 1982). News outlets also have a tendency to favor

8

institutional sources with established legitimacy, which can also often means a bias toward the status quo (E. S. Herman & Chomsky, 2002).

Like policymaking institutions, media institutions are subject to punctuated equilibriums of attention. An issue may get little coverage for months or years, but once it is viewed as an important issue, audiences may tire of the seemingly endless coverage. These are potentially invaluable opportunities for challenger coalitions in their quest for major reform, which becomes clear upon a synthesis of these theories of policy process and of mediated communication.

Synthesizing and Updating: Toward A Theory of Communicating Policy Actors

Taken together, the above theories suggest that different coalitions have an incentive to engage in different communication strategies. As long as the public and the broader political system are paying little attention to an issue, the odds are slim that a policy will undergo major reform. Thus, governing coalitions have little incentive to seek to bring the debate to a wide audience. Previous policy victory has generally come in part because they enrolled at least the tacit support of the relevant specialist policy bodies such as congressional committees and relevant agencies. Governing coalitions will want to stay on good terms with news reporters so that their spin is included in most coverage, but they will rarely clamor for heavy coverage. A low volume of news coverage may be of strategic value if a governing coalition seeks to push for modest policy changes, but the total volume of coverage will ideally be kept well below the point that the issue moves toward the top of the general public's agenda. Especially when governing coalitions have more resources to devote to campaign donations and direct

communication with policymakers—which Lessig and others (Bimber, 2003; Cigler &
Loomis, 2007) describe as the norm—a steady investment in these resources seems the
safest best.

In contrast, challenger coalitions generally want to communicate their messages
as broadly as possible in the hopes of forcing a reconsideration of the policy choices that
have been made. Hoping to create a window in which a new frame, new facts, and new
policy choices can become established, they will constantly seek the attention of the
general public and policymakers from outside the specialized bodies that have natural
jurisdiction. If they can gain this attention, occupying the narrow bottleneck of attention,
they have successfully expanded the scope of conflict and likely inspired alarmed
discovery or rediscovery of an issue. This makes it much more likely that they can effect
what seems to the governing coalition to be a radical overreaction of substantial reform.
Thus, challenger coalitions will seek to communicate their message to anybody who will
listen, which means a constant search for more press coverage and more of the public's
attention.

The internet gives challenger coalitions an exciting new tool for seeking public
attention. The last fifteen years have seen internet adoption explode in the US from
almost nothing (Rainie & Horrigan, 2005) to a majority with home broadband
connections and roughly two thirds total with a home internet connection (Horrigan,
2008). For any group seeking the general public's attention, the web is a fantastic
resource, a virtually costless vehicle for communicating with a potentially large number
of citizens who are or might become interested in a given policy issue. There are limits to

the internet's reach—age, education, and income are among the important factors that shape internet access—but assuming that an issue does not require the mobilization of people who tend to be offline, internet activism can be a powerful way to mobilize an issue public (Bennett & Manheim, 2001; Bimber, 2003).

Since challenger coalitions seek to involve the public, putting pressure on policymakers both within and outside the appropriate specialized committees and agencies, public mobilization via internet advocacy is far more useful for them than it is for governing coalitions. Their goal is to heavily subsidize the online debate with large volumes of frequently updated materials, and if they are successful in mobilizing a sympathetic issue public, incoming links will tend to favor the challenger coalition, paradoxically making them the more authoritative source in the online debate. As challenger coalitions are also likely to have fewer resources than governing coalitions, the internet levels the playing field substantially in their favor, reducing the imbalance of communication power between differently resourced coalitions. A challenger coalition therefore has every incentive to go online, hoping to turn their message into a viral internet phenomenon and push through the narrow bottleneck of public and policymaker attention.

Theorizing the Copyright Debate and Building Research Questions and Hypotheses

The above theoretical background applies to the policymaking process in general. In Chapter Three, I apply this theory to the copyright debate. I then use this application to derive research questions and testable empirical hypotheses.

11

Theorizing the Copyright Debate

Applying the above theories helps to develop a theory of strategic communication within the copyright debate. The governing coalition in this space is the strong copyright coalition. Its core membership is made up of those industries most directly dependent on copyright: media industries such as motion pictures, music, and publishing. This also includes technology firms whose primary revenue streams come from proprietary software such as Microsoft and Adobe. Other generally supportive industries include other media sector businesses such as broadcast networks (which do not want unauthorized copies of their content online), law firms (which get more work as copyright grows), and technology firms that make and sell DRM systems. Recent legislation has consistently expanded the reach of copyright generally (Landes & Posner, 2004) and as a tool for regulating DRM specifically (Litman, 2000), strongly implying that the strong copyright coalition has superior access to policymakers. Likewise, the Copyright Office has been a loyal ally of the strong copyright coalition, especially on DRM policy (Herman & Gandy Jr., 2006). These favorable policy outcomes are consistent with the strong copyright coalition's substantial spending on campaign donations.[3]

In contrast, the strong fair use coalition as it is now constituted was virtually nonexistent 15 years ago; the consumer electronics industry played something of a

---

[3] The strong copyright coalition substantially outspends the strong fair use coalition, and even among the strong fair use coalition groups that donate, copyright law is a much lower priority than for the strong copyright sectors that donate campaign funds. See Chapter Three for more.

challenger coalition role through the 1970s and 1980s, but they were not unified by a strong set of beliefs about the future of copyright law. Since then, the strong fair use coalition has blossomed into a substantial political force. Its core membership includes librarians, a large number of legal scholars and other academics, several information policy advocacy nongovernmental organizations (NGOs), and various education sector groups such as societies of computer scientists. Except for vendors of proprietary software and DRM systems, most groups in the technology sector are also best described as part of the strong fair use coalition. The education sector and technology sector are the coalition's only sources of substantial campaign donations, though these are much smaller than those given by the strong copyright sector, and they are less clearly targeted at copyright law.

Based on these coalitions' positions in the policy space, each is likely to engage in very different communication strategies. The strong copyright coalition has every reason to communicate directly with policymakers and otherwise to seek little public attention for proposed policy changes. This means very regular congressional appearance, communication with major newspapers in enough volume to guide the story being told there, and very little online communication. In contrast, the strong fair use coalition has good reason to seek as much public attention as possible. This means communicating directly with policymakers, though their access is likely less than that of the strong copyright coalition. The strong fair use coalition should seek as much press coverage as possible, though this also may be limited by the strong copyright coalition's superior financial capacity to subsidize information production and the press's tendency to stick

with authoritative sources such as government officials and major industries. This leaves the internet open to near-total colonization by the strong fair use coalition. With opponents having little incentive to go online, not to mention a coalition filled with users and builders of cutting-edge technology, it would be surprising if the strong fair use coalition did not dominate the online debate over copyright law and DRM policy. This gives them a direct opportunity to communicate with the public, seeking to expand the scope of conflict, turn copyright into an electoral issue, and persuade policymakers via their constituents.

Research Questions and Hypotheses

The heart of this study is a quantitative examination of the DRM policy debate as it happened from 1989 to 2006 in Congress and in the pages of *The New York Times* and *Washington Post*, and in more recent years online. Asking whether one coalition has enjoyed better congressional access, I hypothesize that the strong copyright coalition's messages will have been communicated more often than strong fair use arguments in Congress. I also propose that the strong copyright arguments will have been communicated more often than strong fair use arguments in elite newspapers, though their advantage here will not be as substantial as the one they have enjoyed in Congress. Additionally, the formation and growth of the strong fair use coalition give good reason to believe that these proportions will have changed substantially over time. Thus, I predict that the strong fair use coalition's share of messages will have increased over the period under study, in both Congress and the press.

The online debate is a different matter entirely. I predict that the strong fair use coalition will have placed far more information online that is pertinent to the DRM policy debate. Further, I predict that the linking patterns from within the online debate will have strongly favored the strong fair use coalition; following any given link will be more likely to take one to a strong fair use message than a strong copyright message. Finally, I ask whether the online debate will have had a different ratio of strong copyright messages to strong fair use messages than the two offline media studied. The predicted answer is an easy "Yes." The internet will have had a mix of messages far more favorable to the strong fair use coalition than will Congress or the press.

Methodology

As explained in Chapter Four, this study is primarily a quantitative analysis of the DRM policy debate, but it begins with a qualitative case study of the four major debates over DRM policy. For the case study, I provide an overview of each policy proposal, including its contents and any changes as the proposal evolved, relevant external events, and important legislative events, including each proposal's final outcome. This policy case study also tells the story of the evolution of the advocacy coalitions, in particular the birth and growth of what we now know as the strong fair use coalition. This qualitative examination also provides additional context and meaning for the quantitative results that follow.

My hypotheses primarily hinge on the outcome of a quantitative content analysis of congressional hearing documents, newspaper articles, and web documents. Once a document was identified as being relevant to one or more of the debates under study, I

coded it for variables including date and sector(s) represented, as well as the position taken on copyright, which I describe as "rhetorical valence." Documents are coded as representing one or more of 11 sectors. Sector is represented both as a single variable and as 11 nonexclusive variables; the latter was worth the extra effort, as a number of documents—especially newspaper articles—include the voices of more than one sector or represent a jointly authored document. For rhetorical valence, documents were coded as supporting the strong copyright position, the strong fair use position, or a mixed or neutral position. Documents with a mixed valence were then coded by paragraph; each paragraph was coded for relevance and, if relevant, valence. Among documents with a mixed or neutral stance, I quantified each paragraph as follows: "strong copyright" equals one, "neutral or mixed" equals two, and "strong fair use" equals three. I could then average these and get a quantitative estimate of the degree to which a document leans in either direction, and I simply applied the number one or three to documents that were already coded as entirely supporting either position. This means each document's valence can be represented as both a categorical variable (strong copyright, mixed, or strong fair use) and as an ordinal variable (a place on the continuum from strongest copyright to strongest fair use). It is in the latter sense that the term "rhetorical valence" makes the most sense; each document aims either strongly toward either end of the spectrum or toward a point somewhere in the middle.

To identify relevant congressional documents, I used the congressional hearings database in LexisNexis to identify relevant congressional hearings within four-year windows, a time frame chosen because the political science literature suggests this short

window is approximately the duration during which one can command the high level of attention required to enact substantial change. After retrieving all relevant hearings, I coded every document—oral testimony, written supplement to oral testimony, or other written submission—except for question-and-answer sessions. In this manner, I found 17 relevant hearings with a total of 435 relevant documents. I then used the news database in LexisNexis to identify relevant newspaper documents. Despite the surprisingly generous amount of coverage of copyright in general, DRM policy was subject to little coverage— a total of 58 articles from both newspapers across all four policy debates.

Correctly identifying web documents that can speak to the contents of the online debate was understandably much trickier. Yet using well-established characteristics of the web allowed the collection of a set of documents that can be said to represent the online DRM debate with strong authority. The web may be decentralized and seem chaotic, but there is order in the apparent chaos. First, websites tend to cluster thematically, so once one finds a few relevant sites in a topical cluster, the remaining sites are usually within a couple hyperlinks away (Barabási, 2003; Benkler, 2006; Rogers, 2004). Further, across the entire web and within topical clusters, a surprisingly small number of websites get the vast majority of incoming links from other sites, as well as the vast majority of visitors (Barabási, 2003; M. S. Hindman, 2009). Thus, all that is required is a method that can start within a cluster of related sites and find the most authoritative sites—the sites with the most incoming links—within that cluster.

Fortunately, Rogers (2004) developed a method for identifying the authoritative websites within a topical cluster, and his tool for doing so, the Issue Crawler, is available

online (Govcom.org Foundation, n.d.). Starting from a few "seed" sites chosen to represent both coalitions and both nonprofit and for-profit sectors, the Issue Crawler found the sites to which those sites linked, crawled the newly discovered sites, and found sites two links away from the seed sites. I instructed the crawler to repeat the search 12 additional times on a monthly schedule. Each crawl returned nearly one hundred sites, the most-linked of which were consistently at the heart of the DRM policy debate. Looking at the collection of results over this time, I was able to identify the most authoritative sites on this issue. With a list of authoritative websites, I had a population of sites from which to retrieve relevant documents. Here, targeted Google searches proved to be quite useful; using carefully refined sets of keywords, I searched each site for relevant documents, which I immediately saved as PDF files. These were then coded for rhetorical valence, date, and sector represented. Thus, the online debate is represented in the same terms as those used to describe the congressional debate and newspaper coverage, facilitating direct comparisons between the three media.

The quantitative output from the Issue Crawler also provides a treasure trove of valuable data that became part of my analysis. Each site's share of incoming links in any given crawl is a good estimate of its authority within the issue cluster. Thus, each site's average share over the year of crawls—an even better estimate of authority than the results of any single crawl—was used to answer the question about which coalition's websites were more central in the online DRM policy debate.

The DRM Policy Space

This study focuses on four specific debates about the regulation of DRM. Chapter Five provides a modest background into the political machinations that surrounded each of these debates. The first section tells the story of the debate over and ultimate passage of the Audio Home Recording Act (1992) or AHRA, generally regarded as the first US law regulating DRM technologies. The law requires the imposition of a specific form of DRM on digital audio recording devices; this system allows one to make copies of an original recording but not to make copies of copies. The recording industry and songwriters, scared by the invention of the Digital Audio Tape (DAT) in the 1980's, threatened the (mostly Japanese) DAT manufacturers with unending litigation if they imported the machines to the US market. Seeking only to sell their machines without fear of legal liability, the manufacturers quickly agreed to impose this DRM system on imported machines and to pay royalties on the machines and on blank tapes. Faced with a clear inter-industry agreement, Congress passed the AHRA with ease.

The next policy debate was over what became the Digital Millennium Copyright Act (1998), or DMCA. The AHRA was quickly outdated by the evolution of technology, and copyright holders realized the impossibility of prospectively regulating each new media technology with special-purpose DRM requirements. Thus, they sought to give the force of law to any DRM system they might develop in the future. The DMCA gives them this result, making it illegal to circumvent most DRM technologies and to develop or distribute most tools capable of circumventing DRM. Unlike the AHRA, the DMCA was not passed at the unified request of multiple industries. As described more fully in

19

Chapter Five, there was much more opposition to the very idea as well as to several of the specifics, though opponents were unable to stop the bill's final passage.

The third policy debate recounts the attempts to reform the DMCA from 2003 to 2006. These efforts drew a good deal of attention in light of high-profile court cases that started in 2001. The very real threat that computer scientists and programmers could be sued or even jailed for their work helped bring the kind of publicity to the issue that could give credence to the reform effort. More than any single policy issue, the call to reform Title I of the DMCA has helped mobilize and unify the strong fair use coalition. While these efforts came nearest to fruition during this period, the proposals for reform were ultimately rebuffed.

The final policy debate centers on proposals to require that digital television or radio tuners respect a watermark called the "broadcast flag." The proposal to impose this requirement on digital TV (DTV) tuners very nearly came to pass. The Federal Communications Commission (FCC) ruled that this standard should be imposed on DTV tuners beginning July 1, 2005, but a coalition of library groups and NGOs successfully sued to stop the ruling from taking effect (*American Library Association et al. v. Federal Communications Commission and United States of America*, 2005). The court held that the Commission had overstepped its legislative mandate; Congress would need to give the FCC the authority to pass this rule before it could do so. The battle thus moved to Congress, where opposition from the strong fair use coalition solidified and stopped the proposal from becoming law. During this time, a few supporters backed another proposal to impose a similar flag mandate on digital radio tuners, but this proposal did not have

20

nearly as much political backing, technical precision, or realistic shot at passage. Rather, it was an attempt to piggyback on the idea and language of the DTV flag that had gained so much momentum.

<p style="text-align:center">Debating DRM Policy in Congress</p>

The results reported in Chapter Six confirm the hypothesis that the strong copyright coalition's messages were better circulated in Congress than were those of the strong fair use coalition. Across the entire period studied, there were 241 documents calling for stronger copyright compared to 165 calling for stronger fair use. Only 29 documents were neutral or mixed; nearly all participants were clearly in support of one of the two coalitions. Also as predicted, the strong copyright coalition's advantage shrunk over time. From 1989 to 1992, 73% of documents called for stronger copyright, compared to just 23% advancing the strong fair use position. By 1995-1998, the ratio was 56% strong copyright to 41% strong fair use. In 2003-2006, it was basically even, with a very minor advantage to the strong fair use coalition, 46% to 42%. In short, the strong fair use coalition moved from a very small presence to respectability, then to roughly an equal number of appearances.

Counting sectors' representation in Congress helps explain these results. Across all time periods, media companies (123 documents) and technology companies (121) were far and away the best-represented groups, followed by members of Congress (85). Including appointed bureaucrats, the next-best represented group (33), this makes the DRM debate look like a feud between a united media sector and a divided technology sector, moderated by government officials who tend to echo the call for stronger

copyright. NGOs (31 documents), scholars (24), librarians (22), and educators (10)

collectively represent a much smaller voice than either of the for-profit sectors. As with

rhetorical valence, though, this shifted over time. By the 2003 to 2006 period, the media

sector, technology sector, and combined voice of these nonprofit sectors were heard with

roughly the same frequency. As the number of NGOs and scholars within the strong fair

use coalition grew, and as they mobilized and established a consistent voice in Congress,

they helped shift the debate over DRM policy such that now, those calling for a modest

role for copyright have a regular seat at the bargaining table.

Debating DRM Policy in the Press

The results reported in Chapter Seven represent the only findings that were

soundly against one of the proposed hypotheses; contrary to this author's expectations,

newspaper coverage across all time periods was slightly more favorable to the strong fair

use coalition. Where a score of 1 represents strong copyright and 3 represents strong fair

use, the mean valence score for all articles was 2.1, and there were slightly more one-

sided pieces in favor of stronger fair use (16) than in favor of stronger copyright (12).

These represent slight advantages indeed, and the overall tone of coverage in the *Times*

and *Post* was a better reflection of journalistic balance than of either coalition's political

agenda. Also of importance, however, was the very low volume of coverage, and this

favors the strong copyright coalition. Without substantial public attention, the strong

copyright coalition—the governing coalition—was able to effect two substantive reforms

that advanced their goals. In contrast, with just 58 articles spread across two newspapers

and three time periods—an average of just under ten per paper per period—the

newspapers did not help provide enough public attention to help the strong fair use coalition to expand the scope of conflict, produce substantial public pressure, and turn the DRM debate into an electoral issue.

As in Congressional hearings, the strong fair use coalition also saw more clearly sympathetic voices appear with increasing regularity in newspaper articles. In particular, NGOs were quoted in just two newspaper articles from 1989 to 1992 and just one from 1995 to 1998, but appeared in ten articles from 2003 to 2006. These are changes within a small number of articles across all periods, but it reinforces the story of the strong fair use coalition's growth into a major political force, built largely on the birth and development of dedicated NGOs such as Public Knowledge and the Electronic Frontier Foundation.

<center>Debating DRM Policy Online</center>

The results reported in Chapter 8 provide surprisingly strong support for the hypotheses predicting the strong fair use coalition's domination of the online debate. Of the 78 websites that appeared in a majority of the Issue Crawler results—those deemed to be regular, important participants in the online space around this issue—52 had at least one document relevant to this specific debate.[4] Of these, 41 sites were in the strong fair use coalition and just 11 were in the strong copyright coalition. These coalitions also had a sharp difference in the number of documents that were relevant to either the DMCA

---

[4] The high number of sites that did not have DRM policy-specific documents—26 sites— does not necessarily discredit the crawl results. Many sites—such as the Business Software Alliance—had documents related to other aspects of the copyright debate. Other sites—such as Apple and Best Buy—were regularly linked as important examples of a broader debate or as copyright holders' suggestions for places to buy media.

reform debate or the broadcast flag debate. The strong fair use coalition averaged 20 such documents per site, while the strong copyright coalition averaged just 9. Multiplied by the marked advantage in number of sites hosting relevant information, this means something like an eight to one advantage for the strong fair use coalition in total relevant documents. As predicted, the strong fair use coalition put forth much more DRM policy-relevant information online than did the strong copyright coalition.

The patterns of hyperlinks within the online space also substantially favored the strong fair use coalition. Of the top ten sites, accounting for 63% of the share of incoming links, nine had relevant documents that, on balance, advanced the call for strong fair use—eight of them doing so very strongly—and the only exception was a clearly sympathetic site that had nothing to say about the debates under study. These sites have a strong and disproportionate ability to shape the debate, and they all come down cleanly in favor of expanding fair use. Looking at the share of incoming hyperlinks across all sites shows the same pattern. Among all hyperlinks in the set, 85 percent point toward strong fair use sites, 11 percent toward sites without relevant documents, and just 4 percent toward strong copyright sites. Following any given link in this set is more than *forty times* more likely to lead to a strong fair use site than a strong copyright site. These and other comparisons strongly support the hypothesis that websites of strong fair use actors had a higher number of total and average inlinks from within the online copyright policy space than those of strong copyright actors.

Chapter Eight also reports the results of the content analysis along the same lines as those reported for congressional and newspaper documents. Out of 489 documents

coded for valence, 78 percent supported strong fair use, nine percent were mixed or neutral, and just 13 percent supported strong copyright. This large difference is best accounted for by looking at the document counts by sector. Reliably pro-fair use sectors did the lion's share of the work of writing and posting online documents; in particular, scholars authored 160 documents and NGOs authored 140. With 39 documents, libraries came in fourth. In third place was the technology sector, with 67 documents. In offline media, technology groups were often divided on the wisdom of DRM regulations, but only those companies and groups calling for stronger fair use participated regularly online—authoring 62 of the sector's 67 online documents. In contrast, reliable supporters of stronger copyright such as the media sector (28 documents) and appointed government officials (22) used the web quite sparingly. This is a huge numerical advantage for the strong fair use coalition; those who go online to learn more about DRM policy are getting a very one-sided story because the other side mostly ignores the online debate.

Comparing Media

Taken together, Chapters Six through Eight imply substantial differences between congressional communication, newspaper coverage, and the online debate over DRM policy; Chapter Nine systematically explores these differences. Across all three time periods, newspapers lean slightly in the direction of calling for stronger fair use, while congressional documents lean toward calling for stronger copyright; the difference between them is modest but not trivial. These media are not radically dissimilar, but the difference is important.

There are more substantial differences between the either of the two offline media and the population of web documents. The web component of the debate is a relatively recent phenomenon, so in order to ensure timely comparison, I only base these comparisons on documents dated from 2003 to 2006. By large margins, the web is far and away the medium that leans most solidly toward calls for strong fair use. Comparing web documents to newspaper articles shows a medium to large difference, and the same comparison between web documents and congressional documents shows a much larger difference than is typical in social scientific research. These differences and their respective estimates of effect size hold whether one compares rhetorical valence as a categorical or an ordinal variable.

Conclusion

The copyright debate in general and the DRM policy debate in particular may seem hopeless to many advocates of moderate copyright protection—including Lawrence Lessig, who knows the policy space far better than this author. From the perspective of the strong fair use coalition, this dissertation tells a somewhat more optimistic story. This includes the rise of a meaningful strong fair use coalition, a substantial pro-fair use shift in the congressional debate and in the press, near-total domination of the online debate, and a substantial slowing in the march toward ever-stronger copyright.

The strong fair use coalition still has less power than their opponents to pass laws; they still spend a great deal of their energy playing defense against proposals to expand the reach of copyright. They have at least risen to a point of meaningful political capital, however, and this is a substantial change from 20 years ago, when the DRM policy

26

debate began in earnest. An important part of this story is their successful, creative uses

of the web as a political tool. Because of what it says about copyright, but also because of

what it says about the internet's potential as a political tool, this is a story worth telling.

CHAPTER TWO: TOWARD A THEORY OF COMMUNICATING POLICY ACTORS

Many political commentators begin from the assumption that moneyed special interests drive all political decisions. Even serious scholars of the policy process, who generally take a more nuanced stance, have documented how well-funded special interests are disproportionately successful, using campaign donations, advertising, and public relations to drive their policy agendas (see, e.g., West & Loomis, 1999). Substantial evidence illustrates at least some degree of quid-pro-quo trading of policy outcomes for cash. For instance, Cialdini (2001) recounts "the remarkably bald-faced admission by businessman Roger Tamraz at congressional hearings on campaign finance reform. When asked if he felt he received a good return on his contribution of $300,000, he smiled and replied, 'I think next time, I'll give $600,000'" (p. 26). Over at least the past decade, the communications industries alone have devoted millions of dollars in campaign contributions and lobbying resources (Center for Public Integrity, 2006), which is certainly a contributing factor in their significant political influence (McChesney, 1999, 2004b).

Nonetheless, the interest groups that spend the most do not always get what they want. Sometimes, they propose policies that are not adopted. Further, one must account for the rare but recurring shakeups that repeatedly change laws and policies in ways that are contrary to the desires of formerly dominant interest groups. Therefore, a theorist who tries to explain US political outcomes would do well to examine more independent and intervening variables than net cash spent by interested parties. Several theories attempt to

shed some light on policy stability and change; in the first section of this chapter, I draw on three such theories: iron triangle theory, punctuated equilibrium theory, and the advocacy coalition framework.

What these theories do not systematically explicate is the role of the media in policy outcomes; in the second section of this chapter, I introduce some theories of mediated communication. First, I describe four of the media's most important political effects: policy learning, agenda setting, framing, and persuasion. Second, I introduce theories of media economics and information subsidies as a starting point for explaining who has access to the media. Finally, I observe that media attention is, like human and institutional attention generally, subject to cycles of punctuated equilibriums.

This is a lot of theory to have on the table, and in the third section of this chapter, I integrate and update it. I weave together the policy theories with the theories of mediated communication to yield an enhanced model of policy change and stability, featuring a stronger estimate of the media's role, as well as predictions of different coalitions' media strategies. Then, I introduce additional literature that considers how the new information environment created by the internet introduces subtle changes to this integrated theory.

Theories of policy stability and change

This project draws on three traditions of research into policy stability and change. I take the iron triangle as a conceptual and historical starting point, add elements of the advocacy coalition framework, and include the general punctuation hypothesis.

The Iron Triangle

U.S. political scientists often use the term "iron triangle" to describe the cozy relationship between congressional committees, administrative agencies, and the interest groups most affected by their policies. According to some commentators, this theory was at least reasonably appropriate in describing US political system in the middle of the 20th Century, even if that system became markedly more complex—and the theory thus less applicable—in later decades (Gais et al., 1984).

In the U.S. political system, most policy topics at most times will be of interest to a small number of policymakers with specialized expertise and oversight on that issue (Jones & Baumgartner, 2005, p. 39). House and Senate committees specialize, consider topical bills, and oversee the relevant administrative agencies. Interest groups can therefore easily locate and lobby the policymakers who most control their fate, providing electoral support (including campaign donations) to helpful committee members. Constituent groups also lobby Congress on behalf of helpful (and against disagreeable) agencies. In most cases, the theory of the iron triangle would imply, the rest of the legislative and executive branches will pay little attention to this cozy relationship, and a hardened three-way bond of mutual indebtedness will grow over time (see Figure 1). This is a specific theory of subgovernments, the likes of which reach back to the early 20th Century (John, 2001), though the triangular metaphor is today virtually synonymous with theories of subgovernments.

Figure 2.1: The Iron Triangle[5]

Congress

electoral support

funding & political support

friendly legislation & oversight

policy choices & execution

congressional support, via lobby

Interest Group

Bureaucracy

low regulation / special favors
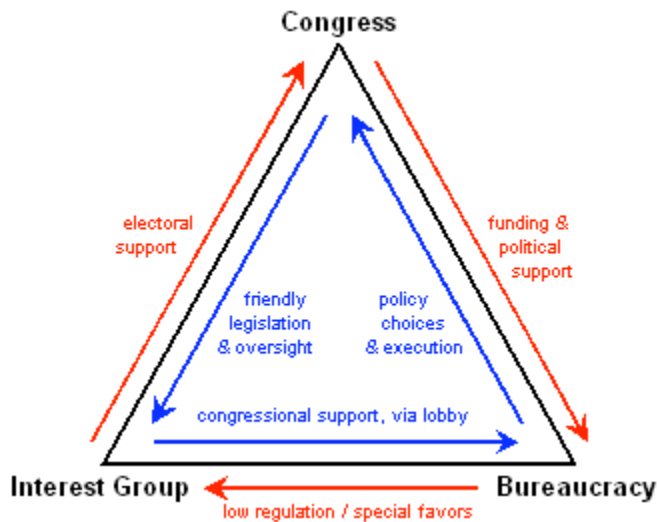
Baumgartner and Jones (1993) observe that most policies are stable most of the time, reflecting "domination of important policy areas by privileged groups of elites" (p. 3). Iron triangle dynamics help explain this stability. Once the members of the triangle decide on the core parameters of a policy—to the benefit of the interest groups in question, of course—those parameters will tend to stay stable as long as the policy topic stays under the purview of the legislators and bureaucrats who helped fashion it. Especially in systems with the very slow turnover characteristic of US federal policymaking, the same policymakers can preserve the fundamental character of the status quo for decades on end.

---

[5] Wikipedia user Ubernetizen donated this graphic to the public domain (*Iron triangle*, 2006).

Iron triangle dynamics can also lead to some major policy changes. For instance, as a result of a legal settlement, tobacco companies agreed to fund a popular and successful pilot program to discourage youth smoking in Florida. Nonetheless, tobacco companies lobbied the state legislature and administrative agencies to dismantle this program in 1999. Because this issue was decided from within the iron triangle, a quick policy change was enacted to the benefit of the tobacco companies and at the expense of public health (Givel & Glantz, 2000).

Many political scientists long assumed the validity of iron triangle explanations (Browne & Paik, 1993), but most now believe things are much messier.[6] Policy change does sometimes occur in ways that flout the wishes of the most resourced interests, and on almost no issues do the most resourced interest groups ever lobby without encountering substantial opposition from other organized groups. Therefore, I also draw on other research traditions that better account for both policy stability and change; the first of these is the advocacy coalition framework.

---

[6] To some degree, this shift in US political scientists' theoretical outlook has reflected genuine on-the-ground changes resulting from Congressional reforms in the early 1970's; these reforms greatly reduced the power of committee chairs and therefore opened the door to regular and substantial influence by grassroots groups and voting constituents (John, 2001).

The place of iron triangle theory in US political science shares a good deal in common with the place of the hypodermic model in studies of media effects (Lubken, 2008). Both present extreme, simplistic causal relationships between the important variables under study. This makes each a convenient jumping-off point; in comparison, nearly all theories seem nuanced. Yet while these theories both serve this role, there is a substantial difference in their historic place within their home fields. Scholars earnestly developed the iron triangle theory of sub-governments beginning in the early 20th Century (Browne & Paik, 1993). As illustrated above, scholars still apply it in earnest to explain policy outcomes. In contrast, Lubken intimates that the powerful effects model underlying the hypodermic metaphor was born a straw opponent.

The Advocacy Coalition Framework (ACF)

The advocacy coalition framework is built around the study of coalitions of policy

actors within policy subsystems. In this section, I first define the framework's basic

terms. Second, I identify ACF hypotheses regarding the means for policy change.

Basic Terms of the ACF

Coalitions are "composed of people from various governmental and private

organizations that both (1) share a set of normative and causal beliefs and (2) engage in a

nontrivial degree of coordinated activity over time" (Sabatier and Jenkins-Smith, 1999, p.

120). The ACF argues that policy analysts can most usefully define policy subsystems as

"those actors from a variety of public and private organizations who are actively

concerned with a policy problem or issue, such as air pollution control, and who regularly

seek to influence public policy in that domain" (p. 119). Within a mature subsystem, the

lineup of coalitions—generally two to four—will tend to stay stable over time, and

"actors within an advocacy coalition will show substantial consensus on issues pertaining

to the policy core" (p. 124), or essential elements of the policy in question.

This is a broadened definition of policy subsystems, relative to the iron triangle of

congressional committees, implementing bureaucracies, and interest groups at one level

of government. The ACF presents a pluralistic view of organized interest groups,

expanding the class from a tightly knit group of well-financed business interests to

include grassroots organizations and other nongovernmental organizations. Additionally,

the framework acknowledges that well-financed business groups often have diverse

interests. In many, perhaps even most subsystems, powerful corporate interests will take

33

conflicting stands. Further, the ACF adds "two additional categories of actors: (1)

journalists, researchers, and policy analysts, who play important roles in the generation,

dissemination, and evaluation of policy ideas... and (2) actors at *all* levels of government

active in policy formulation and implementation" (p. 119).[7]

Means of Policy Change

      As with other theories of policy subsystems, the ACF assumes that major policy

shifts are rare and generally not endogenous to policy subsystems. Persuasion and

learning among a subsystem's players can lead to moderate policy changes. Major shifts,

however, generally require some sort of external event, such as "changes in

socioeconomic conditions, public opinion, systemwide governing coalitions, or policy

outputs from other subsystems" (Sabatier and Jenkins-Smith, 1999, p. 147). The

possibility for external events leading to change is obvious and commonly acknowledged

among many theories outside the ACF. Baumgartner and Jones (1993) explain that a

change in venue—especially a change to a higher level of government, e.g. when the

federal government usurps issues formerly decided by the states—is often associated with

policy change (p. 33). Kingdon (2003)(2003) also argues that outside perturbations, such

as shifting social or economic conditions, create rare "policy windows, the opportunities

for action on given initiatives, [which] present themselves and stay open for only short

---

[7] Sabatier and Jenkins-Smith (1999) put a good deal of stress on what they see as a
hierarchical tripartite of policy beliefs: deep core beliefs (such as those at the root of the
familiar left-right political scale), policy core beliefs (the normative questions at the heart
of a given policy debate), and secondary policy beliefs (less essential beliefs about which
policy actors are more likely to change their mind). This study does not depend upon this
specific hierarchy of beliefs.

periods" (p. 166).[8] In other words, "external perturbations are a necessary but not sufficient cause of change in policy core attributes" (Dudley & Richardson, 1999, p. 227).

These outside perturbations cannot, by themselves, force policy change. Rather, policy actors must translate these outside forces into specific policy outcomes. Kingdon (2003) cites an interest group analyst who uses the metaphor of the surfer waiting for the next great wave. "You get out there, you have to be ready to go, you have to be ready to paddle. If you're not ready to paddle when the big wave comes along, you're not going to ride it in" (p. 165). Kingdon's term for the metaphorical surfer is the "policy entrepreneur" who "advocates for proposals or for the prominence of an idea" (p. 122). For example, European regulation of the steel industry has often waxed and waned in light of steel prices, but not because everyone has agreed to such oscillations. Rather, falling prices have opened a window of opportunity for would-be deregulators—and rising prices for would-be regulators—and policy entrepreneurs have historically succeeded at using market forces as a spark for driving policy changes (Dudley & Richardson, 1999).

The ACF tells us a great deal about the process of paddling out into the ocean and waiting for the next big wave, though less about the translation of outside perturbations

---

[8] Citing Kingdon here implies that he is an ACF theorist. This is not the case. In fact, Kingdon's work preceded that of Sabatier and Jenkins-Smith by a number of years (Kingdon, 2003). Nonetheless, I am not the first to imply through such placement that at least some of his theory provides a nice supplement to the ACF (see, e.g., Dudley & Richardson, 1999, p. 227).

occur into policy core changes. The general punctuation hypothesis is an important contribution that serves this role.

Information Processing and the General Punctuation Hypothesis

On most policy issues, substantial reform never exhibits the smooth, automatic adjustments characteristic of a thermometer. Rather, major change is unthinkable for long periods of equilibrium, each of which is punctuated by brief policy windows during which major change is very possible. Extending their earlier work in which they described this cycle as one of punctuated equilibriums (Baumgartner & Jones, 1993), Jones and Baumgartner (2005) advance the *general punctuation hypothesis* (pp. 17-20): The dysfunctional means by which humans process information, combined with gridlock-producing institutional arrangements, create the cycle of long periods of stability and brief windows of radical change.

Making Decisions and Human Cognitive Limitations

The authors identify three cognitive limitations that lead to dysfunctional information processing. First, they peg the "bottleneck" of selective attention as "far and away the most important cognitive limitation that influences political choice" (Jones and Baumgartner, 2005, p. 16). This bottleneck forces policymakers to leave most issues alone most of the time. Once an issue intrudes on their attention, they tend "to overreact with 'alarmed discovery'" (p. 52) at the sudden recognition of "new or previously overlooked information" (p. 55).

A second cognitive limitation is the tremendous difficulty people have conceptualizing trade-offs (Jones and Baumgartner, 2005, p. 16). Decision makers, whether citizen or elite, have no exchange rate with which to negotiate competing values. This also leaves a door open for disjointed overreaction due to "attribute intrusion," or alarmed overreaction to previously unconsidered angles that become highlighted by "the process of issue redefinition" (p. 55).

The final limitation on cognitive processing is "the manner in which people encode and recover information" (Jones and Baumgartner, 2005, p. 16). People learn and remember information via rules, "and the application of rules to new situations is governed by the ability to view the new situation as analogous to the one in which the rule was learned" (p. 17). This information processing strengthens the bias toward the current policy, which was generally crafted in light of the cognitive rule that policymakers will again deploy in reconsidering the matter.

Institutional Limitations

In addition to cognitive limitations, Baumgartner and Jones (1993) identify institutional biases that lead to punctuated equilibriums. Policymakers perform tasks based on division of labor—the division of the legislature into committees and subcommittees, the division of the executive into departments and agencies, and so on. This helps solve the problem of the lack of attention, giving policy institutions the ability to process multiple issues simultaneously, but it also creates a tendency toward the status quo within policy venues. Committee members and agency employees become experts on

specific issues, forming strong beliefs on a given policy question—cognitive rules that are particularly resistant to change, even in the face of new evidence.

If a disfavored coalition wants to change the core attributes of a policy, they "always have the option of trying to change the policy venue from, say, the national government to subnational units, or from so-called iron triangles to election politics, and such efforts are a constant part of the policy process" (Baumgartner & Jones, 1993, p. 34). In other words, policy actors who are not enjoying success in one venue have an interest in expanding the scope of conflict in search of more favorable venues, including the court of public opinion (Schattschneider, 1960). When they (occasionally) succeed, the policy actors in the newly involved venue focus on the issue with alarmed discovery and are likely to implement a policy disproportionately divergent from the status quo.

For all their value, the ACF and general punctuation hypothesis offer too little that can help us explain and predict how policy actors will communicate. By combining theories of mediated communication with the policy systems literature, I hope to better describe the communication strategies that coalitions will adopt and their expected effects.

Theories of Mediated Communication

The policy systems literature provides some useful theories about how advocacy coalitions will seek to use the media; I draw on these, but I also introduce communication theory about media access and the expected effects of media on the public. First, I argue that media content has several noteworthy effects on audiences: learning, agenda setting, framing, and persuasion. Second, I contend that different types of coalitions will likely

enjoy different kinds of access to the media due to the dynamics of media economics and information subsidies. Third and finally, I argue that patterns of media coverage are similar to those of political attention generally, featuring similar division of labor and cycles of attention.

Why Media Coverage Matters: Theories of Media Effect

For decades, the common wisdom in both political science and media studies reduced the scholarly estimate of the role of the media in the political process. For about the first two-thirds of the 20th Century, political scientists held the view that policies were negotiated from inside autonomous subgovernments, or iron triangles, which insulated policy decisions from the influence of outsiders—including the media. Likewise, in media studies, "decades of academic research into the effects of media-based political campaigns purports to demonstrate that exposure to campaigns mainly reinforces voters' preexisting partisan loyalties" (Iyengar & Simon, 2000, p. 150). These theoretical starting points assume a greatly diminished possibility for media to play an important role in sparking profound political change.

In contrast to this mid-century consensus that media have less power than that ascribed to them by the popular imagination, more recent political science and communication research has continued to open a larger theoretical space for studying the media's role in shaping policy outcomes. The recent policy subsystems theories noted above leave more room for influence by outsiders, including the media and mass audiences, and recent communication theory argues that mediated messages have substantial effects on their audiences. Mediated messages cause policy learning, set the

policy agenda, frame the issues, and persuade some audience members to change their prior attitudes or behavior.

Learning

First, media teach audiences about policy issues. In summarizing the available evidence, Iyengar and Simon (2000) conclude that mediated political campaigns actually do teach voters about candidates. "Contrary to the prevailing wisdom, the information they yield is multifaceted, encompassing the candidates' chances of winning, their personal traits and mannerisms, and most important, their policy and ideological bearings. Media campaigns may appear superficial, but they do educate citizens" (p. 156). For those who are learning about an issue rather than actively working within the policy subsystem, beliefs about as yet unstudied policy issues are much more malleable than pre-existing partisan beliefs, and this difference in malleability plays out in policy learning. Upon learning more via the mass media, voters who were previously unfamiliar with an issue will often identify and adopt their party's viewpoint on the matter (Lenz, n.d.).

Agenda setting

As a second important effect, the media are incredibly effective at setting the public's policy agenda. Like Jones and Baumgartner (2005), McCombs (2004) observes that attention is a very limited resource for reasons both psychological and institutional (p. 38). McCombs also paints a similar picture of the human attention cycle, including the moment of "alarmed discovery" (p. 29). While this may happen in one's personal life

when something unexpected happens, "the news media set the public agenda" (p. 2) to a degree unrivaled by other social institutions. If one wants to move the public to care about an issue, then, one generally must engage the media.

Most news media outlets follow the lead of a few opinion-leading news outlets, amplifying their agenda setting function. For most national and international stories, US news outlets look to the major national daily newspapers: the *New York Times*, *Washington Post*, and *Wall Street Journal* (McCombs, 2004, pp. 113-114). On specialized issues, they also look to specialized publications—for instance, following the lead of well-known science periodicals in their coverage of scientific issues such as global warming (McCombs, p. 114). This process is called inter-media agenda setting.

Framing

In addition to teaching audiences and setting the public agenda, media help audiences to make sense of the world by casting issues within frames, the "*organizing principles* that are socially *shared* and *persistent* over time, that work *symbolically* to meaningfully *structure* the social world" (S. D. Reese, 2001, p. 11). Commonly accepted frames, such as the depiction of elections as a contest between candidates who will do anything to win (Kerbel, Apee, & Ross, 2000), become interpretive schema through which individual news stories and the general public make sense of new information. For instance, because a great deal of election coverage frames elections as horse races, audiences have come to view all politicians' actions as geared toward winning elections rather than as geared at solving social problems, leading to increasing cynicism (Cappella & Jamieson, 1996).

41

Framing depends both on broader social contexts and those forces pushing competing frames. Pan and Kosicki (Pan & Kosicki, 2001) advocate framing analyses that study "two important dynamic interfaces: between the ideational and the sociological dimensions of the framing process, and between and among various actors, as well as their narratives" (p. 49). Without avowing such, their study fits nicely within the ACF tradition.

Persuasion

As a fourth effect, media messages can persuade some audience members to change their beliefs or behavior, at least in the absence of comparably strong counter-messages. Zaller (1996) "maintains that, at least in the domain of political communication, the true magnitude of the persuasive effect of mass communication is closer to 'massive' than to 'small to negligible' and that the frequency of such effects is 'often'" (p. 18). Many effects scholars have studied campaigns, especially presidential campaigns, which "are characterized by roughly equivalent amounts of opposing communication. Thus, the effect of each side may cancel the other out" (Jamieson, 2000, p. 10). In contrast, Zaller (1996) seeks to illustrate powerful effects by locating instances of information imbalance, where one campaign is several times more visible than the opposing campaign. In these circumstances, sizable portion of respondents will exhibit a "reception gap," having gotten one message but not the other (pp. 23-26). This research illustrates that the relationship between media consumption and persuasion is curvilinear. Audiences that consume a middling amount of media will be most likely to get one message but not the other and therefore be persuaded by the louder campaign.

Jamieson (2000) concurs, noting reception gaps benefit advocacy coalitions who can afford television advertising. "When the issue is complex, reporting on it deficient, adwatching minimal, and one side substantially outspends the other, ads … can influence public attitudes" (p. 140). As an extended example, she illustrates the marked success of the tobacco lobby's $40 million advertising campaign, aimed at de-clawing and then defeating unfavorable legislation in 1998. The campaign provides "a textbook example of the power of saturation advertising, [which] allowed the industry to reshape the debate" (p. 136). Advertising, however, is not the only means by which advocacy coalitions seek to communicate with the public, and those with more capital are also better able to use the other means of communication.

Media Economics and Information Subsidies

While the above theories describe media effects, a complete theory of media also includes the economic systems of production (McChesney, 2004a), including questions of who has access and under what terms. Gandy (1982) urges scholars to do so by examining the economics of information. Gathering and analyzing information costs money. Hence, "all who have an interest in such decisions have an incentive to influence the prices faced by others for information related to those decisions. Efforts to reduce the price faced by others for certain information, in order to increase its consumption, are described as *information subsidies*" (Gandy Jr., 1982, p. 8).

Groups that seek to influence policy outcomes will regularly produce policy reports and other documents intended for consumption by policy makers, resources permitting. The practice is so common that a "clamoring diversity of information ...

characterizes modern America. Information on policy matters is supplied by interest

groups, think tanks, political parties, bureaucratic agencies, and congressional

committees" (Jones and Baumgartner, 2005, p. 9), each of which seeks to shape policy

decisions. Naturally, those with more capital can better subsidize policy information,

providing them with an important political advantage.

Information subsidies also shape the content of mass media. Guided by the *rule of*

least effort, Gandy (1982) notes, nearly all news outlets rely quite heavily on bureaucratic

channels for a solid majority of their sources (p. 12). Even though many journalists would

prefer to do original investigative journalism, it makes too much business sense from

management's point of view to rely on information subsidies. "The public-relations

industry [is] now bigger than the press itself" and "provides press facilities, news

conferences, press releases, video news releases, and other similar materials" (Baker,

2002, p. 54). A news outlet can process several ready-to-use news releases for a fraction

of the cost of one investigative story.

Information subsidies have the intended effect: a tremendous advantage for the

subsidizers. Elected officials, government agencies, and corporate public relations

departments dominate news coverage (E. S. Herman & Chomsky, 2002, p. 18) and

therefore dominate the news agenda (McCombs, 2004, pp. 102-113). Even in

investigative pieces, interviewers still rely on these official sources of information

(Gandy Jr., 1982, p. 12). Information subsidization has successfully shaped the majority

of news content for decades (Sigal, 1973). "Without the subsidies routinely provided by

public relations professionals in the public, non-profit, and private sectors, the media

agenda would be considerably different in scope and content" (McCombs, 2004, p. 103),

and the level of reliance on information subsidies is only rising (Davis, 2002, pp. 32-33).

Most civil society groups produce a fraction of the public relations material and

other subsidies, which means that including their voice in the media is comparatively

more expensive and therefore happens less frequently. By subsidizing the cost of

information, the same interests that generally dominate their policy spaces are also

generally able to guide media coverage relevant to their policy subsystems.

Of course, few coalitions can dominate indefinitely. Most core policy changes

come, in part, when challenger coalitions win the media game. Not only is this often part

of their political strategy, in many important respects, these shifts in media coverage

resemble shifts in policy.

Punctuated Equilibriums of Media Attention

The media cover issues in a pattern of punctuated equilibrium, subject to issue

and attribute intrusion as described above (Baumgartner & Jones, 1993, pp. 103-108).

This is not least of all because media institutions have much in common with the policy

institutions described above. Like policymakers, editors and reporters have a narrow

bottleneck of attention, and media institutions divide their labor into specialized sub-

units.

As with policymaking institutions, there is also a division of labor between media

institutions. Major national news outlets disseminate national and international news to

the masses. Many people rely heavily on local news outlets, which devote most of their

reporting resources to local issues and rely on national outlets for news from afar.

Specialized publications have the ear of smaller audiences, such as policymakers or investors (Gandy, 1982). While there is some overlap—DC voters and US Senators alike read *The Washington Post*—there is a clear difference between the content and audience of *USA Today* and that of *CongressDaily*.

Different media outlets and reporters on different beats generally cover different news stories or frame the same stories differently. Policy actors therefore shop for media venues in pursuit of the friendliest possible decision makers. "For the strategically minded policymaker … there are many different types of reporters" (Baumgartner & Jones, 1993, pp. 108-109). This is especially important when another coalition has framed an issue to its benefit in a sympathetic outlet. "One of the most important allies in this process may be a sympathetic reporter who shares the source's interest in generating some new controversy where previously there had been little attention" (p. 109).

One important kind of shift is when a story moves from slow but steady coverage in the business section to a swamp of coverage in the national or local news section. The Enron collapse is a fine example. The company collapsed over the course of 2001, a story covered by the business press and framed as a large but cognizable corporate bankruptcy. Beginning in December of 2001, however, the story gained traction as an item of major national news. It came to stand for "exploitation of consumers (with the company's manipulative energy policies), exploitation of workers by selfish management, and political corruption" (Jones & Baumgartner, 2005, p. 73). Instead of a concern for

employees and investors, it was successfully reframed[9] as a political outrage, and the public took notice. The story moved into media venues with larger audiences and took on a new meaning that would have been difficult in the business pages.

This pattern occurs repeatedly as the media jumps from one politically contentious issue to the next, from Microsoft's business practices (Bennett & Manheim, 2001), to corporate accounting scandals, to proposals for network neutrality (B. D. Herman, 2006b). In each instance, the coverage moves from modest coverage in the business pages to heavy coverage in the national news section. News can also explode by jumping from specialized or local media to national media outlets. This highlights the ACF contention that journalists count as policy actors; sympathetic newspersons can spark an explosive shift of attention.

Synthesizing, updating, and building a new research agenda

This project builds on a unique combination of previously unconnected theories and research traditions, so it is important to integrate these into a cohesive whole. First, this section builds an integrated theory of policy subsystems, synthesizing theories of policy stability and change with theories of communication. Second, it incorporates newer theories of political communication that account for the explosion of digital media technologies such as the internet.

---

[9] Jones & Baumgartner unfortunately fail to assess who was behind such a push, describing its timing as "a mystery" (p. 73). The timing is explicable, a possibility not properly suggested by the term "mystery."

Synthesizing Theories of Politics and Communication

This section presents a synthesis of the above theories, combining them into a theory of communicating policy actors. First, I synthesize the actor theory discussed herein. Second, I consider how theories of mediated communication matter for policy actors and outcomes.

A Theory of Policy Actors

Starting from but adding to the image of the iron triangle, the advocacy coalition framework (ACF) and the general punctuation hypothesis give a fuller picture of policy subsystem dynamics. The combination of all three might look something like this: Coalitions are formed to fight for policies that most closely resemble their policy beliefs. Policy subsystems generally feature one to four advocacy coalitions, each of which may include interest groups, administrators, members of Congress, journalists, researchers, and other policymakers from other levels of government. Policy subsystems will also generally feature policymakers who are not members of any advocacy coalition; these policy brokers can provide enough push to cause policy changes on secondary policy aspects, but political forces outside the subsystem will generally be required to enact a change in the core aspects of a policy. In most cases, the coalition whose beliefs most closely resemble the status quo will deserve most of the credit for the policy changes that implemented it; without the coalition pushing for it, the policy would not have come to be. In these circumstances, a governing or dominant coalition defends the status quo, and any coalitions seeking major changes are challenger coalitions. In most subsystems, the interest group members of the governing coalition will have spent greater resources on

48

the policy process than the interest group members of any challenger coalitions in the same subsystem.

Most of the time, the policy issue at stake will be of little interest to the broader political arena or the general public; outsiders' attention will generally be elsewhere. This greatly benefits governing coalitions, so they will generally seek to draw only specialists' attention to an issue. In contrast, challengers will generally try to incite broad public debate that can undermine or at least problematize the status quo, because substantial attention from those outside the subsystem is a necessary component of substantial policy change. They will also seek to shift venues, such as from the state to the federal level, or to a different congressional committee. Challengers especially benefit from moving the debate from a less visible venue (e.g., a congressional subcommittee) to a more visible one (e.g., the House floor). Such a change gives them a chance to seek out support from other policymakers *and* to bring broader public attention to the issue.

The scarcity of attention and institutional bottlenecks will also obstruct governing coalitions from achieving certain policy changes, especially in the form of legislation. If they can win the support of policy brokers from within the subsystem, those brokers might pave the way for a policy change to move through broader political processes. Yet this may be difficult or insufficient. In those instances, governing coalitions also have an incentive to seek a shift in venue, but it is still not in their interests to shift to a venue that brings heightened visibility to the issue, a process that may invite broader policy debate. Governing coalitions will thus seek to shift to low-visibility, special-purpose venues in which they enjoy a strong political advantage. International forums, such as trade groups

49

and collective defense organizations, often meet these criteria. A coalition can sidestep domestic institutions and push for a favorable international treaty or ruling, allowing them to push national policymakers in the name of compliance and harmonization. This process is called "policy laundering" (Hosein, 2004, p. 187).

A Theory of Communicating Policy Actors

At every stage of the policy advocacy process, policy actors are communicating—within subsystems, with policymakers from outside their subsystems, and with the general public. Some communication strategies will be advantageous for all advocacy coalitions, while some will benefit certain coalitions much more than their opponents. Despite the importance of communication in shaping policy outcomes, however, policy scholars have also given unsatisfactory attention to the role of information and communication in the policy process (Bimber, 2003, p. 12). Thus, predicting the strategic communication decisions of policy actors requires the merging of theories from policy studies and communication, the purpose of this section.

As Gandy (1982) suggests, all policy actors benefit from supplying policymakers with favorable information. Attempts to communicate directly with policymakers include: letters, emails, and telephone calls to policymakers; testimony before policymaking bodies; research reports; and communication with certain elite media, such as the *Washington Post* and the National Journal Group's publications. In other words, all policy actors will view policymakers as important potential audience members. Other factors being equal, the coalition with the greatest ability to communicate its message to policymakers is the coalition most likely to shape policy outcomes. This effect may be

reduced when a policy issue becomes a major concern on the public's agenda; a mass public outcry can sometimes outweigh deep political connections, though this is rare.

Mediated messages are an important part of the ability to communicate within subsystems. While one might assume that nearly all of the important communication between elites happens privately, Kingdon (2003) documents that "media act as a communicator within a policy community" (p. 59). A non-story can become important simply by being covered in a major newspaper (Kingdon, pp. 59-60). Those who have private access to policymakers also use these direct channels. In short, elite media coverage matters for all policy advocates.

For all these commonalities in strategy, a corollary of the general punctuated equilibrium hypothesis is that challenger coalitions have much greater incentive to pursue the widest audience possible, especially in starting a debate about the core aspects of a policy. If the policymakers within a subsystem have already reached their conclusions about a given policy choice, anybody seeking a different choice needs either a vehicle to force those policymakers to reconsider or the ear of different policymakers with some power over the issue. This is done via expanding the scope of conflict (Schattschneider, 1960), generally including an attempt to foment broad public interest and outrage.

Kingston (2003) quotes one activist who views herself as an outsider looking in: "'If there is a strong organized interest in keeping the status quo, you have to overcome it somehow. Your only hope is to go public'" (p. 61). For those who already have preferential access to policymakers, however, media access is also assumed and is just one tactical option. Kingdon quotes one such policy actor who elaborates:

51

> Mass media coverage is not critical. It is one of your vehicles. … We have alternatives of leverage on the system, and we don't have to use the media very much. The media will follow us because what we do makes news. If we get a proposal sold to the White House and it goes to the Hill, then that's news. So I don't think the media adds much. If I weren't here, though, I would worry about it. If I didn't have the levers I have, the media would be very important to me, and I would use it as much as I could. (p. 61)

Members of dominant and challenger coalitions have different reasons to use the media, and media coverage is of much greater importance to challenger coalitions. Inversely, this suggests that dominant coalitions will enjoy relatively greater access to policymakers and be better represented in official policymaking venues, especially venues where participation is determined by policymakers.[10]

Additionally, challenger coalitions are generally much more likely to benefit from news coverage that reaches the ordinary citizen; such coverage helps place the issue on the public's agenda, a potential avenue for punctuating an equilibrium. Internet communication is an excellent means for cheaply and easily reaching ordinary citizens, a point I discuss presently.

Revolutionary Media! Revolutionary Results?

The internet has arrived amidst sweeping proclamations about its inescapable, irreversible impact on society. Many authors (see, e.g., Gillmor, 2004; Hewitt, 2005;

---

[10] Formal gatekeeping is not the only or even necessarily the most important means for reducing the ability of issue publics to engage in face-to-face contact with policymakers. Herman and Gandy (2006) found heavy participation by ordinary citizens in the written submission stages of the 2000 and 2003 hearings to determine triennial exemptions to the DMCA ban on circumventing access-controlling TPMs. In contrast, the live hearings, requiring the time and expense of travel, featured a much higher ratio of industry representatives.

Trippi, 2004) triumphantly celebrate the internet's ability to bypass the "monopoly" of old media. These claims may be overblown, but the internet has lead to at least five politically important changes in the information ecology, which this section reviews in order. First, the internet changes the economics of the production and distribution of information. Second, it facilitates more direct communication with and mobilization of issue publics. Third, it often allows less well-funded groups to define an issue for the general public. Fourth, it sometimes affects offline media coverage. Finally, it inflates the absolute number of mediated messages, reducing the impact of each individual act of communication.

Reduced Costs for Producing and Disseminating Information

The first effect of the mass adoption of the internet is the reduction in costs for producing and distributing mediated content. As Bimber (2003) argues, the internet provides "a multiplication of low-cost channels for the distribution of information by political elites and organizations" (p. 90). This reduction in the cost of producing and disseminating information places increased communication power in the hands of people who lack substantial resources (Benkler, 2006). This increased communication power facilitates dramatically improved communication for less well-funded policy actors, which creates both more effective communication within a coalition and with the general public.

For intra-coalition communication, online media offer cost-effective tools for under-resourced civil society groups, interested scholars, libraries, and other interested citizens to combine their efforts in building an issue network (Marres, 2006). This

includes information sharing, message development, and agenda setting. This part of policy advocacy is often under-valued, but it represents the infrastructural backbone of more public forms of advocacy. New media have come to play an increasingly important role in the formation of issue networks. Marres (2006) supposes, "the rise to prominence of the internet may be *responsible* for the fact that [civil society organizations] increasingly organize themselves as networks" (p. 13). In short, online communication greatly increases the capacity for coalition building among lesser-resourced organizations.

The internet also decreases the cost of communicating with the broader public. As Bimber (2003, p. 100) illustrates, capital gives one the power to communicate one's message to a broad audience, but the introduction of the internet means that each actor's communication power goes up in absolute terms. More capital still buys more communication power, but the internet provides a potentially powerful medium for anybody with a networked computer—powerful as long as one has or develops an audience. As with intra-group communication, this has the greatest relative impact on the public communication of political and social organizations that had the least capital to start with, increasing both the absolute and the relative communication power of these less-well-funded coalitions.

This is not to say that computers and internet connections are free or that their adoption is as total in the US as televisions or telephones. Scholars and commentators continue to bemoan the digital divide, tying it to broader issues of social and political equity (Hargittai & Hinnant, 2008; Mossberger, Tolbert, & Gilbert, 2006; Rodino-

Colocino, 2006; Shade, 2003). Relative to a television ad or a direct mailer, being the *recipient* of an online political message requires greater financial, educational, and technological wherewithal, limiting the reach of online political communication. Yet in an increasingly connected society, those among the general public who may constitute the persuadable targets of policy actors' communication are ever more likely to have regular internet access. This economic change most directly affects those coalitions who seek to mobilize the public—or, at least, the portion of the population most likely to care about their issue of concern.

Better Mobilization of Issue Publics

Networked computing provides policy actors with much more targeted and effective tools for communicating with issue publics, and the power to mobilize these issue publics is a second important effect. It is most effective for the groups who have an interest in drawing massive attention to an issue—generally, challenger coalitions—who can use internet communication to draw as much attention as possible from those who are most likely to agree with their policy beliefs and to invite them to get others involved. Further, as at least one study finds, "the Internet is expanding the numbers of the politically active, specifically in terms of reaching groups that are typically inactive or less active in conventional or offline forms of politics" (Gibson, Lusoli, & Ward, 2005). The number of online-only activists is still small, but there is obviously growth potential, especially among young people. This means that groups do not necessarily need to compete for the same population of would-be activists; the internet can help expand that population.

55

Several examples highlight the power of online public mobilization. Beginning with issue publics and spreading via viral online dissemination, several loose, new coalitions have been able to slow or stop policy changes by well-established governing coalitions. Viral email distribution was largely responsible for stopping policy changes supported by industries as powerful as banking (Bimber, 2003) and broadcasting (McChesney, 2004b), and email and blogging campaigns calling for mandated network neutrality helped stall an industry-backed telecommunications bill (B. D. Herman, 2006b). Each of these is an instance of using "communication technologies to target the smallest audiences likely to be helpful to [a group's] political aims" (Bennett & Manheim, 2001, p. 280). In none of these instances were temporary challenger coalitions able to force a change in the core attributes of a policy; each successfully slowed or prevented important changes in secondary policy matters. Further, major organizations that predated the widespread adoption of the internet, as well as the heavy use of offline communication strategies, also played an important role in each story. Yet they illustrate the ability to use the internet to target messages toward an issue public with immediacy and cost efficiency, a capability generally of more benefit to challenger coalitions.

An important part of the online mobilization of issue publics is the increasing power of ordinary citizens to engage in what Benkler (2006) refers to as "individual and cooperative nonmarket production of information and culture" (p. 2). Armed with a computer and an internet connection, millions of people who would otherwise never produce any media products have begun to do so. Today, the most politically potent genre of nonmarket online production is blogging, giving millions the power to set their

personal news agenda, break news, and reframe issues. Millions of bloggers serve as journalists and commentators with no expectation or hope of payment, taking advantage of their newfound power of political communication. Blogging also serves as an easy entry into political advocacy; one way to measure a coalition's success in mobilizing an issue public is to estimate the number of sympathetic blogs. Once they become bloggers, ordinary citizens can engage in sympathetic investigative reporting, analysis, and further public mobilization.

Public intellectuals are a particularly important subcategory of nonmarket information producers. Many professors have popular blogs on which they apply their expertise to countless topics of political salience. By donating their knowledge to the public sphere, they are effectively subsidizing the blogosphere with highly credible information that helps their cause. Linking patterns reflect these subsidies. As with the decrease in the cost of communicating, this increasing power of viral public mobilization will tend to favor challenger coalitions, who have every incentive to expand the scope of conflict to include a broader public debate (Baumgartner & Jones, 1993; Schattschneider, 1960).

Shaping the Public Definition of Issues

As a third important role for the internet, online advocacy can give some coalitions disproportionate power to shape the general public's perception of an issue. Search engine results and the nature of web browsing both favor websites with a high number of inlinks (Rogers, 2004). A link from one website to another is literally an invitation to visit the other site. From the perspective of the linked-to site, this is an

inlink—an incoming link from another website. If one follows hyperlinks via browsing, one is more likely to wind up at sites with a high number of inlinks; compared to sites with few inlinks, those with many inlinks are easier to stumble into because more sites point visitors toward them. Following the pioneering PageRank technology developed by Google (Google, 2008d), most search engines now use inlinks to measure a site's relative authority. A site with more inlinks has more authority, and if 2 sites are otherwise equally relevant results for a given search term, the site with higher inlink authority appears higher in search results.

To the extent that citizens gain or would seek out an understanding of an issue via the internet, the importance of hyperlinks greatly favors the coalition whose hubs have the most inlinks. Many who create hyperlinks online are aware of this vote-by-link system and deliberately game the system to create search results that favor their worldview (Tatum, 2005). For relatively obscure policy topics, such as the regulation of digital rights management (DRM) technologies, the internet is likely to be the preferred means for most citizens to learn more due to its low cost for end users (Hindman, Tsioutsiouliklis, and Johnson, 2003). Thus, sites that have high inlink authority will have a disproportionate chance to shape fresh opinions on an issue.

Because this authority can be altered by the linking behavior of the thousands of interested people who have related websites, it gives an edge to those coalitions with large numbers of motivated web denizens relative to those coalitions with more financing but fewer sympathetic online voices. On the internet, an advantage in human capital is generally more important than an advantage in financial capital. Because this linking

58

behavior will reflect the extent to which an issue public has been mobilized to add their

voice to the online mix, and because challenger coalitions have a more direct incentive to

mobilize online, the more central sites on an issue will tend to be the leaders among those

coalitions that are challenging the status quo. Since these more central sites have

disproportionate power to define the issue among those learning about it online,

challenger coalitions on obscure issues gain traction as more people use the internet to

investigate their pet issue.

Shaping Offline News Coverage

As a fourth important change, online communication has the potential to shape

offline news coverage. For instance, the agenda-setting relationship between political

blogs and print media is complex and often bidirectional, suggesting that on at least some

issues, newspapers rely in part on the web to determine when an issue merits coverage

(Wallsten, 2007). Additionally, information-rich online policy advocacy provides a

powerful, accessible information subsidy. Thanks to the dramatic increase in journalists'

workloads, journalists are increasingly dependent on external news sources for their

content (Davis, 2002). As one journalist explains, "people are increasingly reliant on the

wire services and Internet and other information coming to [them]" (Davis, 2002, p. 37)[11]

---

[11] Davis conducted this interview in May of 1999. This means that, roughly a decade ago, journalists were already regularly using the internet as a technique for cheap newsgathering. The trends that Davis identifies—reductions in newsroom budgets leading to rising workloads for journalists, and increasing investment in news subsidies by outside news sources—have continued on dramatic trajectories in the years since, with the internet playing a major role on both ends.

rather than investigative journalism. This opens the door for policy actors who can deliver information to fill this vacuum.

If certain policy actors have established themselves online as reliable, expert sources on an issue—if they stand out from the online crowd—a rich online information subsidy makes it easier for offline press to include favorable information and issue frames as developed online. This study does not test the direction of inter-media influence between online and offline media, but the real possibility of influence on offline media is another important reason that groups invest in online communication, from websites and blogs to email listservs. As with the other effects of online communication, this aids those groups that have fewer resources and/or less direct access to policymakers but a highly mobilized, sympathetic portion of the electorate. If reliable online communication comes to compete with or substitute for more expensive forms of press information subsidies—press conferences, Washington, D.C. offices, and so on—this change benefits those with less financial capital and more human capital. Also, recall that those with access to policymakers do not care particularly about getting their messages into the media, but those without such access are desperate to maximize media coverage (Kingdon, 2003). Thus, challenger coalitions communicate online at least partially because they hope this will shape offline media coverage, reaching policymakers' desks via this rather indirect route.

Cheapening of Mediated Political Communication

The above changes wrought by the internet all benefit challenger coalitions, but the cheapening of mediated political communication is decidedly to the advantage of

those who have direct access to policymakers. The internet has radically increased the total volume of mediated political discourse, cheapening the value of each act of communication—much as a deluge of cash leads to inflation, or a decrease in the purchasing power of every dollar. "Information abundance can lead to information fatigue as well as the reasonable calculation by political actors that a message sent cheaply means less than one sent expensively" (Bimber, 2003, p. 107). Most importantly, this information fatigue reduces the efficacy of mediated communication with policymakers. Thanks to email, for instance, it is now easier and cheaper for constituents to send written correspondence to their elected representatives. This means congressional staffers are now inundated with email correspondence, reducing the impact of any given message.

Unfortunately for challenger coalitions, this limits the efficacy of online communication as directed to policymakers. Meeting directly with policymakers is therefore still very valuable and much better correlated with policy success than is online popularity, and face-to-face meetings are as expensive as ever. Online communication may sometimes mobilize and expand issue publics, but face-to-face communication is a much straighter route to successful policy outcomes, and this continues to favor those groups with greater financial wherewithal.

This cheapening of mediated political messages serves to attenuate most of the benefits that online communication delivers to challenger coalitions. Communication within a coalition is still much cheaper and easier, but communication with the general public is less successful than it otherwise might be because countless other groups are

also using these same new media to vie for any given citizen's attention. The

mobilization of an issue public is also much cheaper, but it now takes a higher number of

mobilized citizens to grab congressional attention. This is exactly what happened in the

examples discussed above—in the areas of banking, broadcast media ownership, and

network neutrality—in which hundreds of thousands of citizens emailed their

representatives to express their scathing opposition, stopping well-backed legislation in

its tracks. Yet the ever-increasing number of electronic messages keeps raising the

number of messages required to demonstrate a critical mass of citizen input.

This attenuation in online efficacy also applies to challenger coalitions' greater

ability to define an issue online and to shape offline news coverage. Defining an issue

online only helps to the extent that the public investigates a topic, and the more topics

there are to explore online in ever-greater depth, the less the odds are that any given

citizen will tune into any given coalition's message. While the internet facilitates a

modest increase in the number of political participants (Gibson et al., 2005), that increase

cannot match the exponential growth in messages competing for those new entrants'

attention. Likewise, as more groups use cheap online communication in the hopes of

shaping offline news coverage, a relatively smaller set of them can succeed in any given

news cycle.

None of these limitations change the degree to which challenger coalitions should

try to use online media for coalition building, public outreach and mobilization, issue

definition, and penetrating offline news media. New media technologies still make all of

these goals much more attainable for coalitions that are under-resourced, and they are

powerful tools for any group seeking to expand the realm of conflict—any group seeking to force change on an otherwise static policy subsystem—regardless of resources. Thus, any group seeking major policy changes has every incentive to use the internet for policy activism, and to use it heavily. Yet these groups cannot rely on new media tools alone to effect desirable policy outcomes. To the extent that they can afford to direct precious resources toward old-fashioned policy activism, including DC offices and face-to-face communication, this is an essential component of any complete policy strategy. For a coalition with both a strong online presence and a respectable on-the-ground operation, their online strength can give a tremendous boost to their offline operations, providing ideas, activists, and political pressure on policymakers. The internet gives challenger coalitions much greater odds of success in their attempt to force policy changes on an entrenched policy subsystem, but it does not entirely eliminate the tremendous advantages of inertia, capital, and political connections.

Conclusion

This fusion of theories, drawn from disparate fields of study, is focused on achieving two purposes—the same purposes as this research project as a whole. First, this research serves to investigate the political forces behind the development of DRM policy. Because copyright law generally and DRM policy specifically will continue to play an important role in the future of mediated communication and the development of new media technologies, this purpose alone is worthy of a more substantial investigation. Second, this research serves as a case study—a laboratory for understanding the role of political communication within policy subsystems, especially in light of the recent

upheavals in communication technology. Toward this end, this unique fusion of theory and the following empirical investigation are of interest to anyone who cares about how the policy process, even though many may not know or care about copyright law.

This project carves out a fairly new research program, fusing policy studies, political communication, and the political economy of information. I invoke the policy subsystems literature as the foundation for examining policy outcomes. Yet despite the central role that mediated communication plays in policy advocacy, the political science theorists cited herein generally do not much consider bedrock elements of the study of communication: they ignore the vast literature on media effects, provide only the most basic suppositions about intended audiences, and devote too little attention to the means of information production and distribution. By introducing these elements and fusing them to the policy studies literature, I hope to introduce the communication strategies of advocacy coalitions as a much more important element in the study of policy outcomes. The shift I propose is especially important due to the profoundly democratizing potential of networked communication technologies. In an era when people can gather and distribute political information with unprecedented autonomy, the political significance of communication—as detached from other factors such as access to policymakers and financial backing—is almost certain to grow substantially.

I draw on theories of media effect generally and theories of mediated communication specifically, but this study is also different from the majority of this literature. The most studied genre of political communication is the electoral campaign, and scholars especially focus on the effects of centralized sources of communication—

politicians and the press—on audiences (see, e.g., Graber, 2005). The study of political communication online has also likewise generally focused on partisan or electoral communication rather than issue advocacy. While there are several important exceptions (see, e.g., Bimber, 2003; Dean, Anderson, & Lovink, 2006; Farrall & Delli Carpini, 2004; Napoli, in press; Pickard, 2006; Rogers, 2004), there are still far too few studies examining the internet as a tool for issue advocacy, leaving the research record incomplete.

Candidates, parties, and elections matter, but not for their own sake; they matter because electoral outcomes shape the laws that govern our society. Therefore, campaign communication is just a fraction of the important political communication worthy of scholarly study. The investigative model I propose here begins by treating policymakers not only as communicators, but also as potential audience members. I ask: who tries to communicate with policymakers directly, who does so indirectly via attempts to shape public opinion, and what are the policy outcomes? This is grafted onto theories of policy subsystems, because an understanding of policy subsystems enhances our ability to explain and predict which policy actors will choose which communication strategies.

This dissertation also represents a shift for the study of the political economy of information. Most scholars in this field study the economic and political conditions of the production and distribution of information, as well as the effect of these conditions on producers, distributors, and audiences. Some study the politics behind media policy debates in areas such as media ownership (McChesney, 2004b) and copyright (Litman, 2000), but few (e.g., Galperin, 2004; Hosein, 2004) have applied the wealth of theory

found in policy studies and political communication studies. Through such an application, I seek to explore the policymaking process that is today reshaping copyright into a means of regulating new media technologies.

Finally, this research emphasizes the role that new media technologies such as the internet are playing in the process of policy development. This may represent this dissertation's most substantial theoretical and empirical contribution. The anecdotes cited above, and the dozens like them that could also be cited, give some sense of the internet's power in shaping policy outcomes, but there is still too little work geared toward identifying the role of information—especially as communicated via new media technologies—in shaping policy outcomes. In short, this interdisciplinary field of study desperately needs more testable hypotheses about who uses which media, to communicate with whom, with what effects. This dissertation is an attempt at this kind of research.

CHAPTER THREE: THEORIZING THE COPYRIGHT DEBATE AND DERIVING

RESEARCH QUESTIONS AND HYPOTHESES

This research focuses on the role of communication in the policy process. On one level, it provides new insights into the debate over the future of digital rights management (DRM) policy in the US. On another level—at the broadest level, applicable to policy debates generally—this debate serves as a case study for studying the role of communication in the policy process, especially as that communication has adapted to the introduction of new media technologies. This single project speaks to this very broad question by focusing on a much more manageably narrow question: Who speaks about digital rights management policy in which media? In the last chapter, I brought a wealth of theory to bear on suggesting answers to the broader question. In this chapter, I first apply this broader theory to the copyright debate. In light of this application to the specific policy debate at hand, I then propose specific research questions and testable hypotheses.

Theorizing the Copyright Debate

Before building specific hypotheses about how various policy actors will communicate about DRM policy, this section combines the theory built in Chapter Two with some of the available literature on the copyright debate. First, I consider the historical membership of each coalition. Then, I consider a few specifics about their campaign donations. Third, I briefly discuss their historical track records of getting their

chosen policies passed. Finally, I say something about the status and hypothesized strategy of each coalition.

Coalition Membership

Over the last century or more (Litman, 2000), the most reliable members of the copyright policy space have been groups that profit from holding copyrights (Landes & Posner, 2004). These industries and their representatives have generally worked together in what I call the strong copyright coalition. The most visible and historically longest-standing members are those in media industries who make their money through publishing, songwriting, making and distributing recorded music, and producing and distributing motion pictures (Vaidhyanathan, 2001). Visible media industry groups today include, among others, the Recording Industry Association of America (RIAA), the Motion Picture Association of America (MPAA), the Association of American Publishers (AAP), and the National Music Publishers' Association (NMPA).

In recent years, the strong copyright coalition has also come to include technology firms whose primary revenue streams come from proprietary software. The most visible company with this revenue mix and policy strategy is Microsoft, and the industry association best associated with this strategy today is the Business Software Association, or BSA (Samuelson, 1999). Other generally supportive industries include other media sector businesses such as broadcast networks, law firms (which get more work as copyright grows), and technology firms such as Macrovision that make and sell DRM systems .

While the members of this coalition have long coordinated, over roughly the last quarter century, this coordination has escalated, and their beliefs about copyright have hardened (McLeod, 2005). They have long held common policy beliefs, including especially that copyright increases creative output and sustains the livelihoods of creative people. In more recent years, however, they have more strongly and successfully advanced the belief that few unauthorized uses should be or are permitted by law, as well as the belief that copyright is a right akin to the property right in land or physical belongings (B. D. Herman, 2008; McLeod, 2005, p. 109). Since copyright grants monopoly control that is less total than the right in physical property—with much broader limitations and exceptions, not to mention expiration dates—the metaphor of physical property is a particularly potent undergirding belief which the strong copyright coalition uses in its quest to expand the scope of intellectual property. From the perspective of those who see copyright as a property right akin to that in landed property, copyright provides woefully inadequate legal protection.

While the strong copyright coalition has a long history of coordination within and across sectors, the strong fair use coalition as it stands today has only existed for approximately 15 years. It is not the case that there were no voices against expanding copyright law before this time. For instance, librarians have a long history of reservations over the expansion of copyright; while not alone, the oldest and most visible group is the American Library Association (ALA). Educators have also generally sided with the call for strong fair use. From kindergarten through graduate school, the cost to educate students is lower thanks to copyright exceptions that privilege educators. Especially when

69

compared to librarians, however, educators are generally less up to speed and much less easily mobilized around issues of copyright.

The battle over copyright is often a battle over new technologies, which means well-established interests in the strong copyright coalition are often battling the upstarts who have developed a new technology (Lessig, 2004). In earlier battles with copyright holders who tried to use the courts and Congress to contain the copying capacities of new technologies, innovators such as the inventor of the player piano, early cable television services, and manufacturers of videocassette recorders generally fought their own fights.[12] Some of these fights helped build the beginnings of today's strong fair use coalition. In 1981, sparked by the legal controversy over the VCR, companies that manufacture and sell consumer electronics founded the Home Recording Rights Coalition (HRRC). Over the last twenty years, the Consumer Electronics Association (CEA) has developed much stronger opinions about copyright and about their need to fight for fair

---

[12] Throughout the most of the 20th century, Congress and the courts recognized the value of these new technologies and sought to fit them into existing copyright law. Policymakers either refused to set limits on their uses, as in the case of the VCR (Menell & Nimmer, 2007; *Sony Corp. of America v. Universal City Studios, Inc.*, 1984), or imposed statutory licensing schemes that would ensure users of new media could reproduce copyrighted works at affordable rates—as Congress did for the player piano and cable television (Lessig, 2001, pp. 108-109). Unlike today, fighting against copyright holder control was primarily up to each technology's innovators and the sympathetic public that had grown to love a given media technology.

   The above policy outcomes are remarkable if viewed retrospectively in light of the copyright policy debate as it stands today. The strong copyright coalition has successfully advanced their core beliefs, including especially the property-ness of copyright, into the halls of power and the public consciousness. They have also greatly increased their coordination and efficacy. Thus, the rapid growth and strengthening of the strong fair use coalition has been the only bulwark against an even greater expansion of copyright over the last 15 years.

use on a consistent basis. Except for the proprietary software sector, the technology sector generally has been on the side of stronger fair use; with varying degrees of reliability, this includes computer hardware manufacturers, telecommunications companies, online service providers, and (very reliably) the free software community.

Over roughly the last 15 years, scholar activists and nongovernmental organizations (NGOs) have gone from the margins to the center of the debate over copyright policy. The very titles of the prominent NGOs in the strong fair use coalition, such as Public Knowledge, the Electronic Freedom Foundation, and the Center for Democracy and Technology, correctly suggest some of the coalition's core policy beliefs. These include: The rights to circulate, build upon, and critique pre-existing knowledge and culture are vital parts of a democratic society. Copyright should restrict freedom of speech and press only to the extent necessary to ensure a vibrant public sphere, and it has long since gone too far in protecting copyright holders at the expense of new creators and the general public. New technologies are a potentially democratizing force, and the law should rarely stand in the way of technological innovation. The scholars and NGOs who have joined and helped to build up the strong fair use coalition have contributed substantially to these beliefs.

As discussed in fuller detail in Chapter Five, the proposal for what would become the DMCA (*Digital Millennium Copyright Act*, 1998b), sparked substantial outrage by a number of legal scholars, leading to the 1995 founding of the Digital Future Coalition, or DFC (Litman, 2000). In addition to legal scholars, members included representatives from the library community, the technology sector, educators, and consumer groups. To

71

this author's knowledge, the DFC was the first special-purpose public interest NGO dedicated to ensuring that copyright did not stop technological innovation and innovative uses of copyrighted works. It also represented the beginning of a substantial increase in the policy advocacy role played by legal scholars. The DFC is not very active today, but other NGOs in many ways serve as the rhetorical and organizational core of the strong fair use coalition; these include Public Knowledge (founded in 2001) and the Electronic Frontier Foundation (founded in 1990, but only later coming to spend substantial energy on copyright). It is only with the involvement of these public interest groups that the strong fair use coalition began to coalesce and coordinate to a degree that has allowed them to present a serious counterbalance to the strong copyright coalition.

Campaign Donations

While the theory cited and developed in Chapter Two generally concludes that money is not the sole determinant of political outcomes, it would be naïve to ignore the role that money has played in shaping copyright law. Samuelson (2004) derides recent expansions of copyright law as "the best laws money can buy" (p. 9). Wit aside, she makes a serious and valid point, echoed by Landes and Posner (2004): the benefit of increasing copyright law is concentrated in the hands of those who profit from selling access to copyrighted works, while the cost is widely distributed across the general public and industries that are not as directly affected. This creates problems of collective action for the strong fair use coalition; it is far harder to mobilize a diffuse group of indirectly affected people than a concentrated group of people who are directly affected by a given policy choice.

72

While the available resources on campaign donations do not include precise divisions of donors into each of the industry segments identified above, the data are precise enough to suggest that there are far greater campaign donations behind stronger copyright law than behind strong fair use. The Center for Responsive Politics is a treasured source for this kind of information, which is hosted at their website, OpenSecrets.org. From their list of industries (Center for Responsive Politics, 2009), one can obtain detailed information on the total campaign contributions of each from the 1990 election cycle to the 2008 cycle. Since that is almost exactly the period under study—save the 2008 cycle—this data is quite useful in establishing an estimate of the campaign contributions of each coalition.[13] Table 3.1 provides a summary:

Table 3.1: Federal Campaign Contributions by Sector, 1990 Cycle to 2008 Cycle

Sector	Rank Among All Sectors	Total Contributions
Lawyers & Law Firms	1	$1,013,573,797
TV/Music/Movies	8	$250,271,735
Misc. Manufacturing & Distribution	13	$199,532,204
Computers/Internet	14	$183,298,065
(Computer Software)	(subset of Computers)	($69,588,337)
Education	16	$171,439,513
Telephone Utilities	28	$115,682,460
Printing & Publishing	29	$114,365,977
Telecommunications Services & Equipment	45	$67,627,514
Recreation/Live Entertainment	66	$29,013,747

---

[13] Data are available for each cycle individually as well, but the summary data— including an election cycle that occurred after this study's time frame—is ample to illustrate the very sharp differences in each coalition's investment in campaign donations.

This illustrates the sharp difference between the two coalitions' campaign contributions. Note the second entry in the list, the TV, movie, and music industries. They easily outspent the oil and gas industry, the pharmaceutical industry, and the beer, wine, and liquor industry. Combined, the TV, movie, and music industries have more money to make or lose based solely on the outcome of copyright policymaking than any other industry in this table—perhaps more than everybody else combined—and the very high volume of campaign contributions reflects this. Because of this direct interest in copyright policy, it is their top legislative priority; if the quarter billion dollar donations send one message, it is to maintain and expand copyright law. The other groups that have the most direct interest in copyright, the software and publishing industries, are also substantial donors with a strong interest in strengthening copyright.

Two other industries have an indirect or diluted interest in pushing for stronger copyright. Attorneys represent the very top source of campaign contributions over the last twenty years, and while copyright law is one of a large number of areas of law, it is an important and growing area, and attorneys stand to make more money if copyright grows and increases the number of cases. Nonetheless, despite being the largest donors, one could elide them from this table without substantial loss of the table's descriptive power; the legal industry's direct financial stake in these policy debates is very small compared to the size of their donations. The recreation and live entertainment industry has a somewhat more direct interest in stronger copyright. Most of the top twenty donors are professional sports leagues or teams, and they make a substantial share of their money by selling broadcast rights.

On the other side of the copyright debate are industries with a far less direct stake in copyright law. Miscellaneous manufacturing and distribution is an exceptionally broad category, including a small subset of manufacturers that produce consumer media devices, a small subset of which may become the subject of litigation depending on the outcome of copyright policy. The manufacturing sector's interest in copyright is thus far less direct than the lawyers and law firms, and the latter outspend the former by five to one. Computers and internet companies who do not depend on copyright to sell software—in particular, hardware manufacturers—have a somewhat more direct interest in a modest role for copyright; they can face increased liability for some of their products if copyright grows. Yet as companies who manufacture and distribute products that exemplify the global trade in raw materials, parts, and finished products, they have a long list of legislative priorities that are likely more immediate motivations for these donations. Further, 38 percent of the computing sector's donations come from the software industry, and they generally have enough of a direct interest in strengthening copyright that it is their top legislative priority. Thus, the computer industry is divided, though realistically, the proportion of contributions multiplied by the donor's relative stake in copyright suggests that the total effect of contributions leans substantially toward expanding copyright.

The remaining sectors have a shared, minor stake in moderating copyright law. Educators, telephone utilities, and telecommunications services and equipment companies all face realistic threats of copyright liability. In terms of legislative priorities, however, this is a very minor concern for each. Educators are first and foremost

concerned with federal education and research funding. Any university president worth her salary would gladly accept a ten percent increase in copyright royalty payments if it came with a ten percent increase in federal funding for education. Even less pressing issues such as academic freedom and labor policy are far more important to most education sector donors.

The telecommunications industry is generally far more concerned with other areas of government regulation, though they did face a substantial scare during the debate leading up to the DMCA (*Digital Millennium Copyright Act*, 1998b). Legislation introduced in both the House and Senate in 1995 would have amended the definition of "publication" to include internet transmission (*NII Copyright Protection Act of 1995*, 1995a; *NII Copyright Protection Act of 1995*, 1995b), meaning that every infringing copy of a work transmitted online would become a legal liability for internet service providers (ISPs). These proposals were ultimately rebuffed in favor of what became Title II of the DMCA, which substantially limits ISP liability for infringing copies transmitted by end users. This change most directly addresses the telecommunications industry's narrow interest in copyright; with this proposal attached to what became Title I of the DMCA, which greatly increased the scope of copyright, the industry strongly supported this bill that in total reduced their specific copyright liability (B. D. Herman & Gandy Jr., 2006).

The total campaign donations behind greater copyright law are therefore much larger than the donations behind more modest copyright. The TV, music, and movie industries alone represent such a substantial source of campaign donations that are so directly targeted at increasing copyright as to virtually drown out the competition. The

next most directly interested industries, publishing and software, also provide substantial campaign donations that signal a strong desire for stronger copyright above all else. In comparison, nobody has a comparably direct or well-funded interest in reduced liability. Every other donor in this table—with the possible exception of the recreation and live entertainment industry, which also supports strong copyright—has many legislative priorities that are far higher on their agenda. No candidate for federal office needs to ask how copyright law affects their ability to raise campaign cash; supporting stronger copyright law clearly increases their ability to raise funds. This strongly suggests that the strong copyright coalition is the governing coalition, well connected enough to defend past gains and jockey for future changes. The history of legislative outcomes suggests the same.

Legislative Outcomes

The history of changes in copyright law buttress the claim that the strong copyright coalition is the governing coalition; changes are almost always in their favor. Copyright has grown exponentially over the last century (Litman, 2000; McLeod, 2005; Vaidhyanathan, 2001). Landes and Posner (2004) point out that copyright's growth can be measured in part by measuring the law itself: the number of words in the copyright statute. "The copyright statute had 11,550 words in 1946, 22,310 in 1975, a tripling to 61,600 in 1976 with the passage of the new Copyright Act, and 124,320 words in 2000— a nearly elevenfold increase in fifty-four years" (p. 2). While this is an imprecise measure, it correctly suggests the substantial growth of copyright's reach.

77

When the 61st Congress first convened in 1909, copyright lasted for an initial term of 28 years, with one possible renewal of 14 years. Today, a work created by an individual author lasts for the creator's life plus 70 years, and a "work for hire" usually lasts for 95 years after it is first published. While formerly copyright protection was not granted until a work was registered with the US Copyright Office, the 1976 act (*Copyright Act of 1976*) changed the law so that registration is no longer required; any creative work fixed in a tangible medium is now automatically copyrighted. Penalties have also grown disproportionate relative to any credible measure of market harm. In 2003, four students "were threatened with a $98 billion lawsuit for building search engines that permitted songs to be copied" (Lessig, 2004, p. 185). These claims were based on statutory damages that have also grown exponentially over the last century. Overall, copyright protection has grown much more complicated, terms have grown longer, the number of protected works has grown sharply, and penalties and legal liability have exploded.

Chapter Five details two important examples of the growth of copyright as a tool for regulating technology during the last twenty years: the AHRA (*Audio Home Recording Act*, 1992) and the DMCA (*Digital Millennium Copyright Act*, 1998b). They represent an important change in the kind of regulation implemented by copyright, but they are part of a much longer and broader trend of consistently increasing copyright protection.

Status and Strategy

Especially when combined, the above details about campaign donations and legislative outcomes show that the strong copyright coalition has long been the governing coalition in the copyright subsystem. In contrast, there is scant evidence of the strong fair use coalition having substantial political power. Existing in something like its current form for less than 15 years, the strong fair use coalition is clearly a challenger coalition, attempting to unseat the strong copyright coalition from its perch atop the policy subsystem.

This shapes their communication strategies in fairly dramatic ways. As all policy advocates do, both coalitions will want to communicate directly with policymakers as frequently as possible. Thus, both coalitions will testify in Congress as often as resources and access will permit. Both coalitions will also want to communicate regularly with the mainstream press, though with different degrees of incentive and access. As the established dominant force in the copyright debate, the strong copyright coalition should have nearly automatic access to the news media but with the incentive to see coverage of the policy debate remain low and keep the issue far from the top of the mainstream policy agenda. In contrast, the strong fair use coalition will not have automatic access but will have to fight harder to be included in relevant news stories. They will fight for that access as part of their strong desire to expand the scope of conflict to include the general public and the entire Congress, meaning they will try to land as many news stories as possible about the specific policy debates at hand. This mismatch of incentives, combined with the

journalistic norm of objectivity, should dilute the effect of the strong copyright coalition's near-automatic press access.

Finally, the online debate will almost exclusively reflect each group's relative size and willingness to devote time to making its arguments available for anybody who might care to learn more. Because the strong fair use coalition is a diffuse group of individuals and organizations that are far less well funded than the small group of major corporate players that make up the heart of the strong copyright debate, the wide diffusion of the internet is for them a highly advantageous development. A highly diffuse group with some extra time and energy can collectively shape the online debate far more effectively than can a concentrated, well-funded group, and they can do so while spending little or no extra money (Benkler, 2006). Further, the strong fair use coalition has a unique incentive to seek broad public attention and mobilization, giving the well-organized groups at the core an incentive to post as much relevant information as possible in the hopes that this information will spread throughout the diffuse network of mobilized activists. Compared to the strong copyright coalition, the fair use advocates are particularly likely to have more related sites, more information on those sites, and heavier linking to other sites within the debate.

In contrast, there is little reason to expect a substantial online presence for the strong copyright coalition. They have relatively few mobilized individuals and thus little ability to create a large number of relevant sites. Even the core groups have less incentive to post many relevant documents or to participate in the online debate in the form of heavily linking to other relevant documents. By linking away from other sites within the

debate, they can help delegitimize the strong fair use coalition; linking back to fair use advocates and taking part in the online debate would only add to their credibility.

By applying the theories of the policymaking process in general to the DRM policy debate specifically, one can see how these groups' differences in status and incentives lead to likely sharp differences in communication strategy. Drawing on some additional literature, the next section formalizes these differences into testable research questions and hypotheses.

Research Questions and Hypotheses

With the above theories about the copyright debate in place, the next step is to offer specific research questions and hypotheses that are amenable to empirical testing. This chapter groups research questions and hypotheses into five sections: access to Congress, access to elite newspapers, changing access to Congress and elite newspapers, online communication, and comparisons between the Congress, newspapers, and the web.

Communicating in Congress

First, I am interested in whose voices are better represented in congressional hearings. Legislative change over the time period studied has been almost exclusively to the benefit of those who support stronger copyright law (Landes & Posner, 2004), which suggests that the strong copyright coalition has better congressional access, no matter which theory of policy stability and change one uses. While it would be quite an oversimplification, one could describe this policy subsystem as an iron triangle, with copyright holders having preferential access to the House and Senate Judiciary

Committees and the Copyright Office.[14] Copyright law has moved ever closer to the wishes of the strong copyright coalition, which also suggests better access to policymakers (Sabatier & Jenkins-Smith, 1999). Further, the long period of stability on this issue suggests that the flow of information through congressional channels has favored the status quo (Baumgartner & Jones, 1993; Jones & Baumgartner, 2005).

More directly, the literature on the politics of DRM regulation argues that the strong copyright coalition has better access in Congress (B. D. Herman & Gandy Jr., 2006; Litman, 2000). This should play out in the form of greater access to make their arguments before congressional hearings. This is formalized as:

Q1: **Are one coalition's arguments communicated in Congress more frequently than the other?**

H1: **Strong copyright messages are communicated in Congress more often than strong fair use arguments.**

This study measures this using each discrete contribution to Congressional debate as included in the official hearing record. Most obviously included are speeches delivered at the witness table, but also included are witnesses' written materials—whether written supplements to oral testimony or additional written materials such as policy studies—as well as oral and written statements by members of Congress. Consistent with earlier research (B. D. Herman & Gandy Jr., 2006), question-and-answer sessions are elided;

---

[14] The Copyright Office is technically part of the legislative branch but on this issue serves as something of an administrative agency. For more on the Copyright Office's role in administering DRM policy, see Herman and Gandy (2006).

these present no easily identifiable, discrete statements attributable to a given individual or organization.

The unit of measurement is the single document. Each speech or document will support more copyright protection, expanded fair use, or a relatively neutral or mixed position between these two extremes. If neutral or mixed, paragraph-level coding will assess its proximity to either coalition. This method is discussed in further detail in Chapter 4 and in Appendix B; for now, the important point is that this and most of the remaining hypotheses are tested by measuring the ratio of strong copyright documents to strong fair use documents.

Communicating in Elite Newspapers

In addition to communication with Congress, another important contribution is access to and use of the press, and I am interested in how this compares to congressional access. Here, the thrust of the theory cuts in both directions; there are reasons to suspect that each coalition will be better represented in the press. Due to both the countervailing direction of these forces, as well as the journalistic norm of objectivity, this is the hardest medium for which to predict a clear winner.

On one hand, there are reasons to suspect that the strong copyright coalition will have better access to the press. As discussed in Chapter 2, the power of information subsidies suggests that concentrated economic interests generally have greater access to the media, and the most politically influential form of this access is via elite newspapers such as *The New York Times* and *Washington Post* (Gandy Jr., 1982; E. S. Herman & Chomsky, 2002). In this particular policy debate, the strong copyright coalition has much

more concentrated economic interest in copyright policymaking (Landes & Posner, 2004, pp. 13-21). As a well-established "insider" group and the dominant policy coalition, they will also have an automatic voice in the press (E. S. Herman & Chomsky, 2002; Kingdon, 2003).

On the other hand, there are also reasons to suspect that the strong fair use coalition will have more incentive to subsidize press coverage of DRM policy, leading to greater access for that coalition. The literature on policymaking and attention suggests that the strong copyright coalition will have less incentive to raise public attention than congressional attention; keeping low profile on an issue helps preserve the status quo—which is to the benefit of the strong copyright coalition—while elevating the issue to a matter for broader electoral debate and discussion benefits those who seek change (Jones & Baumgartner, 2005; Kingdon, 2003). In part, the strong copyright coalition has empirically succeeded by avoiding rather than seeking public attention (McChesney, 2004b, p. 233). This does not eliminate or completely contradict the institutional advantage that suggests automatic press inclusion of the voices of these powerful economic actors, but it does suggest that the strong copyright side will not want to see a large number of news stories on DRM policy, favorable or otherwise.

Finally, the journalistic norm of objectivity suggests that the outcome will be balance in the representation of each side's arguments. In the press's view of itself, it does "not operate as a platform for one dominant public voice but, rather, as a 'channel' for a variety of speakers" (Kaplan, 2006, p. 181). In combination with the two countervailing forces identified above, the norm of objectivity suggests that neither side

84

will win a clear victory in press coverage. Since the strong copyright coalition is presumed to have automatic access, their arguments should still appear in the press slightly more often than strong fair use arguments, but the ratio should be more balanced than that in Congress. Formally:

Q2: **Are one coalition's arguments communicated in elite newspapers more frequently than the other?**

H2: **Strong copyright messages are communicated in elite newspapers more often than strong fair use arguments.**

And:

Q3: **Is the ratio of strong copyright arguments to strong fair use arguments different in elite newspapers than in Congress?**

H3: **The ratio of strong fair use arguments to strong copyright messages is larger in elite newspapers than in Congress.**

As with the first hypothesis, I test the second and third hypotheses simply by counting the number of documents representing each coalition's position. Because of the norm of journalistic objectivity, the paragraph-level coding of mixed or neutral documents becomes especially important for measuring newspaper articles.

Changing Access to Congress and Elite Newspapers

This research begins from the belief that the balance in both Congress and elite newspapers will be to the benefit of the strong copyright coalition, but there are also reasons to suspect that this advantage will have deteriorated over time. First, over the course of the time period studied, the strong fair use coalition has blossomed from virtual nonexistence into a thriving political force. In particular, important nonprofit groups that did not exist in 1989 have started and blossomed in the intervening years. While there are several, two are particularly noteworthy. The San Francisco-based Electronic Frontier Foundation (EFF), founded in 1990, began primarily as a voice against internet

censorship and for individual privacy. Over time, however, the group has become an increasingly involved and visible voice in the copyright debate, especially in light of the DMCA. Since its 2001 founding, Public Knowledge has risen to prominence as the pre-eminent DC-based policy advocacy organization within the strong fair use coalition.[15] The widespread adoption of the internet has inspired the birth of a number of additional internet policy groups, many of which are also more or less committed members of the strong fair use coalition. Thus, these groups' entry and rise suggests a very good reason to predict a change over time: the entry of previously unmobilized interests who call for greater fair use.

At the beginning of the time period under study, without involvement from these groups, the iron triangle may actually have been a reasonable representation of the policy space; as long as the industrial interests at the table agreed on a DRM policy, there was no force inside or outside government—no opposing coalition—that stood in the way of its passage or maintenance. In contrast, the long-term development of a serious strong fair use coalition, strongly opposed to DRM regulation, represented an important change

---

[15] Two caveats are in order here. First, I interned for Public Knowledge in the summer of 2006, but this was a choice that reflected my pre-existing belief in the above claim. Second, other groups have exerted at least some influence predating at least the signing of the DMCA. The Digital Future Coalition (DFC) was founded in 1995 as "a unique collaboration of many of the nation's leading non-profit educational, scholarly, library, and consumer groups, together with major commercial trade associations representing leaders in the consumer electronics, telecommunications, computer, and network access industries" (Digital Future Coalition, n.d.). The DFC participated in negotiations over the passage of the DMCA, but the group's public presence has diminished in the decade since. The Center for Democracy and Technology (CDT), a DC group founded in 1997, has also played a role in copyright policymaking, but like the EFF in its early years, CDT is concerned more with issues such as privacy. The EFF has only had an intermittent presence in Washington, DC.

in the policy subsystem. Applying Sabatier and Jenkins-Smith (1999, pp. 128-129, 135-137), this was a period during which it evolved from a nascent policy subsystem, with fluid beliefs and unstable coalitions, to a mature subsystem, characterized by hardened policy beliefs and stable opposing coalitions. Since the strong fair use coalition is the group that rose from virtual nonexistence to powerful membership in the policy subsystem during this time, it would be very surprising if the content of congressional testimony did not move toward the strong fair use end of the rhetorical spectrum during this time.

As an additional reason for suspecting the shift toward the strong fair use coalition, recall that the improvement, cheapening, and widespread adoption of internet technology has made it easier for lesser resourced groups to compete with very well resourced groups in policy advocacy. These new tools have made policy coordination and advocacy far cheaper and easier, reducing the relative advantage of groups with higher concentrations of economic interest (Bimber, 2003). Combined with the entry of new groups, this justifies the prediction of at least a modest change in the balance of power over time, even in offline media such as hearings and newspapers. In other words:

Q4: **Has one coalition gained a greater ability to communicate its message in Congress over time than the other?**

H4: **The ratio of strong fair use messages to strong copyright messages in Congress will have been larger from 2003 to 2006 than from 1989 to 1998.**

And:

Q5: **Has one coalition been more able to communicate its message in elite newspapers over time than the other?**

H5: **The ratio of strong fair use messages to strong copyright messages in elite newspapers will have been larger from 2003 to 2006 than from 1989 to 1998.**

As with the first three hypotheses, I test these by counting the number documents that support each group's position. If the ratios are more favorable to the fair use coalition in the later period in Congress and in the press, then the fourth and fifth hypotheses, respectively, are supported; otherwise, they are not supported.

Online Communication

This research also studies the dissemination of information online. Many claim the internet is a powerful new means of fostering political communication among those who are otherwise effectively muted in formal policy venues and offline media. As noted above, there are many reasons for this claim, including the greatly reduced costs of producing messages online, better mobilization of issue publics, motivated issue coalitions' ability to shape the public definition of an issue, and a newfound ability to shape offline news coverage. In the copyright debate, this upsetting of the dominant communication order means that the strong fair use coalition should fare much better online than offline.

Further, the two coalitions do not have equal incentive to spark the widest possible debate. As the side that has won consistently in the recent past, the strong copyright coalition (which already has the ears and sympathies of the policymakers that have supported them) has every incentive to keep the debate as contained as possible

(Jones & Baumgartner, 2005; Kingdon, 2003). If they have a policy idea, they can call symphathetic policymakers and begin the process from the inside; barring that, as institutionalized authorities, they can call the elite media, a move that is apt to get policymakers' attention.

In comparison, those who have the most incentive to gain the attention of a broad issue public—in this case, the strong fair use coalition—are likely to provide more information online. They are less able to contact policymakers directly, and as the political outsiders, they also must work harder to get their message into the elite media. But through online communication, they can contact sympathetic issue publics (Bennett & Entman, 2001) and, if successful in making enough noise through them, they may be able to change the unfavorable conditions Congress and the press. Thus, the strong fair use coalition has a much stronger incentive to provide rich information subsidies online, and the low cost of doing so provides a great opportunity for lesser resourced groups and symphathetic citizens—the mobilized issue public—to leverage their collective energy.

For all these reasons, the strong fair use coalition should have a much higher number of inlinks. Those groups leading the online charge for stronger fair use will likely have a high number of sympathetic websites pointing in their direction; as described in the previous chapter, this number of incoming links is termed "inlink authority" and it represents a fairly accurate quantification of a given site's power to define an issue. Because the strong fair use coalition has the incentive to communicate more information online and to mobilize sympathetic citizens to join this online chorus, and because the low cost of doing so makes it possible to act on this motivation, the strong fair use

coalition should have more relevant information online and have higher inlink authority than the strong copyright coalition. Formally:

Q6: **Will different coalitions provide different amounts of online information that is pertinent to ongoing policy debates?**

H6: **On average per site, strong fair use actors will provide more information pertinent to ongoing policy debates than will strong copyright actors.**

And:

Q7: **Will different coalitions receive different degrees of inlink authority in the online issue space?**

H7: **The websites of strong fair use actors will have a higher number of total and average inlinks from within the online copyright policy space than those of strong copyright actors.**

Both of these hypotheses depend upon fairly specific operational definitions; the circumstances under which each hypothesis is supported are more fully described in the following chapter on methodology.

Comparing Media

Taken together, the hypotheses above suggest that the online representations of the DRM debate will have been much more favorable than the offline representations in Congress and elite newspapers. The strong copyright coalition has historically enjoyed better access in Congress (B. D. Herman & Gandy Jr., 2006; Litman, 2000), so the strong fair use coalition has every incentive to seek broader public attention (Baumgartner & Jones, 1993; Schattschneider, 1960), including via internet mobilization. Thus, it would

be quite surprising if the online communication was not much more favorable to the strong fair use coalition than congressional communication. Likewise, the institutional advantages of incumbency and industrial largesse should give the strong copyright coalition a certain degree of automatic access to newspapers (Gandy Jr., 1982; E. S. Herman & Chomsky, 2002), so even though the strong fair use coalition will have stronger incentives to seek publicity around the issue, one has little reason to suspect that they will dominate newspaper coverage. In contrast, the volume of online communication produced by a policy coalition is primarily driven by sheer enthusiasm; neither cost nor institutional gate keeping serve to preclude the coalition that is most directly motivated to seek public attention from producing as much information online as they wish. Thus, the strong fair use should also do much better online than in the newspapers.

Formally, this means comparing the ratio of strong fair use arguments online to those in print and in Congress:

Q8: Will the ratio of strong copyright arguments to strong fair use arguments be different online than in the Congress and elite newspapers?

H8: The ratio of strong fair use arguments to strong copyright messages will be larger online than that in Congress.

H9: The ratio of strong fair use arguments to strong copyright messages will be larger online than that in elite newspapers.

These hypotheses capture the heart of this investigation. If the internet presents a truly disruptive medium for political communication, the logically primary step in this disruption is that it enables communication that is different than offline media. For those

who study specific policy debates, this represents an interesting test case. For those who study copyright policy, this might help explain the profound disconnect in each coalition's perceptions about the debate.

Conclusion

These questions and hypotheses follow from applying the broader theory of the policymaking process to the copyright and DRM policy debate. If the strong copyright coalition has the characteristics of a governing coalition, they should have superior access to policymakers, near-automatic access to the press, little desire to see a high degree of public attention, and little online presence. Likewise, if the strong fair use coalition has the characteristics of a challenger coalition, they should have inferior access to policymakers, strong incentive to work their way into the press, a strong desire for attention, and a strong presence on the web. As the strong fair use coalition has grown in numbers and strength, they should have gained on all these counts, reducing the comparative advantage of the strong copyright coalition. The next chapter details the methodological tools for putting these assumptions to the test.

CHAPTER FOUR: METHODOLOGY

In order to answer the research questions outlined above, I deploy three methods. First, I conduct a case study of the two laws and two proposed laws under study; I describe each below. Second, I perform a web graph analysis of the online copyright issue space. Finally, the bulk of the research consists of a content analysis of congressional documents, print media, and online sources.

Case Study

The first step in this project is a case study of the most significant reforms or attempted reforms in the history of the regulation of digital rights management (DRM). This consists of a short overview of each bill's history and details, including:

- An overview of the proposal

- A description of the external events surrounding the proposal

- A description of any key permutations in the proposal as it evolved,

- Important dates as the bill moved from proposal to legislation or rejection

- Important legislators and their roles in the debate

The purpose of the case studies is primarily to provide context for the reader's better understanding the results of the quantitative content analysis. To a lesser extent, it may also suggest the role of policy communication in shaping policy outcomes. Neither of these is central to testing this project's hypotheses, but both are important reasons the case study analysis rounds out the final dissertation.

In the evolution of US policy on digital rights management over the past 20 years, the four most important legislative proposals are:

- The Audio Home Recording Act (AHRA), passed in 1992 (*Audio Home Recording Act*, 1992)

- Title I of the Digital Millennium Copyright Act (DMCA), passed in 1998 (*Digital Millennium Copyright Act*, 1998b)

- Attempts to reform Title I of the DMCA, both in the 108th Congress (*Digital Media Consumers' Rights Act of 2003*, 2003) and in the 109th (*Digital Media Consumers' Rights Act of 2005*, 2005)

- Attempts to mandate the adoption of a technology called the "broadcast flag" in both the 108th Congress (no formal bill introduced) and 109th (*Audio Broadcast Flag Licensing Act of 2006*, 2006)

Here, I provide a very brief description of each. The next chapter provides a fuller description of each as part of the broader policy case studies.

Audio Home Recording Act (AHRA)

The AHRA was a legislative reaction to the recording industry's fear of the Digital Audio Tape (DAT) deck, developed by Sony in the 1980s (Lee, 2007, pp. 451-460; Menell & Nimmer, 2007, pp. 20-21). Unlike analog cassette tapes, DAT technology allows for a nearly infinite number of perfect clone copies. The recording industry therefore used a combination of congressional lobbying, threatened and actual litigation (Menell & Nimmer, pp. 19-20), and market pressure to retard the importation of DAT machines. The recording industry and the electronics manufacturing industry settled out

of court, and the AHRA represented a legislative embodiment of the settlement. The act permits manufacturers to sell DAT decks and other digital audio recording devices (e.g., audio component compact disc burners) on the condition that they equip consumer-level decks with a DRM system that would permit users to make a copy of an original but not of a copy. "For the first time in copyright history, the AHRA imposed technological design restrictions on copying devices and established a royalty system on the sale of recording devices and blank recording media" (Lee, 2007, p. 451). This would set the table for future proposals by which copyright industries would propose further design restrictions.

Title I of the Digital Millennium Copyright Act (DMCA)

The AHRA implemented a specific form of DRM on a specific class of devices; in an ever-changing technological environment, this represented an inherent limitation on the bill's efficacy. The copyright industries knew this, so they sought legal protections for DRM systems yet to be invented—to put the force of law behind solutions that they would develop and deploy in an effort at self-help. Thus, Title I of the DMCA regulates almost all forms of digital rights management. It bans the act of circumventing DRM to gain unauthorized access to copyrighted works. It also bans trafficking in tools that circumvent most forms of DRM. Further, it bans the removal or alteration of copyright management information—digital identifiers such as watermarks inserted into works to mark the identity of the copyright holder and communicate other information.

The bill imposes very stiff civil liability and, for those who violate the law for commercial gain, criminal penalties of up to one million dollars and 10 years in prison.

Further, the bill provides for few exemptions for noninfringing uses such as fair use. The strong copyright coalition came out in full support of the DMCA; supportive witnesses came from industries including the music, movie, publishing, and software, as well as the Copyright Office and the Patent and Trademark Office. The members of the strong fair use coalition opposed Title I or pushed for it to be substantially weakened. These included educational institutions, computer hardware manufacturers, librarians, and the nonprofit sector.

DMCA Reform

The strong fair use coalition desperately wants to change Title I of the DMCA, and Representative Rick Boucher, Democrat of Virginia, has led the congressional charge. In the 108th and 109th Congresses, he introduced bills (H.R. 107 in 2003-2004 and H.R. 1201 in 2005-2006) that would reduce the scope of the anticircumvention provisions of Title I of the DMCA (17 U.S.C § 1201). It would nullify the ban on circumventing copy controls as applied to otherwise legal activities. For instance, assuming that it is a fair use, a hobbyist could circumvent the encryption on several DVDs to make a video remix and play it at home for friends and family. The law would still prohibit circumventing DRM en route to selling bootlegged copies of copyrighted works on the subway; one who did so would be subject to the DMCA's civil and criminal penalties *in addition to* the civil and criminal penalties that apply to the acts of infringement. Because infringement is already illegal, opponents decry Boucher's bill as a gutting of section 1201.

Broadcast Flag

Accompanying the transition to digital radio and television broadcasts, the music, television, and movie industries express trepidation about the potential for viewers to record perfect digital copies of broadcasts, edit out the commercials, and post them online. For them, the broadcast flag is a potential solution to this problem. The flag is a very small addition to the digital broadcast signal by which a broadcaster can tell compliant devices which programs may be recorded, which cannot, and what can be done with permitted recordings. Device makers have every incentive to make noncompliant devices; many consumers will more highly value tools that permit any and all recording, and incorporating flag-compliant technology is an additional manufacturing expense. Hence, the music and motion picture industries seek to impose a federal mandate that all digital radio and TV tuners comply with instructions contained in a broadcast flag.

Web Graph Analysis

In addition to the case studies of the four policy areas described above, I also perform a web graph analysis, both for its own sake and as a tool for identifying a population of relevant online documents, via snowball sample (Farrall, 2005), to content analyze. Here, I first explain the technology of the web graph and justify its use. Second, I explain how I use it to locate websites from the population of those that participate in the online debate about digital copyright law. Third, I consider why using this technology is a superior research strategy to simply using the more straightforward method of conducting targeted internet searches in a search engine such as Google. Fourth, as a

means of establishing the validity of the specific web graph results I use, I describe the websites identified and compare these results to other web crawls.

A Brief Introduction to Web Graph Analysis

Relative to congressional and print media documents, identifying a population or sample of topic-specific websites for analysis is a rather difficult task. Constructing a representative sample of content is already difficult (Krippendorff, 2004), a problem that is only exacerbated by the boundless amount and variety of online sources (Herring, 2004). It is virtually inconceivable to imagine constructing a sample of equally informative websites or web pages from within the population of those that are working on an issue. In place of statistical representativeness, another worthy goal is authority. As discussed in Chapter 2, internet sites tend to cluster around issues of mutual interest, and attention within such clusters is disproportionately focused on an even smaller cluster of central sites, deemed by the group's collective linking behavior to be more authoritative. In every category of political websites, these tiny handful of websites dominate the issue space (M. Hindman, Tsioutsiouliklis, & Johnson, 2003). Thus, as long as a method permits one to identify and study the few dominant websites, one can confidently say something about the views of the other sites in the same formation.

Web graph analysis is just such a method. Developed by Richard Rogers (Rogers, 2004) and further refined by Kenneth N. Farrall and Michael X. Delli Carpini (Farrall & Delli Carpini, 2004), web graph analysis uses Rogers' Issue Crawler software (Govcom.org Foundation, n.d.). Farrall (Farrall, 2005) explains the process:

The Issue Crawler ... builds the web graph from a seed of URLs provided by the

researcher. The seed is expected to include significant websites within the issue of

interest. The software scans the seed documents for links pointing to external

domains and stores these links in a matrix. Any links [that] are not present in at

least two of the seed documents are thrown out. The linked documents that remain

are then scanned again for external links, with the same criteria for throwing out

solitary links (a process known as co-link analysis). The process is usually carried

out two or three times (iterations) and may also involve the retrieval of deeper

links within domains (depth).

The software produces two types of output that are very useful for describing the issue

network. First, it produces a list of websites ranked by number of inlinks from the web

pages of other domains in the network.[16] Second, it produces a network map of each site

that is included in the network. Each site is represented as a circle, and the size of the

circle reflects the number of inlinks. Arrows between the circles represent each link from

one site to another. One can view all of these links in concert or hide all those except the

links into and out from a selected site. Further, the location of each site's circle represents

the frequency with which it is linked in common with the sites around it. If pages that

---

[16] One can rank them by number of linking pages or number of linking sites. For instance, if seven different Public Knowledge pages link to one or more pages on the Progress and Freedom Foundation domain, that counts as seven votes for PFF in the former case (ranked by page) and one vote in the latter case (ranked by site). This project uses the rankings by page for two reasons. First, it represents more data; it provides a finer-grained description of the degree of variation between the lowest- and highest-ranked sites. Second, it is the means by which the Issue Crawler tracks changes in site ranks over time. (To access these over-time comparisons, from the "Network Details" page, choose "Compare the networks in this schedule.")

link to Site A almost always include links to Site B, then the circles for A and B will be very close. If Site C is almost never linked from the pages that link to A and B, the circle for C will be very far from A and B (see Rogers, 2004, p. 25).

Consider a hypothetical example of how the crawler would work in practice. Imagine a crawl that starts with two seed URLs, the Annenberg schools' websites: http://www.asc.upenn.edu and http://annenberg.usc.edu. The first iteration would crawl both sites for hyperlinks, following those links for sites to crawl in the next round. Further, suppose both Annenberg sites link to the websites for the International Communication Association (ICA), the National Communication Association (NCA), and the Association for Education in Journalism and Mass Communication (AEJMC). In the second round, the crawler would examine these associations' websites, looking for sites to which ICA, NCA, and AEJMC link. As part of their membership lists, these sites may provide links to many other colleges' websites. These would then be crawled in the third round.

One can specify both the crawl depth in which each root domain is examined and the number of iterations outward from the original seed URLs. One can also privilege the seed URLs, meaning that they remain after the first round even if less than 2 other sites link back to them. For identifying issue networks, Rogers recommends 2 iterations outward and a crawl depth of 2,[17] and he recommends not privileging the seed URLs; the

[17] If one intends to use the Issue Crawler, it is important to note that the number assigned to the parameters—the number of iterations of drilling into a web page's root domain, and the number of iterations outward—reflects the number of rounds of crawling minus 1. For instance, each can be set to zero, which would crawl the seed URLs only—one

crawls in this study use these settings. Finally, one can schedule regular crawls, allowing

one to study the network's changes over time or to summarize the results, minimizing the

impact of temporary fluctuations.

Identifying and Ranking DRM Policy Advocacy Websites

Web graph results from the Issue Crawler are not, by themselves, adequate to the

task of describing an online issue space. Without studying the websites themselves, one

cannot learn what an organization represents, how much their website has to say about a

given issue, or why it is likely to have been included in the web graph results. Farrall

(2005) urges the use of content analysis of links to better understand their context; the

nested links are the point of interest, and the textual context of the link helps explain

those relationships.

While this project reports a broad description of the interlinking between sites, it

primarily uses the Issue Crawler as a tool for constructing snowball samples. Where

Farrall (2005) uses content analysis primarily as a means of better understanding the links

between sites, this project uses the links between sites primarily as a means of better

identifying authoritative websites to content analyze. The goal is to quantify which side's

arguments are best represented online, and the Issue Crawler is the first step in

identifying a sample of the documents that bear these arguments.

---

round deep and one round outward. This is akin to the European system of numbering
floors in a building. In the US, the first floor is the ground floor. In Europe, floor number
one is one story above the ground—what Americans would identify as the second floor.
As Rogers is a professor of media studies at the University of Amsterdam, this metaphor
may offer some mnemonic value.

I begin with the sites of five organizations I know to be intimately connected to the copyright policy space: the US Copyright Office, the Recording Industry Association of America (RIAA), the Consumer Electronics Association (CEA), Public Knowledge, and the Progress and Freedom Foundation (PFF).[18] Other than the Copyright Office, each group is also substantially involved in other policy areas. Thus, rather than starting with the root URL for each, I chose copyright- or intellectual property-specific pages for the other four organizations. These are:

1. The website for the U.S. Copyright Office, http://copyright.gov

2. From the RIAA website: http://www.riaa.com/issues/piracy/default.asp

3. From the CEA website:

   http://www.ce.org/aboutcea/ceainitiatives/viewInitiativesOverview.asp?na
   me=253&title=Fair%20Use/Preserving%20Betamax

4. From the PFF website: http://www.ipcentral.info/

5. From the Public Knowledge website:

   http://www.publicknowledge.org/articles/49

The first crawl was on October 20, 2006. After verifying that this produced a network that included the vast majority of policy advocates I considered important to the debate, I scheduled it to run roughly every month until October 23, 2007, creating thirteen web

---

[18] This list reflects the deliberate attempt to create balance among the seed websites. The US Copyright Office is the obvious source of official government positions on copyright. The RIAA and CEA represent generally opposing views for their respective industries, and PFF and Public Knowledge are nongovernmental organizations (NGOs) who are also generally at odds. This web crawl uses the default settings for the web crawl. For more, see Farrall (n.d.).

graphs and—more importantly for the sake of formal analysis—thirteen data sets. The resulting networks contained between 86 and 95 domains.

The median and mean ranks of each site over the duration of the crawling period provide a convenient way to summarize this ranking data. Yet there are many occasions on which many of the sites are not included as being in the network. Thus, a reasonable calculation of mean and median requires estimating a ranking equivalent of not being in the data set—a penalty for failing to make the grade. Since those at the very bottom of the rankings in any given crawl received links from just 3 or 4 separate pages, compared to top numbers of hundreds or thousands, it makes sense to choose a "missing" number that does not provide undue additional penalty. Thus, in all cases where a site is not ranked, they have been assigned a rank of 99. Each site with a median rank of 99, then, was missing from the network more often than not; since I am only interested in sites that are a consistent part of the debate, my analysis here discards these sites. Using this criterion, 78 sites remained for analysis, and 132 sites were discarded.

The strategy of choosing more specific web pages for each seed website—relative to using the root domain (e.g., http://publicknowledge.org) for each organization— appears to have come with neither substantial strengths nor weaknesses. All five URLs were the single best locations for finding the site's recent, relevant content when the crawl started. As of March, 2008, this is no longer true for three of the sites, creating modest concerns with two. First, the URL for the CEA came to redirect users to a site labeled "CEA Initiatives," which is at http://www.ce.org/Membership/3495.asp. This page directly linked to their new "Fair Use/Preserving Betamax" page, at

http://www.ce.org/AboutCEA/CEAInitiatives/3631.asp. Assuming the Issue Crawler ignores such redirects in counting the depth of crawls, there are no problems here; even if not, since this page links directly to the "Fair Use/Preserving Betamax" page, the core of the site's discussion of the issue is still within the Issue Crawler's reach at the specified depth setting.

More potentially problematic is the case of the RIAA. The site above became a dead link over the course of the study. Yet even this appears not to have created any problems. RIAA.com was included in 11 of the 13 crawls, and in the two in which it was missing, RIAA.org was included. Thus, the organization's place in the network was not compromised by its not being included as a seed URL; other sites identified it as relevant. This helps assuage concerns about the impact of the choice of seed URLs, a point discussed in more detail below.

The third seed in which the chosen URL has proven less effective is Public Knowledge. The group's bloggers stopped using the tag for "intellectual property" during the study, so while the URL above was still valid, it remained unchanged for months. This would have created a third problem site, but the last post therein was on October 15, 2007, just days before the final crawl used in this study. Even were this not the case, the Issue Crawler would have been able to dig through the site to find many other new, relevant pages within the specified depth.

These results have a very high degree of face validity; they conform to my previous understanding of the important actors within the online network of policy actors. Table 4.1 lists the top ten websites, sorted by mean rank over time:

Table 4.1: Top Ten Web Crawl Results by Mean Rank

Rank	Organization/Individual	URL	Mean Rank
1	Creative Commons	creativecommons.org	1.77
2	Electronic Frontier Foundation	eff.org	2.38
3	Free Software Foundation	fsf.org	3.85
4	Lawrence Lessig (Stanford Law)	lessig.org	6.69
5	Center for Democracy & Technology	cdt.org	14.46
6	Public Knowledge	publicknowledge.org	16.00
7	Thomas, Library of Congress	thomas.loc.gov	18.54
8	Stanford Center for Internet and Society	cyberlaw.stanford.edu	18.77
9	Consumer Project on Technology	cptech.org	21.31
10	Berkman Center for Internet & Society	cyber.law.harvard.edu	22.69

I provide a more formal test of the validity of these results below, as part of the test of seed dependency. I also provide a full list of results, a description of their relationships to each other, and an accounting of the number of relevant documents per site in Chapter 8. For now, it is enough to say that 9 of these 10 sites are highly involved in the debate over copyright law and are highly linked from the rest of the internet—not merely other sites in this population. The only exception, Thomas, is an official repository for current legislative information and is thus an important reference point for federal policy debates in general. Even before formal testing, these results suggest the validity of this project's web graph analysis.

Choosing Web Graphing Over Simple Searching

The skeptic might ask, "Why not just use a search engine such as Google to search for items related to each policy in which you're interested?" I describe this method, often colloquially referred to as "Googling," as "simple searching." As described

below in the section on content analysis, this project uses Google as a second-stage method for identifying relevant documents within each website included in light of the Issue Crawler results, so I clearly hold Google searches to be of some value. Yet Appendix C documents the many perils of using Google even in this context, and for these and other reasons, a web crawl is a superior first-stage means of identifying websites to study.

First and foremost, a simple search would be a poor reflection of the means by which the bulk of online policy advocacy takes place, which is in clusters of densely interlinked sites that pay close attention to specific issues (Benkler, 2006; Bennett & Manheim, 2001; Rogers, 2004). Online issue advocates seek to affect policy outcomes through highly interested issue publics (Bennett & Manheim, 2001), and these audiences do not need to Google the entire internet for relevant information once they have already discovered the core advocates' websites. For this reason alone, choosing a method that seems to capture the heart of an issue network is vastly preferable to one that dilutes that network's collective opinion with that of thousands of other loosely interested actors—at least it is preferable if, as in this project, one is primarily interested in studying the use of the internet for issue advocacy.

Second, as discussed in Appendix C, Google returns at most 1,000 results for any given search. As described below, this research project investigates up to roughly 100 documents from each of the 78 websites studied for each of the two policies studied. As a result, this research sorts through several thousand documents per topic, providing superior coverage. When combined with the greater focus on the network of sites that are

107

explicitly dedicated to advocacy on these issues, this higher level of coverage is superior for identifying and classifying the highest number of relevant documents.

Third, the Issue Crawler permits the inclusion of valuable data: a sophisticated description of the relationships between relevant sites. No simple search can provide this, but the inclusion of even a few maps from the series of web crawls provides a helpful illustration of the network of online actors—let alone the detailed quantitative measures of inlink authority discussed in Chapter Eight.

Fourth, the results of a test simple search using Google reaffirms this study's findings and highlights the relative strength of the web graph method. Using the same search instructions identified below for searching individual sites for broadcast flag-relevant documents, a search of the entire web[19] is instructive. Of the first 100 results, 44 are from sites included in the web crawl results. Of the remaining 56, 23 are not relevant to the debate. This leaves 33 relevant documents from other websites. These 33 documents show the same bias toward the strong fair use coalition as do the documents identified during the web crawl; 11 are neutral or mixed, and 22 support strong fair use. This reinforces the conclusion that the strong fair use coalition dominates the web debate.

These 33 documents are almost entirely from sites that are more authoritative according to the entire web but are less engaged in the regular debate over DRM regulation. These include online versions of print media sources such as *USA Today* and *Wired*, and David Pogue's *New York Times* blog (pogue.blogs.nytimes.com). These also include visible internet sources such as Wikipedia, the O'Reilly Network, G4, and CNet.

---

[19] Conducted March 10, 2008.

Other less-visible sites include blogs such as TechLiberation.org and Copyfight (copyfight.corante.com) that feature regular commentary on the copyright debate. Their inclusion in the web graph results would have been unsurprising, but they are not central parts of the online copyright advocacy network. The only site in this set that represents an interest group that is regularly engaged with policymakers in the fight over copyright regulation is the Library Copyright Alliance (librarycopyrightalliance.org). This is a minor omission at best, however. The Alliance is comprised of 5 library groups—most notably the American Library Association (ala.org) and the Association of Research Libraries (arl.org), both of which are included among the 78 websites studied.

To the extent that the results from this test are included in the results from the web graph analysis, the choice between methods is a wash. To the extent that both tests suggest the same overall rhetorical direction, the simple search results reinforce those of the web graph analysis. This overlap even suggests that the results in terms of rhetorical valence are not particularly dependant upon methodological choice.

Still, there is a real difference between these methodological choices. A simple search may actually be a better indicator of what the average person would discover upon first researching an issue. Likewise, any potential third-person effect of the internet—for instance, if a congressional staffer tries to search the internet to investigate the range of opinions on an issue—is more likely to occur via simple searching. Yet the web graph method is superior for identifying the regularly participating policy actors and measuring their levels of participation and relative authority within the community. Major periodicals and highly visible technology sites may have a larger audience, but the

109

literature on political mobilization via the internet (see, e.g., Bennett & Manheim, 2001; Benkler, 2006) suggests that the goal of online communication is not just to reach a large audience, but to mobilize issue publics—the kind of people who already have a level of familiarity with an issue that reduces the need for general searching.

Sites such as David Pogue's blog and Wired may get citizens interested in copyright policy, but if they get more directly involved, it will most likely be via groups such as the Electronic Frontier Foundation. Web graph analysis better reflects the way internet activism works: within clusters of densely interlinked sites, and largely in communication with those who need the least generalized information about an issue.

Establishing Validity of Seed URLs

It may seem as though the seed URLs chosen for the repeated crawls can exert an arbitrary or capricious control over the result that is produced. This can be described as a concern about seed dependency: do the results depend on the seed URLs chosen? Can an online issue network be misrepresented, intentionally or unintentionally, based on the choice of seeds? To a certain extent, this is possible. For instance, if one wanted to create the impression that the only voices on the issue are pushing for more copyright, one could enter the websites of the RIAA and similar groups, such as the Motion Picture Association of America and the Business Software Alliance. These groups link to each other, link to other friendly organizations, and almost never link to the organizations on the other side of the debate. Thus, the resulting network would be misleading.

As an initial move to suggest face validity, the seed URLs include organizations from both coalitions. As long as these seeds include at least one group in each coalition

that links heavily within the debate, and as long as two included organizations that

disagree on the issues nonetheless agree to some extent on what other organizations are

important participants, the resulting graph will generally represent the issue network at

hand with a reasonable degree of fidelity.[20] That is the case here. The Copyright Office,

RIAA, and CEA are not particularly heavy participants in the online debate; they produce

fewer online documents than the NGOs in the network, and they tend not to link back to

the rest of the debate. In contrast, the Public Knowledge and PFF sites both feature

frequently updated blogs, countless links to other groups' websites, and links to pages

and organizations with views opposite their own. In short, these two sites are the vital

links in the seed URLs chosen. They have helped identify the overall character of the

online issue network and the bulk of important participants.

To test the inclusion of important copyright actors, I ran 10 test crawls. I ran these

in rapid succession, reducing the effect of actual changes within the network. For the first

five test crawls, I used five combinations of URLs, each including four of the five

original seed URLs. For test crawls six through ten, I ran five combinations of URLs

excluding the original seeds; as a means of identifying other relevant websites, I used the

list of websites returned during the series of scheduled crawls from October 2006 through

July 2007, sorted by average rank. I ran crawls of sites with the following ranks:

---

[20] There are other scenarios in which a representative graph could also result. For
instance, if the seed URLs from different coalitions do not generally agree on which sites
are important to the debate, the inclusion of two or more seeds from each coalition that
are similarly active linkers could also produce good results.

- 1 to 12, excluding Public Knowledge (then ranked #4) and the Copyright Office (#11).

- 13 to 23, excluding the RIAA (18)

- 24 to 35, excluding the PFF (26) and CEA (31)

- 36 to 45

- 46 to 55

All crawls were run between August 22 and August 25, 2007. I compare these results against the averaged results that build on the year's worth of crawls.

As explained in Chapter 2, online networks feature a power law distribution of attention, with the most-linked sites exerting highly disproportionate influence on the rest of the group. Thus, in testing these supplemental crawls against the scheduled crawls, it is important to look at the most authoritative members of each test network; I operationalize this as the ten most-linked sites. Table 4.1 above recounts the ten most-linked sites over the year of data collected. These results make for instructive comparisons with the results from the ten test crawls. The key questions in each comparison are:

- Does a test crawl do a better job prioritizing sites that are truly central in the copyright debate?

- Does a test crawl identify sites that are fairly described as substantially participating in the online debate that were *not* identified by the scheduled crawls?

- Would the inclusion of any missing sites change the overall understanding of the network?

It would be hard to top the results from the scheduled crawls over time; a test crawl would need to have all 10 of its top sites consist of relevant, highly linked policy advocates. Further, as shown below, few relevant advocacy sites are identified as central by test crawls but not included in the final 78 sites studied. Even those few sites that are missed do not lead to a different understanding of the network as represented by the scheduled crawls; rather, they fit in nicely with those sites already included.

Again, the first five validity tests explored the results when one of the five original seed URLs was dropped. For instance, the first crawl dropped CEA, so the crawl's four seeds were Public Knowledge, the Copyright Office, the Progress and Freedom Foundation, and the RIAA. The top 10 results (up to 12, including all sites tied for 10th) for three of these test crawls are recounted in Table 4.2:

Table 4.2: Seed Validity Tests One, Two, and Three

Rank	Test 1, w/o CE.org	Test 2, w/o Copyright.gov	Test 3, w/o RIAA.com
1	thomas.loc.gov	eff.org	eff.org
2	eff.org	creativecommons.org	copyright.gov
3	uspto.gov	loc.gov	creativecommons.org
4	fcc.gov	thomas.loc.gov	ala.org
5	creativecommons.org	copyright.gov	publicknowledge.org
6	copyright.gov	uspto.gov	thomas.loc.gov
7	icann.org	wipo.org	fcc.gov
8	cyberlaw.stanford.edu	icann.org	epic.org
9	publicknowledge.org	ala.org	icann.org
10	wipo.org	epic.org	uspto.gov
11		cyberlaw.stanford.edu	cyberlaw.stanford.edu
12		publicknowledge.org	

These results are clearly consistent with the scheduled crawls over time. Each of these sites is included among the 78 studied; the lowest rank among them is 31

(wipo.org). In short, dropping any of these three sites does not substantially change the types of sites represented as being important actors in the network.

In contrast, consider the results from dropping either Public Knowledge or the Progress and Freedom Foundation, the two seed sites that are active participants in the online debate. These are shown in Table 4.3:

Table 4.3: Seed Validity Tests Four and Five

Rank	Test 4, w/o PublicKnowledge.org	Test 5, w/o IPCentral.info
1	nashvillesongwriters.com	ascap.com
2	bmi.com	bmi.com
3	grammy.com	grammy.com
4	cisac.org	fcc.gov
5	usa.gov	soundexchange.com
6	musicunited.org	riaa.com
7	mp3.com	eff.org
8	songwritersguild.com	musicunited.org
9	artistdirect.com	nashvillesongwriters.com
10	soundexchange.com	usa.gov
11		regulations.gov
12		uspto.gov

Compared with crawls conducted using the full set of seed URLs, these sets of websites are inadequate to the task of capturing the heart of the online copyright and DRM debate. For instance, none of the 10 sites in the first list, based on the crawl without Public Knowledge, have a single page that would be coded as relevant to the DMCA reform debate.[21] In the top results from test crawl five, several sites are members of the

---

[21] See below and Appendix B for a fuller description of how this decision of relevance is made.

online debate—the FCC, the RIAA, the EFF, and the US Patent and Trademark Office—but each of these is also represented in the 78 sites included in this study. This list, however, contains a lower ratio of sites with DMCA-reform-relevant documents (3 of 12) than the top results from any of the individual scheduled crawls (minimum 7, maximum 10 of 12). While these lists diverge substantially from the results of the scheduled crawls, the contrast only serves to highlight the relatively high quality of the scheduled crawl results—especially when averaged over time.

It is perhaps somewhat less surprising that these first five test crawls would not substantially challenge the results from the scheduled crawls. The first three tests used the two best seed URLs and still produced reasonably good—and similar—results, while the last two used only one good seed and thus produced results of little value. Similarly, the results from the next five test crawls—using groups of ten sites excluding the original five seed URLs—suggest that the scheduled crawls capture the essence of the network. For instance, consider Table 4.4, displaying the results from the first of five crawls using seeds not included among the five chosen seed URLs:

Table 4.4: Seed Validity Test Six

Sites Used	Top 10 Results
eff.org	creativecommons.org
thomas.loc.gov	eff.org
creativecommons.org	icann.org
cyber.law.harvard.edu	cyber.law.harvard.edu
www.fcc.gov	w3.org
uspto.gov	publicknowledge.org
cyberlaw.stanford.edu	gnu.org
apple.com	fsf.org
wipo.org	epic.org
icann.org	lessig.org
	cyberlaw.stanford.edu

In these results, 8 of the 11 sites are included in the top 24 websites in the scheduled crawls. The sites icann.org, w3.org, and gnu.org are not included in a majority of the scheduled crawls, so they are not included in this research. There is no concern that icann.org is excluded; it has zero documents that are relevant to either debate, so it can hardly be identified as central to the DRM debate. In contrast, gnu.org and w3.org are false negatives; despite being visible and relevant, they were rarely returned in the results of scheduled crawls.

The more troublesome of the two false negatives is gnu.org, the website for the GNU software development project, which is best known for the GNU/Linux free operating system. It is both visible and relevant. The site has a Google PageRank of nine out of ten, a mark of substantial online distinction. As of March 8, 2008, it features one document strongly condemning the proposal for a broadcast flag mandate and several documents that are relevant to the DMCA reform debate.

Despite these reasons for thinking that the exclusion of gnu.org is a problem, the scale of the problem is slight for two reasons. First, gnu.org is but one of two highly visible sites run by the Free Software Foundation; the other, fsf.org, is ranked number three in the summary results from the year's worth of scheduled crawls. Thus, the Foundation's views have been included in the results of this research. Fsf.org, which also has a PageRank of nine, is much more explicitly dedicated to discussing the political and social implications of copyright law, including DRM policy. In contrast, gnu.org is the hub for learning about GNU software, including free distributions of GNU/Linux. Gnu.org also contains a surprising number of documents expressing the Foundation's beliefs regarding the perils of DRM and the regulations that give DRM teeth, but these are of the same general character as those on fsf.org: vehemently opposed to laws that limit end user's rights to use and modify copyrighted works.

The second reason the site's exclusion is of little concern is that, in taking this strong fair use position, it is of the general character of the online issue network described by this research. As discussed in Chapter 8, the strong fair use coalition so thoroughly dominates the online discussion that the exclusion of a few more strong fair use coalition sites does not create a substantial risk of Type II error.

The second false negative here is the World Wide Web Consortium site, w3.org, which only appeared in two of the scheduled crawls. The site is even more visible, being one of just a handful of sites with a PageRank of ten. While the Consortium is an important part of the network of actors responsible for determining the future of the internet, it is not an important part of the online debate about DRM policy. The site

117

contains no documents relevant to the broadcast flag debate and only 2 that are relevant

to the DMCA reform debate. Both of these argue against legal restrictions on the

circumvention of DRM. Again, since this is true of the majority of documents included in

the studied population, this site presents little concern about the validity of this study's

findings.

The results from the next two test crawls posed similarly minor issues. The top

websites from test crawls seven and eight are in Table 4.5:

Table 4.5: Seed Validity Tests Seven and Eight

Test 7: Sites Used	Results	Test 8: Sites Used	Results
aclu.org	eff.org	cdt.org	usa.gov
ala.org	epic.org	ftc.gov	icann.org
fsf.org	aclu.org	house.gov	ftc.gov
lessig.org	creativecommons.org	cptech.org	thomas.loc.gov
mpaa.org	thomas.loc.gov	hrrc.org	uspto.gov
epic.org	icann.org	dfc.org	ntia.doc.gov
chillingeffects.org	w3.org	cato.org	eff.org
loc.gov	cdt.org	fairuse.stanford.edu	w3.org
cpsr.org	publicknowledge.org	wired.com	whitehouse.gov
futureofmusic.org	ftc.gov	bmi.com	bbb.org
	cyber.law.harvard.edu		fcc.gov
			consumer.gov

In these results, two sites that contain some relevant documents are missing from

this study's final analysis. The first is w3.org, discussed above. The other false negative

is the site for the National Telecommunications and Information Administration,

ntia.doc.gov. The DMCA requires the NTIA, which is a part of the Department of

Commerce, to provide additional input for Congress and for the Librarian of Congress on

118

the effects of the restriction on circumventing DRM. The website provides no documents

relevant to the broadcast flag debate, but it does feature five DMCA reform debate

documents. These are a microcosm of the population of documents from the sites under

study: one is opposed to DMCA reform, one is mixed, and three support DMCA reform.

This site's exclusion therefore also creates little concern.

The only source of minor concern is that government actors' voices will be

misrepresented. The NTIA is certainly more skeptical of the value of ever-stronger

copyright law than other government agencies—most notably the Copyright Office and

the Patent and Trademark Office. Yet the internet is not the best location for identifying

these agencies' policy positions. Their websites often serve as repositories for public

filings, so even the Copyright Office domain contains a remarkably high proportion of

pro-fair use documents. Rather, congressional testimony is an excellent place to identify

an agency's position on a given issue, and an agency's relative frequency of testimony

suggests its relative centrality in the debate. The fact that the NTIA rarely testifies on

copyright issues, while the Copyright Office is constantly represented, correctly suggests

the NTIA's minor role in this debate.

Finally, consider the results from test crawls 9 and 10. These are in Table 4.6:

Table 4.6: Seed Validity Tests Nine and Ten

Test 9: Sites Used	Results	Test 10: Sites Used	Results
youtube.com	eff.org	eldred.cc	usa.gov
acm.org	epic.org	freedom-to-tinker.com	hhs.gov
bsa.org	aclu.org	citizen.org	whitehouse.gov
sesac.com	thomas.loc.gov	emusic.com	grants.gov
archive.org	icann.org	wto.org	firstgov.gov
digmedia.org	ftc.gov	musicnet.com	cdc.gov
mediaaccess.org	cpsr.org	intel.com	regulations.gov
musicunited.org	fcc.gov	hp.com	pandemicflu.gov
wipo.int	publicknowledge.org	pewinternet.org	thomas.loc.gov
napster.com	ntia.doc.gov	www4.law.cornell.edu	nih.gov
			section508.gov
			fedbizopps.gov

These results uncover no new problems. From Test 9, only 1 of the results is not included: ntia.doc.gov, which was discussed above. Remarkably, the results from Test 10 are all irrelevant government websites; unsurprisingly, cdc.gov and pandemicflu.gov are not filled with statements about the DMCA reform debate.

These test crawls suggest little chance of skewed results caused by seed dependency. The most useful results resembled the results from the scheduled crawls, and deviations from the scheduled crawl results were generally much less useful. This can be formalized by taking each result above as one data point—including duplicates, since each repeated instance of an included, relevant site is an additional reason to support the final results and each instance of the same missing, relevant site is an additional reason for skepticism. Thus, the test crawls yield 111 data points. 63 of these sites have at least one relevant document. Of these, only 6 are relevant but excluded from the final results: gnu.org, w3.org (three times), and ntia.doc.gov (twice). Thus, the scheduled crawls

retrieved 57 of 63 of the relevant results from the test crawls, or just over 90% success. Further, for the reasons discussed above, none of the excluded sites are particularly problematic because none represents a uniquely important part of the debate. Based on these results, the scheduled crawls appear to have captured all of the most important sites from the issue network. Thus, the over-time results from the web graphs can be used as a basis for locating relevant sites to content analyze, and these results may be compared in a meaningful way with those from newspapers and congressional hearings.

Content Analysis

The goals of the content analysis include quantifying the number of relevant documents in each medium, documenting each document's basic characteristics such as date and type of organization represented, and assessing each document's overall valence along the continuum from strong copyright to strong fair use. In this section, I first discuss reliability and validity. Second, I describe the means for identifying relevant congressional documents. Third, I describe the inclusion of newspaper articles. Fourth, I detail the means for identifying relevant documents from the sites identified by the web graph analysis; here, I describe a new method for documenting the recall and precision of Google searches. Fifth, I describe the basic details for which each document is coded. Sixth, I describe the process of coding for rhetorical valence.

Content Analysis: Reliability and Validity

An important component of content analysis is establishing high intercoder reliability; this ensures that outcomes are not dependent on the subjective views of a

single coder. I measure intercoder reliability using Krippendorff's alpha (Krippendorff, 2004), signified simply as α. Krippendorff suggests researchers can generally rely on variables with an α of at least .8 and treat with caution variables with an α of less than .8 but at least .667 (p. 241). I follow these guidelines here and report α for each variable used. Each α represents the comparison of my codes with one other coder who was trained for that variable. Because of the large number of variables—primarily, medium- and topic-specific instructions for whether to include or exclude a given document or paragraph—several coders were used, each coding for one or more variables. Except where noted, second coders examined a random sample of coding units. In constructing a sample of documents for calculating relevance, I followed Krippendorff's guidelines (p. 240) for estimating the requisite sample size.

All groups of documents presented here are retrieved via computer-assisted search tools. These present substantial challenges in establishing recall and precision of search methods; while few researchers report estimates of recall and precision, these benchmarks are important in establishing the validity of the search strategy. Recall is a search term's ability to retrieve relevant documents, and precision is a term's ability to exclude irrelevant documents (Stryker, Wray, Hornik, & Yanovitzky, 2006).

For instance, imagine that a database contains 100 relevant newspaper articles. A search that retrieves 50 documents, including 48 of the 100 relevant articles, has very high precision; 96 percent of the retrieved documents are relevant. In contrast, it has very low recall, retrieving just 48% of the total relevant documents in the population. Another search strategy that retrieves 190 documents, 95 of which are relevant, has a recall of

95% but just 50% precision. For contexts where the number of relevant documents is likely to be in the thousands, Stryker, Wray, Hornik, and Yanovitzky (2006) have developed a valuable method for estimating recall and precision. I call this the "Stryker method." Precision is simple to calculate: ask what percentage of the retrieved documents are relevant. Estimating recall with the Stryker method is more challenging. The authors refer to a researcher's targeted search term as the "closed" term and recommend testing this against a representative sample of documents retrieved using an "open" term intended to retrieve literally every relevant document. Because the required sample of relevant documents will generally be in the hundreds, this method for calculating recall is less appropriate for populations in the dozens.

For most of the populations of documents in this study, the number of relevant documents is so small that the Stryker method is inappropriate. In these cases, I conduct a search using the open search term—for instance, requiring merely that newspaper articles' headline, lead paragraph, or search terms contain the word "copyright"—and sort through documents by hand. In these cases, recall is very nearly 100%, and while search term precision is very low, hand coding ensures that only relevant documents are used. The search for relevant internet documents, in contrast, scours through thousands of documents. Thus, as described below, I adapt the Stryker method to internet search strategies and produce estimates for recall and precision of closed search terms.

Identifying Relevant Congressional Documents

As uncovered using the LexisNexis Congressional database, I take a census of relevant congressional hearings from within the four-year windows that best capture the

peak of debate over each of the policy issues under study. As one of Kingdon's (2003) policy actor respondents argues, "In Washington, the world of ideas is like the world of fashion. Ideas don't last for more than four or five years" (p. 105). Because four years also represents two sessions of Congress, this makes for a naturally clean break. As operationalized in this study, these windows are: for the Audio Home Recording Act (AHRA), 1989-1992; for the Digital Millennium Copyright Act (DMCA), 1995-1998; and for the two recently stalled attempts at legislation—one to reform the DMCA, and the other to impose a radio or television broadcast flag—2003-2006.

Using the open term strategy identified above, I searched for any hearings over the last 20 years that have "copyright" in the subject and do not have "appropriation" in the title. On November 19, 2007, this search retrieved 199 hearings, including 128 total from the 3 time periods under study: 48 from 1989-1992, 30 from 1995-1998, and 50 from 2003-2006. Of these, a total of 17 were coded as potentially relevant: 4 from 1989-1992, 6 from 1995-1998, and 7 from 2003-2006. Coders agreed perfectly ($\alpha = 1$) on coding of relevance for all 199 LexisNexis hearing summaries.

These hearings in hand, I then separated each speech, written testimony by a witness who also spoke, and other written submission for separate coding. If an item stands on its own as making a complete set of arguments about one of the policies at hand, it is included. If an item is clearly paired or coupled with another item such that its significance is only as an accompaniment—for instance, a brief letter to a committee chair introducing a journal article into the written record for a hearing—such items are considered as a single document.

Again, each was coded for relevance before applying additional codes. Intercoder agreement on document relevance was $\alpha = .92$ (71 pairs) for the AHRA, $\alpha = .86$ (70 pairs) for the DMCA, $\alpha = .86$ (70 pairs) for the broadcast flag, and $\alpha = .80$ (70 pairs) for proposals to reform the DMCA.

<p style="text-align:center">Identifying Relevant Newspaper Articles</p>

I also study print news stories that discuss the four bills within the same four-year windows. In the belief that the two newspapers have disproportionate influence on the discourse around policy options, this study focuses on *The New York Times* and *The Washington Post*. While neither role is exclusive, the former is commonly understood to set the national media agenda (McCombs, 2004, p. 113) and the latter to be a direct route to policymakers' ears (Gandy Jr., 1982, p. 13; Kingdon, 2003, p. 59). In order to ensure virtually complete recall of a very small number of relevant articles, the search term required only that articles have "copyright" in the headline, lead paragraph, or search terms. The results included 284 articles from 1989-1992, 896 from 1995-1998, and 1431 from 2003-2006. Remarkably, only the smallest of the three groups—the earliest— contained enough relevant articles (25) to establish intercoder reliability ($\alpha = .89$, 61 pairs). For the other two groups, the number of relevant articles was sufficiently low to justify the inclusion of articles from *Billboard* simply to have enough articles to estimate an α. For the DMCA debate from 1995-1998, I supplemented 14 relevant articles from the *Times* (6) and *Post* (8) with eight articles from *Billboard* to reach $\alpha = .93$ (61 pairs). For the broadcast flag debate ($\alpha = .96$, 70 pairs), I added nine *Billboard* articles to ten

from the *Times* (6) and *Post* (4). The DMCA reform debate featured 15 articles in the *Times* (9) and *Post* (6); I thus added five from *Billboard* to reach α = .93 (70 pairs).

<div align="center">

Identifying Relevant Internet Documents

</div>

As discussed above, I use the results of the web graph analysis to identify websites that are part of the population of sites that seek to influence copyright policy. This still leaves the problem of identifying individual documents on each domain that are relevant to the debate at hand. This is a much trickier problem than identifying relevant newspaper articles or congressional hearings; internet documents are troublesome to identify reliably, as discussed in Appendix C. Here, I briefly outline the goals and overall strategy for identifying individual documents. I then provide an extended explanation of my adaptation of the Stryker method (Stryker et al., 2006).

Google Searches: Goals and Strategy

The web graph analysis identify 78 web domains as central in the online debate about copyright. None of these are exclusively devoted to discussing the recent policy debates under study. Thus, it is important to quantify and characterize the individual internet documents in which authors make arguments. As with most of the web, these are most often HTML files, but they are also in formats such as PDF, Microsoft Word, PowerPoint, or Excel. [22]

---

[22] I do not include multimedia files such as video, audio, or flash animations. These are often difficult to archive and share for intercoder reliability checks. A very small proportion—far less than 1%—of total documents fall into this category.

In order to identify relevant documents, I use Google to search each web domain. For instance, in order to identify documents that relate to the broadcast flag on the EFF domain, I use the following search term:

copyright (audio OR video OR radio OR broadcast) flag site:eff.org

On November 7, 2007, Google said this search retrieved "about 835" documents. The results for the same search on the Public Knowledge domain, publicknowledge.org, returned about 2,350 documents. A very high number of these are false positives, but that was true of the newspapers as well; what differs is the sheer scale of the results. With 76 more websites to examine on just one of the two topics studied, there are already more documents on the table (3,185) than the entire collection of newspaper articles described above (2,611). Using a similar "code everything" method was thus impractical.

On the upside, this large population of documents is amenable to sampling in a way that the smaller piles of newspaper articles and congressional hearings are not. Further, as discussed next, Google has a strong ability to front-load the most relevant results. What is needed, then, is a means of maximizing the quality of search results and estimating the number of documents left behind on each given domain. In short, what is needed is an estimate of search terms' recall and precision.

Testing Recall and Precision of Google Searches

The quality of a search term can be expressed as a combination of its recall and precision. As Stryker et al. (Stryker et al., 2006) explain:

> Recall and precision are both proportions. Recall is an estimate of the conditional probability that a particular text will be retrieved, given that it is relevant, calculated by dividing the number of relevant items returned by a search phrase

by the total relevant records in the database. Precision is an estimate of the conditional probability that a particular text is relevant, given that it is retrieved, calculated by dividing the number of relevant items by the total number of items returned in a specific search. (pp. 414-415)

A search term should retrieve as many relevant documents as possible while excluding as many irrelevant documents as possible. Testing precision is the more straightforward task, asking: Of the documents retrieved, how many were relevant? Testing recall, however, requires that one "capture any and all [documents] relevant to the topic of interest" (p. 416) for the sake of comparison. Since examining every document from within a database is generally impractical (otherwise, using search terms would be unnecessary), this method tests a narrow search term, or "closed" search, against the results of a much broader "open" search.

The Stryker method was developed and tested on search terms within the LexisNexis database of newspaper stories. In order to avoid the problem of identical or similar stories, the authors test search terms on one randomly selected newspaper (Stryker et al., 2006, p. 20) as a reasonable representation of all the newspapers in the sample. Since their project studied newspaper coverage of cancer, each newspaper would be certain to contain a large supply of relevant stories.

This project studies online documents that discuss either of two specific debates, one over proposals to mandate a video or audio broadcast flag, and the other over proposals to reform Title I of the 1998 Digital Millennium Copyright Act (*Digital Millennium Copyright Act*, 1998b). These are both part of the much broader debate about copyright policy. Thus, the simple term "copyright" is an excellent criterion for a broad search against which one can test the recall and precision of narrower search terms.

Stryker et al. (Stryker et al., 2006) offer the conceptual tools with which to develop a method for testing the quality of internet search terms. The authors "do not claim that the proposed method may be used in evaluating internet searches, as the Web is that much more mutable, not finite or defined in the way databases are" (Stryker et al., 2006, p. 424). Despite this caveat, the concepts of recall and precision are applicable, and the goal of calculating them as a means of estimating search error is important if one is to try to make substantive quantitative claims based on the results of internet searches. Toward this end, I adapt the measures of recall and, to a lesser extent, precision to the internet environment.

This adaptation proceeds in the following stages. I first discuss the selection of a single domain for testing search terms. Second, I make some observations about using Google; these and more are covered in greater detail in Appendix C. Third, I describe the method at a conceptual level. Finally, I apply this conceptual understanding to this specific research project.

Selecting a test domain

As an important difference between online and offline research, consider the selection of a source site for testing search terms. While selecting one major-market newspaper at random as a representative for dozens like it is a fairly straightforward decision for Stryker and her co-authors (2006, p. 8) as they tested their search terms, the random selection of a website among many other websites within this community would have been a poor choice. In Hypothesis 6, I predict that the strong copyright coalition puts far less information online than does the strong fair use coalition. Further, few

websites have a newspaper-like interest in providing a balanced view of the debate; most explicitly advocate for one position or the other, creating the real possibility that search terms would be tested against only one side of the debate. In order not to rig the test to substantially disfavor either coalition, recall and precision must be tested on a site that is reasonably neutral—where many documents featuring the rhetoric of each side are likely to reside.

There are almost no websites in the population that solve both problems—adequate number of relevant stories and viewpoint balance—but thankfully, the website for the US House of Representatives, House.gov, does so. It contains several dozen relevant results for both debates, and both coalitions have enough documents on the site that searches consistently pull results from both sides. Thus, search terms are tested on House.gov, which serves the same role for this project as the randomly selected newspaper does in the original test of the Stryker method.

Using Google

Searching websites for relevant documents is trickier than searching within a closed database. The Stryker tests are all conducted within LexisNexis. In contrast, this project uses a third-party search engine, Google, to search specific websites. Google is perceived as the most authoritative search engine (Rogers, 2004, pp. 41-42), and many web designers optimize their sites for Google searches (Masum & Zhang, 2004), making it a natural choice for querying each site for relevant documents. Yet Google is remarkably disobedient. For more on this, see Appendix C. Of particular import for this methodological problem, one cannot ask for every search result; after 1,000 documents—

or less, depending on the search—Google simply stops returning results. As the site explains, this is by design. "We try to make your search experience so efficient that it's not necessary to scroll past the first ten listings. We understand that some users would like to see more than 1000 results, but this is fairly rare, and it would heavily tax our system to provide these results for everyone" (Google, 2008c). One might call this the "1,001 problem". So even though House.gov may have *thousands* of documents with the word "copyright", Google simply will not list them all.

As a result, even an "open" search will often miss some of the relevant documents in the target population. If a topic is a particularly hot embodiment of a broader debate— as is the case with the DMCA reform debate, relative to the broader copyright debate in the House—these documents may be particularly likely to appear within the first 1,000 hits. Yet this contingency cannot be taken for granted; a researcher needs to account for the number of relevant documents missed by the closed term *and* those not returned within the first 1,000 documents by the open term.

Calculating recall: The method at a conceptual level

To conduct a modified Stryker test for online recall, the researcher conducts a search on the test website using an open search term and codes the results for topical relevance. Because of the ceiling of 1,000 documents or less, the sampling procedures in Stryker (pp. 418-422) will rarely apply; it will generally be necessary to code all results for relevance. One can then conduct multiple searches using more specific closed search terms until a search retrieves an adequately high proportion of the relevant documents

131

retrieved by the open search term. Because both the test website and Google's search algorithm might change at any time, speed is of the essence.

This method for estimating recall might not seem to account for the substantial imperfection of the so-called "open" term. That is, if the open term fails to retrieve all relevant documents, one might object to using it in this way as though it does retrieve all relevant documents. Yet the best estimate of a closed term's recall remains the number of relevant documents retrieved by both the closed and open terms divided by the number of relevant documents retrieved by the open term. This is despite the very real possibility that the closed term will retrieve relevant documents that were not retrieved by the open term. What follows is an illustration of the reasoning that leads to this conclusion.

Once one has chosen a closed term, there are three sets of known relevant documents, each represented by a cell in Table 4.7: the number of relevant documents retrieved by both searches, a, the number retrieved by the open search only, b, and those in the closed search only, c.[23] This acknowledges that there are as-yet-unidentified relevant documents. For the sake of simplicity, this table only includes relevant documents.

Table 4.7: Open and Closed Search Term Recall of Relevant Documents

Relevant Documents	In closed results	Not in closed results
In open results	a	b
Not in open results	c	d

[23] This table is not to be confused with that in Stryker et al. (2006, p. 415), which assumes that the open search term retrieves all relevant documents. In their example, this assumption is tenable, eliminating the need for cells c and d in this Table. Online, this assumption is untenable.

Now, let us fill in the table with some hypothetical results. Imagine a scenario where an open search retrieves 60 relevant documents, a closed search retrieves 70 relevant documents, and these searches share 50 relevant documents in common. In other words, $a = 50$, $b = 10$, $c = 20$, and d is unknown. These figures are shown in Table 4.8:

Table 4.8: Hypothetical Example of Open and Closed Search Term Recall

Relevant Documents	In closed results	Not in closed results
In open results	$a = 50$	$b = 10$
Not in open results	$c = 20$	$d = ?$

In this example, there are 80 known positive documents. The closed search failed to retrieve 12.5% of known positives (10 of 80), while the open search failed to retrieve an even larger proportion of known positives, 25% (20 of 80). In LexisNexis, this would be virtually impossible without substantial error on the researcher's part. The open search is designed to capture 100% of the relevant documents, and the researcher measures the recall of the closed search against this ideal. If Google returned all of the thousands of results for the open search, this unexpected outcome would be unlikely.[24] Accepting the tool for what it is, we are left to estimate how many relevant documents lie beyond number 1,001; we need an estimate for d. Within the sub-sample of 60 relevant documents retrieved by the open search term ($a + b$), we know the closed term retrieved

---

[24] It would not, however, be impossible. Google sometimes returns search results from broader searches that exclude documents returned by narrower searches. Asking very specific queries can sometimes unearth documents not revealed by broader queries that, in a strict application of Boolean search logic, would necessarily include all results of narrower queries, even when neither returns over 1,000 documents. Again, the costs and benefits of using Google search results are discussed in further detail in Appendix C.

50; that is, the closed search retrieved 5 / 6 of this set, or .833. In this case, the ratio of (a / b) = (5 / 1). By analogy, it is reasonable to estimate the same ratio between documents retrieved by the closed term only (c = 20) and relevant documents missed by the open search (c + d); this means the same ratio for (c / d).[25] Put more formally, I postulate that:

$$a / b \approx c / d$$

Solved for d:

$$d \approx bc / a$$

In this example, $d \approx 10 * 20 / 50 \approx 4$. In other words, there are approximately 4 relevant documents that have not been retrieved by either search.

Now, we are ready to estimate[26] the recall of the search term based on this estimate of the total number of relevant documents. The formula is:

$$\text{Recall} = (a + c) / (a + b + c + d)$$

In this hypothetical example:

$$\text{Recall} = 70 / 84 = .833$$

---

[25] This is not a logical necessity, but it follows from what we know of the samples. Within the sample of relevant documents retrieved by the open search, the closed search retrieves the ratio (a / a + b) of these documents. To the extent that this sample is representative of the broader population, we would estimate the same ratio between *all* documents retrieved by the closed term (a + c) and *all* relevant documents [(a + c) / (a + b + c + d)]. The same would hold for documents retrieved by the open search *within* the population of relevant documents retrieved by the closed search; thus, (a / a + c) is a sound estimate for [(a + b) / (a + b + c + d)].

[26] For readability, I drop the use of the character $\approx$ in describing the calculation of recall, but all formulas for recall are estimates.

Thus, to the extent that the researcher believes this site is representative of the broader population, one would estimate that this closed search term reveals approximately 83% of relevant stories.

This figure, .833, may sound familiar. If we begin from the assumption that $a/b \approx c/d$, there is no need to estimate d in order to calculate recall because $(a + c) / (a + b + c + d)$ is equal to $a / (a + b)$.[27] In this case, $a / (a + b)$ also equals .833. In other words, the best estimate of recall from a closed Google search is the number of documents retrieved by both searches divided by the number of relevant documents retrieved from the open search term; this is even when the open search retrieves fewer documents.

A recall rate of .833 is a relatively low figure compared to what is achievable in a closed database. Stryker and her coauthors (2006) insist on a recall rate of .93 for their research project (p. 21). Yet this high level of recall may be difficult to reach with a single online search. As described in Appendix C, Google does not strictly follow the Boolean search logic of the terms entered, so expanding the reach of search terms to include some of the documents from b will often result in the dropping of relevant documents that were retrieved by the original closed search, potentially including results from both a (in both the original closed search and in the open search) and from c (in the

---

[27] First, start with:
$$(a + c) / (a + b + c + d) = a / (a + b)$$
Multiply each side by both denominators to get:
$$(a + c)(a + b) = a(a + b + c + d)$$
$$a^2 + ab + ac + bc = a^2 + ab + ac + ad$$
Subtract a^2, ab, and ac from both sides to get:
$$bc = ad$$
Which is synonymous with our starting premise:
$$b / d = a / c$$

original closed search only). As the authors explain, however, the purpose of the Stryker method is not merely to ensure the high quality of search terms against a single standard. Rather, it provides a means "to estimate the sampling error around an online search term" (Stryker et al., 2006, p. 423). If a researcher's hypotheses tolerate a retrieval of less than 90% of relevant documents—for instance, if one believes the differences between groups of websites to be so substantial as to tolerate lower recall and the resulting larger margin of error—one could use this method and still be reasonably certain about the results.[28] This research project meets these conditions, so a recall percentage of 80% or higher is acceptable.

Applying the method: Calculating recall and increasing precision

Moving to the actual measures of recall for this project, here are the recall results of the search for documents related to the broadcast flag. The open search used these terms:

copyright site:house.gov

This retrieved 793 documents,[29] 14 of which were relevant to the broadcast flag debate. I originally tried a closed search using the following search terms:

flag copyright audio OR video OR radio OR broadcast site:house.gov

---

[28] Were this not the case, one could creatively deploy multiple searches on each website studied. By combining the results of several searches, one could presumably reduce the collective estimate of b, increasing recall to reach a targeted rate.

[29] Google reported "about 8,140" documents, but would return no more than 793. This is actually an example of the 1,001 problem. Even asking to "repeat the search with the omitted results included" still returned 793 results. In other words, one cannot even rely on getting all of the first 1,000 documents. Strangely, the same search on March 14, 2008, did retrieve 1,000 results. Again, please see Appendix C for more.

This retrieved 235 documents, 26 of which were relevant. These results did not include 4 positive results from the open search results. For this closed search term, $a = 10$, $b = 4$, and recall is $10 / 14 = .714$.

In an effort to get recall above .8, I conducted searches using new terms, ultimately settling on:

copyright (audio OR video OR radio OR broadcast) flag site:house.gov

This is logically equivalent to the original search, but the order of words matters, and this search retrieved two additional relevant documents from the open search results; unfortunately, it also dropped two of the relevant open term documents retrieved by the first search, meaning no actual improvement. Choosing to repeat the search with the omitted results included, however, did much better. Out of 900 documents,[30] the search results netted 12 of the 14 relevant documents from the open search. For this search, then, $a = 12$, $b = 2$, and recall is thus $12 / 14 = .857$. This exceeds the standard of .8, so this search term is used on all sites.

Next, here are the results of the search for documents discussing the potential reform of the DMCA. From within the same open search described above, of the 793 documents, 46 were relevant for the DMCA reform debate. My first attempt at a closed term that could retrieve an acceptably high proportion of these documents was:

copyright (dmca OR "digital millennium copyright act") (boucher OR encrypt!

OR 1201 OR hack! OR DRM OR "digital rights management") site:house.gov

---

[30] Google reported "about 1,110" documents, but would return no more than 900.

This retrieved 130 documents, with 39 of them relevant, including 38 of the 46 relevant documents retrieved by the open term. Here, $a = 38$, $b = 8$, and recall is $38 / 46 = .826$, so this search has an adequate level of recall.

Having established adequate recall for both searches, I was ready to calculate precision. This could also be done using the simple ratios of the Stryker method, calculating the ratio of relevant documents to the number retrieved. However, this would fail to leverage the power of Google to put relevant results toward the top of search results. Once one has coded every document from the closed term search results, logistic regression will describe the distribution of relevant documents within search results. This will greatly increase the researcher's ability to concentrate scarce resources on coding the portion of search results most likely to contain relevant documents.

For the broadcast flag search, search rank is an excellent predictor of a document's relevance to the broadcast flag debate. Of 900 documents retrieved by the final closed search described above, 42 were coded as relevant, including 6 duplicates. Where the first or very top result is represented as 1, and the last or very bottom is 900, the lowest-ranked (or highest-numbered) relevant document was at number 77, and 28 of the first 29 documents were relevant. This represents a very strong relationship between a document's search rank and the likelihood of its relevance. Where the X-axis represents a document's search rank and the Y-axis represents the natural log of the odds that it is relevant, the calculated β from a logistic regression is highly significant (-.085, constant = 3.591, both $p < .001$), as is the Chi-Square (265.9, $p < .001$). This model also holds a great deal of explanatory power (Cox & Snell = .256, Nagelkerke R Square = .814).

138

Calculating the predicted odds based on this model, the first result is predicted to be relevant in 97% of searches, and the odds drop to less than 1% for all results after number 96.

Logistic analysis of the DMCA search yields similar results. Of 130 results, 39 were coded as relevant; 30 of the first 45 documents were relevant, including the first ten, and the lowest-ranked relevant document was number 84. Again, the calculated β is highly significant (-.060, constant = 2.263, both $p < .001$), as is the Chi-Square (64.3, $p < .001$). Additionally, the model explains a substantial amount of the variance in document relevance (Cox & Snell = .390, Nagelkerke R Square = .554). This model predicts the first result would be relevant in 90% of searches, and the odds drop to 1% or less for all results after number 114.

This gives a good indication of how deep into the search results to plow. Unless the original calculated recall for a closed search is barely above the acceptable rate—in this case, .80—it is reasonable to accept a minor decrease in effective recall to shave off what may be hundreds of irrelevant documents per website. For the broadcast flag search, if one searches a website with this search term and examines the first 96 results, one would begin with the originally calculated recall rate of .857, multiply it by .99, and estimate an effective recall of .848. Within this set at the top of search results, we would estimate precision at 42 / 96, or .438; due to hand coding, this low rate is acceptable, but it is a far better use of human coding resources than a search with a precision of 42 / 900, or .047. For the DMCA reform search, multiplying the originally calculated recall rate of .826 by .99, one would estimate the first 114 results to have an effective recall rate of

.818. Here, precision is 39 / 114, or .342. Note that this model does not extend to sites with many more relevant documents, of which there are a few in the population; in order to remain within the limits of the model, I code only the first 42 relevant documents from each broadcast flag search and the first 39 from each DMCA reform search. I use exactly this method to collect relevant internet documents from the sites identified by the rolling web graph analysis.

Having identified a group of relevant documents to code for relevance and valence, I tested internet documents for relevance as I did with offline documents. For broadcast flag relevance for internet documents, $\alpha = .90$ (130 pairs), and for the broadcast flag, $\alpha = .93$ (84 pairs).

Content Analysis: Document Details

Before coding for a document's position in the copyright debate, each is coded for basic details. These include:

- Relevance for one of the four policy debates (specific instructions for each population; if irrelevant, stop coding)
- Year of authorship
- Category of organization represented

Relevance was described above. Because of minor differences in each medium, relevance was tested separately for each, but for basic details and valence, documents from all three media were lumped together. Unsurprisingly, it was easy to reach high agreement on year of authorship ($\alpha = .999$, 162 pairs).

The category of organization represented is, for instance, an industry sector such as "media" or "technology." These categories are deliberately large in an effort to make them very easy to apply at the expense of precision because this variable does not play a role in testing any of my hypotheses. For the same reasons, these are tested for intercoder reliability as one mutually exclusive variable; this is even though I report sector results as coded both exclusively and nonexclusively. In an effort to better describe the evolution in representation of documents, especially in documents such as newspaper articles that generally include the voices of more than one sector, such nonexclusive coding is a useful descriptive tool. As documented elsewhere (Herman & Gandy, Jr., 2006), the two copyright debate coalitions generally break down along industry lines, and tracking their participation across venues provides a rough indication of which coalition has access to which venues. For a basic descriptive account of who is speaking in which communication forums, the follow list of organizational categories were used:

1. Media (e.g., Paramount Pictures, Recording Industry Association of America)

2. Legal associations (e.g., American Intellectual Property Law Association)

3. Appointed government officials in intellectual property-related offices (e.g., the Register of Copyrights, the Commissioner of Patents and Trademarks)

4. Congresspersons

5. News (applied to print news and to online documents produced by news organizations, with the singular exception of opinion pieces by authors who fall into one of the other groups on this list)

6. Technology firms and trade groups (e.g., Dell, Microsoft, the Consumer Electronics Association, the Business Software Alliance)

7. Scholars (professors or advanced graduate students who speak from their own expertise and not on behalf of their institution, e.g. Lawrence Lessig)

8. Nonprofit groups (e.g., Public Knowledge, the Progress and Freedom Foundation)

9. Libraries (e.g. the American Library Association)

10. Education (e.g., the Association of American Universities)

11. Other (e.g., unaffiliated bloggers, other individuals with no cited organizational affiliation)

These categories are roughly in order of strongest-copyright to strongest-fair use supporters. As tested using mutually exclusive categories, $\alpha = .86$ (179 pairs). Because this variable is not central to hypothesis testing, documents were also coded nonexclusively for the sake of identifying documents that represent more than one sector.

Content Analysis: Valence

After documents are identified as relevant and coded for the basic details described above, the final step in coding is to identify the document's valence along the rhetorical continuum from strong copyright to strong fair use. In other words, this coding identifies which side of the debate a document advances. There are two sides to this debate; one side advances stronger copyright law (and thus, weaker exemptions such as fair use), and the other side argues for stronger fair use (and thus, weaker copyright law).

A coding unit is characterized as falling into one of three categories. Either it is clearly in the strong copyright camp (valence = 1), clearly in the strong copyright camp

(3), or in some sense in between these two extremes (2). This third option, neutral, may mean that a unit advances no position, and it may also mean that it makes an earnest attempt to consider both the strong fair use position and the strong copyright position. A unit is not coded neutral if it describes the arguments of one side en route to rebutting these arguments. For instance, several strong copyright advocates discuss one or more arguments of the strong fair use side and then rebut these arguments, and vice versa. In this case, the document is coded as advancing the cause of one of the two coalitions.

Coding for valence proceeds in two stages. First, each document is coded for valence. For document valence, $\alpha = .95$ (120 pairs). If a document falls into the strong copyright or strong fair use categories, coding is complete. If a document is neutral, however, then the document is coded at the paragraph level. Each paragraph is coded as relevant or irrelevant for the four debates: AHRA ($\alpha = .90$, 58 pairs), DMCA ($\alpha = .92$, 65 pairs), the broadcast flag ($\alpha = .92$, 106 pairs), or DMCA reform ($\alpha = .89$, 106 pairs). If a paragraph is relevant, it is then coded for valence ($\alpha = .90$, 139 pairs). The valence of all relevant paragraphs is then averaged to create a total valence score for the document. Thus, a document that is coded as neutral may ultimately have a score that is greater or less than 2.

These valence scores can then be analyzed as ordinal variables to characterize which side is winning the copyright debate in which media—at least as measured by whose arguments appear most frequently.

Conclusion

As described above, this research answers the research questions primarily using quantitative measures. In a broad sense, the main question is: Who speaks about digital rights management policy in which media? This research design does not seek to provide answers as to the effects, but it does proceed from the assumption that communication has some chance to affect policy outcomes. Thus, before plowing into the results, I provide some background on the policy space under study. This involves both a general background and a brief summary of the evolution of each of the four policy proposals under study. Then, I proceed to describe the results of the quantitative methodology outlined above.

CHAPTER FIVE: THE DIGITAL RIGHTS MANAGEMENT POLICY SPACE

Over the course of this study, Congress has transformed copyright law into an important vehicle for regulating the development of digital media technologies. Like the legislative process generally, this has hardly been straightforward. This chapter sets forth the twists and turns that characterized each of the policy debates considered in this dissertation. In each case, I consider the technological, economic, and political background within which each debate took place, the specifics of the policy proposals set forth, and some of the political forces that helped shape each outcome.

Audio Home Recording Act (AHRA)

In the early 1980's, electronics manufacturers began developing what would become the Digital Audio Tape (DAT) system. DAT represented an innovation that promised consumers the perfect audio fidelity of compact discs (CDs), but with the added ability to make their own recordings—keep in mind that CD recorders were not introduced until 1990. The music industry had already coined the complaint, "Home taping is killing music" (J. Sullivan, 1988, p. 51), so they were extremely concerned about DAT's promise to make perfect copies—and copies of copies. As the *New York Times* observed in 1986:

> The advent of digital audio tape—somewhat like the development of Xerox copying—will be a test of the extent to which those who push back technical frontiers can or will prevent their discoveries from being used in ways economically harmful to themselves or others. It will also test the extent to which

the music industry can impose its will on manufacturers and consumers in the face of technological advance.

            Concern over the digital tape runs high enough to prompt a bitter denunciation of the new technology by the president of the Recording Industry Association of America, Stanley Gortikov, who in a guest editorial last month for the trade journal Billboard characterized the Japanese-dominated audio hardware industry as an "assassin" bent on destruction of the largely American recording industry. "We are already losing billions to home taping," Mr. Gortikov said recently in a telephone interview. "Imagine what it will be like if the tape copy is equal to the original." (Crutchfield, 1986, p. A1)

The recording industry used a combination of congressional lobbying, threatened litigation, actual litigation against Sony (Menell & Nimmer, 2007, pp. 19-20), and market pressure to retard the importation of DAT machines. On all counts, the recording industry had a substantial effective advantage over Sony and other electronics manufacturers. They had much greater political clout in Congress (Litman, 2000), especially considering that most would-be DAT manufacturers were based in Japan. Also, record labels could refuse to release music in DAT format, reducing the consumer market for the new machines exclusively to those interested in recording music.

Starting in 1987, the recording industry attempted to get favorable legislation passed in Congress that would ban the importation or sale of DAT recorders that did not include copy control technologies (Lee, 2007, p. 452). The first such control technology that was developed actually altered the audible sound; the change likely would have been inaudible to most listeners, but it was audible for the very audiophiles who were the primary target market for the expensive new high-fidelity machines (see, e.g., Pollack, 1988). Hearings considering early attempts at such legislation met with substantial resistance from the electronics industry (*Digital Audio Tape Recorder Act of 1987*, 1987; *Digital audio tape recorders*, 1987), and the failure to produce a solid inter-industry

146

consensus around a workable technology kept these proposals from serious consideration. Tensions between the recording industry and electronics manufacturers eased to some extent when Sony consummated its purchase of CBS Records in January of 1988 (Boyer, 1988). Still, the legal threats around DAT prevented manufacturers from importing and selling the new tools.

In 1989, the electronics industry and record labels came to terms, apparently clearing the legal cloud that hovered over attempts to sell DAT in the US (Pollack, 1989). The terms of the agreement required DAT decks to include an early form of DRM that would allow consumers to make a perfect digital copy of a recording but not to make copies of copies. This system, the Serial Copy Management System (SCMS), does not alter the sonic output of sound recordings, and allowing only first-generation copies represented a compromise between the industries. In return for this limitation, record labels agreed not to pursue DAT manufacturers or their customers over in-home recording of copyrighted music.

Both industries sought quick codification of the deal, getting behind legislation (*Digital Audio Tape Recorder Act of 1990*, 1990b) and making supportive statements in hearings (*Digital Audio Tape Recorder Act of 1990*, 1990a). Unfortunately for DAT manufacturers, record companies were not the only copyright holders with the power to threaten litigation; songwriters and music publishers were not satisfied with the SCMS requirement and thus did not sign on to the agreement. As *The New York Times* reported in 1990:

> The National Music Publishers Association [NMPA], a New York group representing music copyright holders … thinks the mechanism does not restrict

147

copying enough and can be circumvented easily. The organization favors charging buyers of tape machines and blank tapes a royalty fee that would go to compensate the songwriters and music publishers. (Pollack, 1990)

The NMPA advanced litigation against Sony in the name of songwriter Sammy Cahn (*Cahn v. Sony Corp.*, 1990) in which they accused Sony of contributory infringement (Menell & Nimmer, 2007, pp. 19-20), hoping to impose legal liability on DAT manufacturers for the potentially infringing uses of their products.

The NMPA sued even though the odds were against an ultimate victory in court. In principle, the precedent set by an earlier case, in which Sony was exculpated for customers' potentially infringing uses of the company's VCRs (*Sony Corp. of America v. Universal City Studios, Inc.*, 1984), was a powerful legal tool for Sony. In the earlier *Sony* decision, the US Supreme Court had ruled that "copyright law did not impose such secondary liability where the device in question was capable of substantial noninfringing uses (and that the VCR was such a device)" (R. A. Reese, 2006, p. 197). Had the case gone to trial, the outcome likely would have been similar, but Sony opted out of another extended legal fight—likely in large part because they had already come to terms with the record labels. With Sony anxious to sell DAT decks without the threat of litigation, "the parties settled about a year into the litigation" (Menell & Nimmer, 2007, p. 20) in June of 1991 (Lee, 2007, p. 452).

In addition to implementing SCMS to limit copying, the manufacturers agreed to pay a copyright royalty on DAT decks and blank tapes; a portion of this revenue would go to songwriters and publishers. Further, they agreed to support a modified version of the proposed legislation that would force all digital audio recording devices to be manufactured under the same rules; both the SCMS requirement and the royalties would

gain the force of law (Shapiro, 1991). In return, the bill gave consumers the explicit legal

right to make noncommercial recordings for personal enjoyment, and it gave

manufacturers the legal right to help them do so. The law reads:

> No action may be brought under this title alleging infringement of copyright
> based on the manufacture, importation, or distribution of a digital audio recording
> device, a digital audio recording medium, an analog recording device, or an
> analog recording medium, or based on the noncommercial use by a consumer of
> such a device or medium for making digital musical recordings or analog musical
> recordings. (*Audio Home Recording Act*, 1992, § 1008)

This clause has come into play in many legal disputes in the years since, as the industries

and the courts have grappled with the hundreds of unforeseen technologies that are at

least capable of recording and playing digital audio (Lee, 2007).

The AHRA was outdated very quickly after it was signed. By the mid 1990s,

computer CD burners allowed music fans to engage in unlimited serial copying without

paying a cent of royalties, and the PC as home entertainment center was becoming a

reality (Stets, 1996). The AHRA was obviated by these developments, as it only regulates

special-purpose audio devices, "the digital recording function of which is designed or

marketed for the primary purpose of, and that is capable of, making a digital audio copied

recording for private use" (*Audio Home Recording Act*, 1992, § 1001(3)). The act

specifically excludes from regulation blank media that are "primarily marketed and most

commonly used by consumers … for the purpose of making copies of nonmusical literary

works, including computer programs or data bases" (§ 1001(4)(B)(ii)), and no medium

"in which one or more computer programs are fixed" (§ 1001(5)(B)(ii)) counts as a

regulated recording.

DAT decks were never widely adopted, but this bill is historically significant as the first regulation of digital rights management; in this case, the law requires it be installed into certain products. The AHRA was seen as the solution to a pointed disagreement over the direction of copyright as a vehicle for the regulation of technology.

This dispute plants the seeds of each coalition's stance in the DRM debate, and it includes a partial list of the participants in today's dispute. Copyright holders—in this case, the music industry—sought to impose a requirement that manufacturers implement a specific DRM in their product design. Electronics manufacturers—here, Sony and other DAT manufacturers—wanted only to be able to design and sell their DAT decks without being sued; even though they found the need for legislative protection objectionable, they grudgingly accepted it as a lesser evil than unending litigation. By the 101st and 102nd Congress, the industry was practically begging for Congress to pass this requirement so that they could finally import this product that had been available for years in other countries (see Chapter 6). In short, this is the first legislative battle between members of the strong copyright coalition and the still-nascent strong fair use coalition.[31]

However, the battle looked different from today's DRM debate in important ways—differences that are quantified in later chapters. Importantly, the only potential source of substantial support for the strong fair use side was the consumer electronics

---

[31] Sony is a very visible exception to the overall rule that organizations stay soundly within a coalition; as noted above, they have become copyright holders, which over time has led them to support strong copyright law. "As the mandala of history has continued to revolve, Sony itself has mutated into a motion picture studio. It therefore now finds itself allied with the likes of Disney and Universal, its erstwhile adversaries from the *Sony* [*v. Universal* Supreme Court decision]" (Menell & Nimmer, 2007, p. 20, n. 96).

industry, and they ran to Congress seeking DRM regulation as soon as it seemed likely to help them sell DAT decks. Other than policymakers, there were very few participants in the debate except those in the consumer electronics and music industries. General-purpose consumers' groups participated lightly; the National Consumers League participated in one hearing, and a quotation from Consumers Union appeared in one *Washington Post* article, but there was not yet any NGO that saw opposition to the expansion of copyright as among its core goals. Likewise, scholarly opposition to the growth of copyright was just getting started compared to today's cottage industry of multidisciplinary thinking about copyright in the digital age. Law professor Jessica Litman participated in one hearing, as did Philip Greenspun, then a research assistant at MIT. In short, the only potential mobilized opposition would have come from the technology sector, and while they were not of the strong copyright bent by nature, DAT manufacturers chose expanded copyright over endless copyright litigation. In this context, the few voices of genuine opposition to the AHRA were easily dismissed, and the bill sailed through Congress.

Digital Millennium Copyright Act (DMCA)[32]

The Digital Millennium Copyright Act, or DMCA (*Digital Millennium Copyright Act*, 1998b), is the most sweeping revision to copyright law of the last 30 years "and arguably represents the most dramatic change in the history of US copyright law" (Gillespie, 2007, p. 177). The act was an effort "to bring US copyright law 'squarely into

---

[32] Substantial portions of this section are adapted from Herman and Gandy (2006).

151

the digital age,' … [and] the primary battleground in which the [Act] achieved this goal is its first title" (Nimmer, 2000, pp. 681-682). This title (*WIPO Copyright and Performances and Phonograms Treaties Implementation Act of 1998*, 1998) was billed as an implementation of the World Intellectual Property Organization treaties, which were "signed by the US in 1996 and … by most other major industrialized countries" (Fishman, 2003, p. 13/5) shortly thereafter. There is much more to this story, which begins with fears over the internet.

The internet presented copyright holders with a substantial legal and technological challenge that could not be addressed through AHRA-style legislation. The AHRA imposed control on stand-alone media devices. When DAT was developed, few people envisioned a day when general-purpose personal computers could become the means for the reproduction and worldwide distribution of copyrighted works such as music and movies. During the 1992 presidential campaign, however, "the 'Information Superhighway' suddenly sprang into the news and became a media darling" (Litman, 2000, p. 89). Most people were still not online, but beginning in 1994, the World Wide Web exploded in popularity (Kelty, 2008, p. 223), putting a graphical browser at the heart of the online experience and quickly drawing tens of millions of new subscribers online. Over the course of the 1990's, the internet evolved from a tool primarily for academics, researchers, and technology enthusiasts into an everyday fact of life in the homes of most Americans. During this evolution, copyright holders became increasingly anxious about the internet's capacity to empower infringement. Along with sympathetic policymakers, copyright holders began devising legal remedies to this perceived threat. Policymakers

generally had little or no online experience themselves, leaving them open to a view of

the internet that portrayed it as in need of content—as hostage to long-established content

creators' willingness to post information online—even at a time when the content online

was growing exponentially without participation from major media companies (Litman,

2000, pp. 93-94).

The policy actor who gets the most credit for the push to address this threat via

something like Title I of the DMCA is Bruce Lehman, who was Patent Commissioner

from 1993 to 1998. In 1993, the White House

> appointed an 'Information Infrastructure Task Force' to formulate government
> policy related to the [internet]. … Content issues were delegated to the
> Information Policy Committee, which appointed a Working Group on Intellectual
> Property chaired by Patent Commissioner Bruce Lehman. Lehman had
> represented the computer software industry on copyright issues before his
> appointment to the Patent Office; his senior staff included former copyright
> lobbyists for the computer and music recording industries. They maintained
> extensive informal communication with private-sector copyright lobbyists as they
> geared up to formulate administration copyright policy. (Litman, 2000, p. 90)

In a Green Paper draft report (Information Infrastructure Task Force: Working Group on

Intellectual Property Rights & Lehman, 1994), and again in the White Paper final report

(Information Infrastructure Task Force: Working Group on Intellectual Property Rights &

Lehman, 1995), the group expressed the view that copyright holders should develop and

deploy DRM systems to retard infringement. Because DRM can be circumvented, "the

Working Group concluded that the law should be amended to prohibit any circumvention

of anticopying systems, and forbid the creation or sale of any device or service intended

to defeat such systems" (Litman, 2000, p. 93). This idea became the heart of Title I of the

DMCA: Prohibit the circumvention of DRM, and prohibit the development and

distribution of the tools of circumvention.

153

The Working Group reports were part of Lehman's broader push for federal legislation to turn this idea into law. He was also pushing for a treaty that would implement the idea internationally. This was before there was a well organized and identifiable strong fair use coalition, but the Working Group reports caused "dismay among libraries, composers, writers, online service providers, … and the makers of consumer electronic devices and computer hardware" (Litman, 2000, p. 93). Several of these sectors—particularly libraries and the hardware side of the technology sector—have become vital parts of the strong fair use coalition. Several law professors were also opposed to the proposals made in the White Paper. Immediately following its release, "Peter Jaszi, a law professor at American University in Washington, held informal consultations with like-thinking law professors and representatives of library organizations to see whether there was any possibility of mounting an effective opposition to the White Paper's proposals" (Litman, p. 123).[33]

Jaszi recruited other White Paper opponents, including "library organizations, online service providers, telephone companies, computer hardware and software manufacturers, consumer electronics companies, and civil rights and consumer protection organizations" (Litman, p. 123). This group of interests agreed to work together, calling themselves the Digital Future Coalition, or DFC (Digital Future Coalition, n.d.). The DFC succeeded in mobilizing substantial—and, from the standpoint of Lehman and the content industries, unexpected—opposition to the proposed legislation to implement the

---

[33] The other law professors at the earliest meetings included Jessica Litman, Pamela Samuelson, James Boyle, Lolly Gassaway, Bob Oakley, Julie Cohen, and David Post (Litman, 2000, p. 145).

154

White Paper recommendations (*NII Copyright Protection Act of 1995*, 1995a; *NII Copyright Protection Act of 1995*, 1995b).

Both the House and Senate versions contained a categorical ban on the importation, development, and distribution of any tool to circumvent DRM (*NII Copyright Protection Act of 1995*, 1995a, § 1201; *NII Copyright Protection Act of 1995*, 1995b, § 1201). The bills also contained bans on the removal or alteration of copyright management information—data, whether plainly obvious or carefully hidden (e.g., digital watermarks), that identifies the copyright holder and related information (§ 1202). They also contained provisions stipulating civil penalties, giving a victorious plaintiff the choice of actual damages or statutory damages of up to $2,500 per violation of the ban on trafficking in tools that circumvent DRM (the ban contained in section 1201) or up to $25,000 per violation of the section 1202 ban on removal or alteration of copyright management information (§ 1203). Finally, the bill stipulated criminal penalties of up to $500,000 in fines or 5 years in prison for anybody convicted of violating "section 1202 with intent to defraud" (§ 1204).

Most of the groups represented by the DFC saw this bill as a looming legal liability that could threaten what they considered to be their regular course of business. Thus, opposition mobilized and prevented the bills' easy passage through Congress. This development surprised Lehman, who was so confident of domestic passage that he had already begun pushing for an international treaty with similar provisions via the appropriate United Nations agency, the World Intellectual Property Organization (WIPO). Litman (2000) explains how he used this mistake to his advantage:

> The domestic legislation, however, was not moving. The commissioner, therefore, decided to attack the problem the other way around. He focused his attention on getting his agenda adopted by the WIPO member nations, reasoning that when the United States signed the treaty, Congress would be obliged to adopt implementing legislation in accord with the White Paper's recommendations. (p. 129)

The language of the final treaty that was signed in 1996, the WIPO Copyright Treaty (*WIPO Copyright Treaty*, 1996), was much weaker than Lehman had hoped (Litman, 2000, p. 131), requiring only that signatory countries "shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures" (*WIPO Copyright Treaty*, 1996, Article 11). Negotiators from other countries were far less interested in the sharp expansion of copyright Lehman envisioned, and the compromise language reflected this hesitation to move as sharply in the direction of DRM regulation.

Two important differences between the Treaty language and the domestic bills warrant discussion. First, the Treaty requires the regulation of the act of circumvention, whereas the bills regulate tools and services that aid in circumvention. Thus, the Treaty is much closer to the traditional contours of copyright, which regulates behavior but generally holds technology manufacturers blameless, providing their tools also have substantial noninfringing uses (Menell & Nimmer, 2007; *Sony Corp. of America v. Universal City Studios, Inc.*, 1984). Thus, while Lehman had hoped for treaty language that regulated technology, resistance from other delegates led to a compromise that was a much less radical divergence from then-current copyright law.

Second, the Treaty imposes a rather low standard for domestic legislation, that it "provide adequate legal protection and effective legal remedies" against circumvention. Existing US law arguably met the Treaty's standard. It was already illegal to circumvent

DRM if conducted as part of a copyright infringement, and manufacturers of "black box" devices that only served to circumvent DRM had already been subjected to legal liability for facilitating infringement (Litman, 2000, p. 131). The "Clinton Administration initially considered whether the WIPO Copyright Treaty might even be sent to the Senate for ratification 'clean' of implementing legislation" (Samuelson, 1999, p. 13). Herman and Gandy (2006) describe how strong copyright advocates—including congressional allies—made a more sophisticated use of the treaty:

> Instead [of seeking ratification without implementing legislation], Congress used the Treaty as an excuse to implement a much more sweeping ban on circumvention. In short, Lehman and the bill's congressional supporters used WIPO to launder their own interests, running their political capital through the bank of international credibility and treating the final bill as something required by international law. (p. 131)

The Treaty became a rhetorical device by which the strong copyright coalition could set aside many of the normative questions that normally arise during the legislative process. Despite the coalition's disappointment with the relative weakness of the final treaty, they urged passage of much stronger legislation in the name of compliance with our new international treaty obligations. In hearings in the 105th Congress in 1997 and 1998, at least 10 witnesses made this argument, a clear case of "policy laundering" (B. D. Herman & Gandy Jr., 2006, p. 133).

The bans in the bills proposed in the 105th Congress and in the final legislation are even more expansive. Section 1201 implements three different bans. The first ban (or the "basic ban") prohibits circumventing DRM to gain unauthorized access to copyrighted works. As passed into law, it reads, "No person shall circumvent a technological measure that effectively controls access to a work protected under this title" (*Digital Millennium*

Copyright Act, 1998b, § 1201(a)(1)(A)).[34] For example, it is a violation of this provision to defeat a software installer's requirement for a unique serial number. While doing so for the purpose of infringing copyright was already illegal, this clause bans it for nearly any reason—even if one has misplaced the serial number for a legally purchased software package and intends to install it on just one computer. Even a librarian who wants only to preserve a decaying digital artifact is prohibiting from circumventing an access-controlling DRM to do so.

The second ban prohibits manufacturing, importing, and trafficking in tools that would assist one in the kind of circumvention covered by the basic ban (*Digital Millennium Copyright Act*, 1998b, § 1201(a)(2)). This ban (the "access trafficking ban") prohibits computer-repair services from assisting a librarian in the above scenario, and it prohibits librarians from developing a technology to facilitate circumventions. The third ban (the "additional violations ban") prohibits trafficking in tools to facilitate the circumvention of DRM if that DRM protects any copyright holder's rights, notably the exclusive right of reproduction (*Digital Millennium Copyright Act*, 1998b, § 1201(b)(1)). An example of a technology that would violate this ban is a program that defeats a DRM system that prevents the copying of some CDs; again, not even librarians can develop such a technology.

The basic ban, which prohibits the circumvention of access control technologies, is the major difference between section 1201 in the bills as introduced in the 104th

---

[34] The original House and Senate bills contained one minor difference, banning the circumvention of a "technological protection measure" (*Digital Millennium Copyright Act*, 1997; *Digital Millennium Copyright Act*, 1998a).

Congress (which had no such ban) and the 105th Congress. Whereas in the earlier bills, section 1201 only prohibited trafficking in circumvention devices, the later bills added a ban on the circumvention of access control technologies.

In the House bill as introduced (*Digital Millennium Copyright Act*, 1997), a very brief section 1201 contained the three bans with no explicit exceptions. It contained the following caveat, which is also included in the final legislation: "Nothing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title" (*Digital Millennium Copyright Act*, 1997, § 1201(d); *Digital Millennium Copyright Act*, 1998a, § 1201(c)). The section does not increase the scope of what constitutes infringement, so this clause does not necessarily mean that fair use is an affirmative defense against charges that one has illegally circumvented DRM or trafficked in illegal circumvention devices. Otherwise, the bill contained no other exemptions or caveats.

In the face of mobilized opposition from a number of sectors, the bill's supporters made several narrow concessions, each creating a limited reprieve from one or more of the three bans. The final legislation includes exemptions from one, two, or all three bans for nonprofit libraries (*Digital Millennium Copyright Act*, 1998b, § 1201(d)), law enforcement (§ 1201(e)), reverse engineering computer programs to achieve interoperability (§ 1201(f)), "good faith encryption research" (§ 1201(g)(2)), and protecting personally identifying information (§ 1201(i)). The legislation contains an exemption from the access trafficking ban for devices with the sole purpose of preventing "the access of minors to material on the Internet" (§ 1201(h)(2)), and one is exempted

from the basic ban and the access trafficking ban in order to engage in security testing of one's own computer (§ 1201(j)).

Nearly every one of these caveats addresses specific concerns by one or more concerned opposition sectors. Librarians spoke up and got a very limited exception; they are only allowed to circumvent DRM "to make a good faith determination of whether to acquire a copy of that work" (*Digital Millennium Copyright Act*, 1998b, § 1201(d)(1)). Software designers and encryption researchers spoke in opposition and got some more substantial breathing room to do their jobs. Privacy advocates testified in opposition, so section 1201(g) gives them the right to circumvent DRM in order to protect their personal information. The list of exemptions reads like an honor roll of opposition groups, each of which got a concession in rough proportion to their political capital. Rather than opening the legislation to a general-purpose exemption for otherwise noninfringing uses—and technologies that are capable of substantial noninfringing uses—the bill's backers chose this patchwork effort to address the specific concerns of the most vocal opposing constituencies, giving would-be DRM hackers fewer legal defenses.

The basic ban is also subject to one general-purpose exemption of sorts. Every three years, under the supervision of the Librarian of Congress, the US Copyright Office solicits written proposals for classes of works that would be exempted from the ban on circumventing access-controlling DRM systems (*Digital Millennium Copyright Act*, 1998b, §1201(a)(1)). Exemptions are determined based on these proposals, as well as written reply comments and in-person testimony. As Herman and Gandy (2006) note, the procedure moves questions of fair use away from relatively fair use-friendly federal

courts and into the hands of the Register of Copyrights, a clear member of the strong

copyright coalition—a shift of venue that substantially favors the strong copyright

coalition (pp. 143-144). They further argue that, though several exemptions have been

granted in each rulemaking, the statute and the Register of Copyright's interpretation of

the rules for determining exemptions are heavily stacked against proposed exemptions

(pp. 187-190), though in the Register's defense, the 2006 rulemaking was somewhat less

objectionable in terms of both procedure and outcome (B. D. Herman, 2006a).[35]

---

[35] Herman and Gandy (2006) were critical of the decision to make exemptions that made
no reference to specific uses and users. For instance, in 2003, the Register recommended
an exemption for electronic books in which text-to-speech functionality had been turned
off. The exemption was targeted at supporting the efforts of those who sought to make
audiobooks for the visually impaired, but the class of works was strictly tethered to traits
that were inherent in the works themselves. Thus, any noninfringing use of these works
justified circumvention for the 3 year period of the exemption, even though the
exemption was granted with just one particular use in mind.

This unnecessary reading of the statute—that exemptions have to be tethered to
qualities in the works independent of types of uses or users who are seeking an
exemption—limited the likelihood of recognizing exemptions that could not be limited in
reach by being tethered to specific kinds of works. For instance, Herman and Gandy
(2006) argue, the socially valuable work of media critics was hampered by the Register's
2003 refusal to recognize a limited right to circumvent CSS, the encryption on DVDs (pp.
171-174, 187-188), a decision that was rendered at least in part because the Register saw
no way to limit the population of those allowed to circumvent CSS.

In the 2006 ruling, in contrast, exemptions were also tethered to specific uses and
users in order to limit the reach of exemptions granted to would-be circumventers of
DRM systems that are of substantial commercial significance. This resulted in
exemptions for classes of works that would have been incredibly unlikely under the
former interpretation, including especially a class that permits the circumvention of CSS.
Herman (2006a) explains:

Two of the [2006] exemptions are particularly noteworthy. They are:

1.   Audiovisual works included in the educational library of a college or
        university's film or media studies department, when circumvention is
        accomplished for the purpose of making compilations of portions of those
        works for educational use in the classroom by media studies or film
        professors.

Ironically, the bill's opponents may have been better off had they allowed the

original bill to pass without the explicit exemptions that were later added. Litman (2000)

explains:

> The original Lehman bill granted copyright owners sweeping new rights, but its silence on available exceptions invited the courts to apply copyright's traditional limitations [such as fair use]. The DMCA [as passed into law] also grants copyright owners sweeping new rights. Its laundry list of narrow exceptions, however, discourages the inference that the classic general exceptions and privileges apply. (p. 145)

This inapplicability of general exceptions has become cause for much political wrangling

in recent years.

---

6.   Sound recordings, and audiovisual works associated with those sound recordings, distributed in compact disc format and protected by technological protection measures that control access to lawfully purchased works and create or exploit security flaws or vulnerabilities that compromise the security of personal computers, when circumvention is accomplished solely for the purpose of good faith testing, investigating, or correcting such security flaws or vulnerabilities.

This represents a substantial shift in the Copyright Office's interpretation of Section 1201. In 2000 and 2003, they vocally rejected any and all classes of works that were defined, even in part, by reference to specific users or intended uses. Classes of works had to be defined strictly in terms of the qualities of the works themselves.

In 2006, however, the intended use and/or user are part of 4 of the 6 granted exemptions. Of particular note, read pp. 12-24 of the Register's recommendations detailing the reasoning behind the film & media studies professors' exemption.

Under the rules as constructed in 2000 and 2003 and applied by the Copyright Office, exemptions covering DRM-bearing CDs and DVDs would have been almost impossible. The change in rules, and these exemptions, represent at least a modest improvement— albeit one that hardly obviates the substantial critique of the rulemaking procedure and the anticircumvention provisions.

At the time of this writing, the Register's 2006 recommendations (Peters, 2006) are available at: http://www.copyright.gov/1201/docs/1201_recommendation.pdf. The documents associated with all of the triennial rulemakings are at: http://www.copyright.gov/1201.

Compared with the AHRA, the passage of the DMCA represented a much more significant shift in copyright law as a vehicle for the regulation of technology. The AHRA regulates only one small class of technologies—stand-alone digital audio recording devices—in which the bulk of innovation may have already come to pass. In contrast, the DMCA regulates a potentially infinite number of devices—including those that exist now and countless more that are yet to be developed—that are touched by DRM systems. Every copyrighted work that can be digitized can be wrapped in encryption and flagged by copyright management information. Doing either or both gives copyright holders rights of exclusion substantially exceeding those available under traditional copyright law, so those whose business models depend upon control enforced by copyright[36] are generally quite happy to apply DRM to everything they can. Those who design or experiment with technologies to handle such copyrighted works are therefore on fairly thin legal ice unless they work within the limits set by copyright holders. This discourages even legitimate academic encryption research (McLeod, 2001, pp. 261-262) despite the exception in section 1201(g). Further, those who hope to make unauthorized but legal uses of DRM-protected works are also discouraged from doing so, both by the illegality of their behavior and low availability of the tools of circumvention (Herman & Gandy Jr., 2006, p. 132). This represents a substantial departure from prior copyright law that "shares neither the logic nor the strategy of copyright" (Gillespie,

---

[36] Not everyone who creates and disseminates copyrightable works depends on enforcing this control. For instance, Benkler (2006) provides an insightful topology of reasons that one might produce and disseminate original creative works, most of which do not require—and many of which would discourage—the strong enforcement of one's copyrights (pp. 41-48).

2007, p. 177). Opposition to the DMCA as enacted has become one of the central issues

driving the growth of the strong fair use coalition in the decade since its passage, and the

strong copyright coalition has been ready for the battle.

DMCA Reform

Following the DMCA's passage, Title I quickly earned the ire of people from

many sectors—primarily the same kinds of groups that mobilized against the White

Paper, though with increasing numbers and organization. Several incidents have

accelerated this mobilization. In 2001, a team of computer scientists at Princeton faced

legal threats for their research into a DRM system then in development. The Secure

Digital Music Initiative (SDMI), a coalition of recording industry and technology firms,

was developing the DRM system, and SDMI issued the legal threats. Lessig (2004) tells

the story:

> Using encryption, SDMI hoped to develop a standard that would allow the
> content owner to say "this music cannot be copied," and have a computer respect
> that command. The technology was to be part of a "trusted system" of control that
> would get content owners to trust the system of the Internet much more.
>   When SDMI thought it was close to a standard, it set up a competition. In
> exchange for providing contestants with the code to an SDMI-encrypted bit of
> content, contestants were to try to crack it and, if they did, report the problems to
> the consortium.
>   [Princeton Professor Ed] Felten and his team figured out the encryption
> system quickly. He and the team saw the weakness of this system as a type: Many
> encryption systems would suffer the same weakness, and Felten and his team
> thought it worthwhile to point this out to those who study encryption. …
>   And though an academic paper describing the weakness in a system of
> encryption should also be perfectly legal, Felten received a letter from an RIAA
> lawyer that read:
>   "Any disclosure of information gained from participating in the Public
> Challenge would be outside the scope of activities permitted by the Agreement
> and could subject you and your research team to actions under the Digital
> Millennium Copyright Act." (pp. 155-157)

Felten's team was threatened with legal liability for attempting to share the results of their research at an academic conference; this story drew substantial negative publicity for the DMCA. The image of an Ivy League computer scientist facing legal liability for conducting research is offensive to most in a country such as the US with a strong tradition of free speech. The negative publicity led the SDMI attorneys to drop the case, though not before the ordeal wreaked substantial professional havoc in the lives of the researchers (Felten, 2006). Supported by the Electronic Frontier Foundation, Felten filed suit seeking to have his research declared legal, but since the recording industry had dropped the threat to sue, the New Jersey Federal District court dismissed the case, and Felten declined to pursue an appeal (Electronic Frontier Foundation, 2002).

Also in 2001, Russian programmer Dmitry Sklyarov endured an even more substantial legal ordeal. During a visit to the US, he was arrested and jailed for nearly a month, charged with criminal violations of the DMCA. Sklyarov was a PhD student researching cryptography and an employee of Russian software firm Elcomsoft. He helped create a program called the Advanced eBook Processor (AEBPR), which removed the restrictions embedded into Adobe Systems' eBook software. When he gave a presentation about the software at the 2001 Defcon hackers'[37] convention in Las Vegas,

---

[37] As portrayed in mainstream media, "hackers" are often people who use their technological skills to commit crimes such as stealing personal financial information or breaking into carefully protected commercial or government databases. As used here, a "hacker" is better thought of as a tinkerer—somebody who enjoys the challenge of taking things apart, seeing how they work, and perhaps modifying them. Their political core values are best captured in the name of Felten's blog, Freedom to Tinker (http://freedom-to-tinker.com). In this sense—and for most people who identify as hackers—the primary goals are satisfying an inherent curiosity and improving things, including security.

FBI agents arrested him and charged him with trafficking in a circumvention device for profit, a criminal offense as defined in section 1204. After spending several weeks in jail, Sklyarov was released on the condition that he testify against Elcomsoft, and in 2002, the jury found the company not guilty on the grounds that they believed the company's defense of not knowingly violating the law (Richtel, 2002).

Many other cases have been used to highlight criticisms of the DMCA, but the Felten and Sklyarov cases instantly came to symbolize opposition to the law. More importantly from a political mobilization standpoint, the Felten and Sklyarov cases quickly grabbed the attention of those whose livelihoods depend on research, such as academics, programmers, and inventors. These groups were already natural allies in the fight to reform the DMCA, but these stories added thousands to the ranks of the newly or potentially mobilized.

The years immediately following the DMCA's passage also saw an explosion in the number of nonprofit groups investing substantial time and energy in building and growing the strong fair use coalition. In addition to the continued involvement of the Digital Future Coalition, 2001 saw the founding of DC-based Public Knowledge, a group that quickly came to be known for its dedication to advocating for copyright law that better fits consumers' interests. This period also saw substantial investment in the copyright debate by pre-existing technology groups such as the Center for Democracy and Technology (CDT) and the Electronic Frontier Foundation (EFF). Consumers Union

---

Technology enthusiasts object to the DMCA because it criminalizes what they see as socially valuable tinkering.

also took a substantial interest and, in addition to several of their own submissions,

coordinated with Public Knowledge and other groups on a number of submissions to the

congressional hearing record. The hearings on DMCA reform are peppered with this

substantial variety of NGO voices, giving them quite a different tenor than those leading

up to the DMCA's passage. This period features several voices from the NGO sector[38]

calling for DMCA reform.

Would-be DMCA reformers also had a handful of allies in Congress, led by

Representative Rick Boucher, Democrat of Virginia. In the 108th and 109th Congresses,

he introduced bills (H.R. 107 in 2003-2004 and H.R. 1201 in 2005-2006) that would

reduce the impact of the anticircumvention provisions of the DMCA (*Digital Media*

Consumers' Rights Act of 2003, 2003; *Digital Media Consumers' Rights Act of 2005*,

2005). Also in the 108th Congress, Representative Zoe Lofgren, Democrat of California,

introduced a similar DMCA reform bill, cosponsored by Boucher (*Benefit Authors*

without Limiting Advancement or Net Consumer Expectations (BALANCE) Act of 2003,

2003). These bills would have nullified the basic ban on circumventing copy controls as

applied to otherwise legal activities. They also would have substantially reduced the

reach of the anti-trafficking provisions, allowing companies to develop and sell tools with

---

[38] This period also saw the Progress and Freedom Foundation (PFF) take a more substantial interest in copyright policy. As far as this author knows, the PFF is the only NGO in the DRM policy debate that is in the strong copyright coalition. Because the strong copyright coalition had many members, strongly shared policy core beliefs, and strong coordination since well before the DMCA's passage, the PFF plays a far less substantial roll than, say, Public Knowledge. Further, PFF plays a far less active roll in Congress. In the 2003 to 2006 period, PFF contributed just one written submission, while Public Knowledge authored or co-authored a total of 12 speeches and written submissions in six different hearings.

substantial noninfringing uses. For instance, here is the provision in the 2005 bill that

would have added the following limitations on the basic ban and trafficking bans:

> Fair Use Restoration- Section 1201 (c) of title 17, United States Code, is
> amended--
> (1) in paragraph (1), by inserting before the period at the end the following: `and
> it is not a violation of this section to circumvent a technological measure in order
> to obtain access to the work for purposes of making noninfringing use of the
> work'; and
> (2) by adding at the end the following new paragraph:
> `(5) Except in instances of direct infringement, it shall not be a violation of the
> Copyright Act to manufacture or distribute a hardware or software product
> capable of substantial noninfringing uses.'. (*Digital Media Consumers' Rights Act
> of 2005*, 2005, § 5(b))

The bill would have tethered charges of illegal circumvention to charges of infringement,

and it would have applied the *Sony* standard (*Sony Corp. of America v. Universal City

Studios, Inc.*, 1984) to the development and distribution of tools capable of

circumvention—namely, that they also be capable of substantial noninfringing uses.

The effect of the DMCA would be quite different were such a reform

implemented. The law would still forbid hacking DVDs en route to selling bootlegged

copies on the subway; because a would-be bootlegger would have circumvented the

encryption, he would be subject to the DMCA's civil and criminal penalties *in addition to*

the civil and criminal penalties that apply to the acts of infringement. As reformed,

however, the DMCA would no longer apply to the would-be circumventer who seeks to

transfer a copy of a legally purchased DVD onto her laptop.

Those who sought to develop and sell the tools for circumvention also might have

faced little legal threat for doing so after such a reform, as long as these tools were

capable of substantial noninfringing uses. Since DRM systems often prevent otherwise

noninfringing uses such as fair use (Netanel, 2008, pp. 74-75), most circumvention tools

are likely capable of substantial noninfringing uses. This would have been quite a legal shield for would-be makers of circumvention devices, putting the tools to circumvent most major kinds of DRM into the hands of those with little technological sophistication.

If Boucher's bill had become law, the DMCA would also be somewhat less frightening for the likes of Ed Felten and Dmitry Sklyarov. Copyright holders would be somewhat less likely to threaten encryption researchers and even less likely to prevail in court, though even a substantial reduction in the odds of a lawsuit or an unfavorable verdict might not provide encryption researchers with much comfort. Even under the current law, which contains an exemption for encryption research (*Digital Millennium Copyright Act*, 1998b, § 1201(g)), Felten likely would have prevailed, but SDMI sued anyway. Since the court's dismissal of Felten's suit, no researcher has faced similar legal threats for academic encryption research. The reform may have possibly increased any such researcher's odds of success in court, and it would have greatly increased the odds of a successful outcome for somebody in Sklyarov's situation—researching encryption for academic purposes and turning this knowledge into a marketable product. Yet this change in odds may provide little comfort when one's own career and financial well-being are on the line. As Felten (2003) explains, "For me and my colleagues, probably wasn't enough. Even a 99% chance of getting to keep our houses and savings wasn't enough. Nor should it be" (Felten, 2003).

While Boucher's bills may not have delivered everything for which the strong fair use coalition could have hoped, they garnered serious attention and support from virtually every significant member of the coalition. The policy documents examined in more detail

later in this research are filled with such support from a diverse array of groups that got behind Boucher's attempts to reform the DMCA. Multiple witnesses, in at least five separate congressional hearings, seriously discussed the bills. Further, in the 109th Congress, the bill's 13 bipartisan cosponsors included House Committee on Energy and Commerce Chair Joe Barton, whose co-sponsorship gave immediate political capital to H.R. 1201.

Copyright law as a vehicle for regulating technology is under the natural jurisdiction of no one congressional committee; Judiciary is the proper place for considering copyright law, but Commerce is the natural home for issues pertaining to the regulation of electronics. This opens the door to substantial venue shopping for both sides. In general, the Judiciary committees in both the House and Senate have been quite friendly venues for the strong copyright coalition on the issue of DRM. In contrast, the Commerce committees have proven more skeptical of DRM regulation, especially in the House. Barton's chairmanship sharpened this divide; he gave Boucher three hearings that were partially or completely dedicated to airing the DMCA reform proposal.

Despite the substantial push, however, the bill died in committee due to highly mobilized opposition from the strong copyright coalition. The motion picture industry, recording industry, and a substantial portion of the technology sector expressed strident opposition to the idea of weakening the DMCA bans on circumvention and trafficking in circumvention technologies. Their allies in Congress joined in, in enough numbers to stop the passage of any reform. On this count, the strong copyright coalition substantially limited Barton's influence by isolating discussion of the bills to his committee. All three

other policy proposals discussed in this research were the explicit subject of hearings in both Commerce and Judiciary hearings; in contrast, neither Judiciary committee held a hearing on Boucher's bills. While it might have had the votes in Barton's committee, it never came to a vote. The attention given to Boucher's proposals, however, marked a watershed moment in the history of the DRM policy debate. For the first time, the strong fair use coalition was on the offensive and gaining substantial traction in their effort to roll back the regulation of DRM. The effort may have stalled, but it shows how seriously the coalition had grown between 1998 and 2003.

Broadcast Flag

The 2000's saw a number of proposals for increasing the reach of DRM regulation. One, the Consumer Broadband and Digital Television Promotion Act (*Consumer Broadband and Digital Television Promotion Act*, 2002), "would have prohibited the manufacture, sale, import, or provision of any 'interactive digital media device' that didn't incorporate certain security technologies" (Gillespie, 2007, p. 196). This bill, also known as the "Hollings Bill" in honor of its sponsor, Senator Fritz Hollings (D-SC), drew howling opposition from the newly well-organized strong fair use coalition. Another, the 2005 Digital Transition Content Security Act (*Digital Transition Content Security Act*, 2005), would have prevented the re-digitization of analog content (Gillespie, p. 197).

While these and other proposals have merited a reasonable amount of attention, the most noteworthy proposal was the effort to impose a DRM system called the "broadcast flag" on devices capable of receiving broadcast digital television (DTV)

signals. This proposal was the result of sophisticated negotiations between multiple industries and other stakeholders. After a rulemaking, the FCC passed a broadcast flag mandate, only to be rebuffed as having exceeded their jurisdiction as granted by Congress. The decision left open the chance that Congress might give them that jurisdiction, and as discussed in the next chapter, this proposal made some headway in Congress. The proposal for a television flag even had enough momentum that a far less developed proposal for a similar flag mandate to be imposed on radio receivers was advanced in an attempt to piggyback upon the TV flag proposal.

Consumers have long been able to record broadcast radio and television signals, first with analog tape, and now with sophisticated digital recording devices. For this entire period, this ability has been a source of genuine anxiety among the content industries. Even though copyright holders have tried to sue manufacturers for enabling home recording, the *Sony* decision (*Sony Corp. of America v. Universal City Studios, Inc.*, 1984) recognized home taping as at least potentially noninfringing; in the case, the court recognized a right to record television programs and watch them later. This leaves content owners with few options for imposing DRM on broadcast media content—without additional government intervention, that is.

Section 1201 of the DMCA is of little use for copyright holders seeking to restrain home taping of broadcast media; the statute only applies in situations in which media creators can apply DRM before media products are distributed and expect that only authorized uses will be enabled. Broadcast content does not meet these requirements. The standards for broadcasting analog signals have been in place for decades (Barnouw,

172

1990), well before ordinary consumers were capable of substantial reproduction and distribution of media content. Under these standards, analog signals are sent, unencrypted, to unsophisticated receivers that do little more than reproduce the sights and sounds carried on a given frequency. This presents a substantial obstacle to a DRM scheme in the form of political and economic inertia; it would not be possible to impose any DRM scheme on broadcasting without completely revamping the standards for broadcast receivers and thus obviating all broadcast technology deployed under the old standards.

The transition to digital broadcasting increased the level of copyright holders' anxiety over home recording; digital recordings of digital broadcasts are generally of superior quality and more easily shared than analog recordings. Yet this transition also offered a unique opportunity to push for stronger technical and legal control over content, potentially squelching an activity that was already of concern in the analog era. Motion picture studios in particular seized on this opportunity, hoping to recreate the success of the relatively sealed environment offered by DVD distribution. Their best weapon was the threat to withhold content; they argued that, unless they could be reasonably assured that their content would be protected, they would withhold their high-value content from broadcasting, sabotaging the transition to digital television broadcasting (Gillespie, 2007, p. 200).

With an eye toward working DRM standards into the new set of standards to be deployed as part of the transition to DTV, the studios began building an inter-industry coalition to develop a mutually acceptable, well-developed technical solution that could

be proposed as the basis for a government mandate. There was no political will for

encrypting content at the source—as is the case with DVDs, for instance—so the next

best choice was to force a mandate that all tuners encrypt content before passing it along

to other media devices. In his very detailed review of the broadcast flag process, Gillespie

(2007) explains the basic premise of the proposed broadcast flag DRM system as follows:

> Digital broadcasts would be accompanied by a mark that indicated whether the
> owner of that content would permit it to be redistributed or not. Any digital tuner
> that transformed this signal into a displayable form would be required to check for
> and honor this flag. If the content was flagged, the tuner would allow it to be
> recorded only in specified formats—formats that would preserve the broadcast
> flag if that copy were passed to another device … after encrypting it using one of
> a limited set of authorized encryption technologies. (p. 202)

In this way, only authorized forms of reuse would be allowed, but broadcast content

would still be sent unencrypted—preserving at least the rhetoric that broadcasting is a

public service. This proposal represented a profound reshaping of consumers' ability to

record and reuse broadcast media, and it nearly became the law of the land.

In 2001, Fox Broadcasting Company proposed the DTV broadcast flag as a

technical standard and began building an inter-industry coalition to put the force of law

behind the standard. This push led to the formation of the Broadcast Protection

Discussion Group (BPDG), including representatives from the major motion picture

companies, as well as "consumer electronics corporations, … information technology and

software companies, … companies specializing in existing forms of copy protection, …

and consumer and public advocate groups" (Gillespie, 2007, p. 203). Despite initial,

vocal objections by some participants—in particular, NGOs such as the Electronic

Frontier Foundation—the process of developing the DTV flag standard was a reasonably

smooth process within the BPDG. "The premise of the flag and how it would work was

already agreed upon at the start, or agreed upon by enough of the major players that critics could be pushed aside" (Gillespie, p. 204). Many groups that actually opposed the flag mandate continued to participate in order that they might play some role in steering the process. All parties behaved as if doubts about the need for a mandate would not be adequate in force to stop its passage.

While important differences remained on the question of how new encryption schemes would be approved, the BPDG was able to present the DTV flag proposal to the FCC as the result of nearly unanimous inter-industry agreement. With nearly everybody seemingly on board, the FCC quickly initiated proceedings to consider a proposed rule implementing the broadcast flag as a required standard for DTV receivers. In August 2002, the Commission began a proceeding to consider the proposal. By November 2003,

> the FCC ordered that broadcast flag regulations be imposed on all digital
> television devices. … The FCC did, however, issue a "further notice of proposed
> rulemaking" in which it extended some of the questions about its proper
> application, particularly about the way the "downstream" DRM encryption
> technologies would be authorized. (Gillespie, 2007, p. 211)

The rule was to take effect July 1, 2005.

In 2004, a coalition of four NGOs and five library groups filed suit to stop the broadcast flag rule from taking effect. Among NGOs, Public Knowledge led the charge, joined by the EFF, Consumers Union, and Consumers Federation of America. Library groups included the American Library Association (ALA), Association of Research Libraries. American Association of Law Libraries, Medical Library Association, and

Special Libraries Association. In May 2005, the DC Circuit Court ruled on behalf of the petitioners, holding that the FCC had exceeded its jurisdiction under current law (*American Library Association et al. v. Federal Communications Commission and United States of America*, 2005). The FCC has the right to regulate receivers, but the 3-judge panel unanimously held that current statutory authorization does not grant the FCC the "authority to regulate receiver apparatuses after the completion of broadcast transmissions" (Lee, 2007, p. 411). This last-minute intervention prevented the broadcast flag standard from ever becoming law regulating broadcast receiving equipment—just two months before the regulation would have gone into effect.

The court ruling left open the possibility for congressional intervention; if the FCC needed authorization from Congress before acting, Congress could provide that authorization. In May 2006, then-Senator Ted Stevens, Republican from Alaska, included Senate Bill 2686, "a comprehensive telecommunications reform bill covering a wide range of telecommunications issues, including the issues of the digital television transition, Internet and universal service fund, video competition, and community broadband deployment as well as digital content protection" (Lee, 2007, p. 406, n. 5). One section of the bill authorized the FCC to adopt a broadcast flag mandate, specifically invoking the Commission's 2003 ruling (*Communications, Consumer's Choice, and Broadband Deployment Act of 2006*, 2006, § 452); this was part of the subtitle known as the Digital Content Protection Act of 2006 (§§ 451-454). The bill was the subject of congressional hearings and a relatively high volume of attention, but the broadcast flag

was only part of the cacophony of debate over the bill, which passed committee but never came up for a final vote in the Senate.[39]

Stevens' effort stalled in large part due to public demands that network neutrality be part of any comprehensive telecommunications reform act (Reilly, 2006).[40] These caveats aside, the strong fair use coalition expressed serious opposition to the flag mandate in congressional debate. In particular, the same groups that won the legal battle to stop the FCC ruling—NGOs such as Public Knowledge and the EFF, and library groups such as the ALA—came out in full force against broadcast flag proposals in both the House and Senate. Having won a stay of the FCC mandate in federal court, the strong fair use coalition—including groups that had been more or less supportive members of the BPDG—became very vocal in the halls of Congress, expressing sincere opposition to the flag mandate. It is unclear whether these forces alone could have stopped either the whole bill or the broadcast flag portion of it, but the congressional record suggests the strong fair use coalition was emboldened by the court's ruling and felt free to express the depths of its opposition.

---

[39] The Stevens bill passed the Senate Committee on Commerce, Science, and Transportation as H.R. 5252, which was the number assigned to the telecommunications bill authored by Joe Barton (R-TX) that had already passed the House on a vote of 321 to 101 (*Communications Opportunity, Promotion, and Enhancement Act of 2006*, 2006). Had the Stevens bill passed the Senate, this change would have enabled a conference committee to work out the substantial difference between the two proposals.

[40] In the interest of full disclosure, the perception that the network neutrality debate killed this bill is in this author's best interests. I fought for network neutrality as an intern for Public Knowledge in 2006, and I even recorded Stevens' infamous "Series of Tubes" speech that helped bring a great deal of additional attention to the issue.

While the DTV broadcast flag nearly became the law of the land, proposals for a digital audio broadcast flag gained much of their viability from piggybacking on the DTV flag effort—and even then, they were abandoned while still in a much more formative stage of development. No similar inter-industry coalition developed a proposed technology for implementing the flag, and even members of Congress who supported the DTV flag derided the audio flag proposal as half-baked (see Chapter 6). Despite this, it was contained in two bills, and the similarities between the proposals—strategically employed by audio flag proponents—meant it was taken more seriously than it would have been had there been no DTV flag proposal.

In addition to the provision granting the FCC the authority to adopt a video flag mandate, the Stevens bill also included an audio flag provision, albeit a much more prospective one than the DTV flag authorization. While the DTV authorization (*Communications, Consumer's Choice, and Broadband Deployment Act of 2006*, 2006, § 452) directs the Commission to begin a rulemaking process specifically to implement its original 2003 mandate—with some minor modifications—the audio flag authorization (§§ 453-454) gives the FCC the power to implement a similar rule, but only if a similar inter-industry process leads to substantial agreement. The bill orders the FCC establish a "Digital Audio Review Board" (§454(b)), composed of all the relevant industries, as well as "public interest organizations" (§454(b)(10)). The Board would have had up to 18 months to reach substantial consensus and offer a draft regulation to the FCC, at which point the FCC would be urged "to give substantial deference to the proposed regulation submitted by the Board" (§ 454(d)(2)(A). Were there no substantial inter-industry

consensus to emerge, the bill directed the FCC not to issue an audio flag mandate, but to submit recommendations to the House and Senate commerce committees (§454(d)(3)). Thus, the Stevens bill recognized that a good deal of discussion was necessary to reach anything like consensus on an audio flag. If government-ordered discussion led to such a consensus, the FCC was to implement a rule giving that consensus the force of law; otherwise, the Commission was not to act.

Also in 2006, New Jersey Republican Representative Mike Ferguson introduced legislation granting the FCC the authority to require audio flag compliance for digital radio tuners (*Audio Broadcast Flag Licensing Act of 2006*, 2006). Whereas the audio flag provisions of the Stevens bill would have required a substantial inter-industry consensus, the Ferguson bill made no such stipulation; it simply granted the Commission the authority to impose such a mandate. While the omnibus Stevens bill had a great deal of political muscle behind it and was close to passage, the much more targeted Ferguson bill never gained much traction. The lack of a preexisting system for inter-industry negotiation certainly weighed substantially against its passage. Another factor also weighed against the audio flag proposal: the recording industry, as the primary copyright holders advancing the proposal, does not have the right to withhold content. While broadcasters seeking to use movies and TV shows must negotiate with copyright holders, radio stations—terrestrial, satellite, and internet alike—all have the statutory right to use music for a price set by law. This compulsory licensing system gave the recording industry far less leverage in getting the electronics industry and other interests to the bargaining table. Unlike the motion picture industry, which had the ability to force a

negotiation by threatening to withhold their content, the recording industry had no threat to levy.

Like the DMCA reform bills, the audio and DTV broadcast flag bills provide excellent opportunities to see the strong copyright and strong fair use coalitions in action. Both efforts warranted substantial attention from all interested parties. Further, like the seriousness with which the DMCA reform effort was taken, the fact that the broadcast flag proposals stalled further suggests the growth of the strong fair use coalition. Without some degree of resistance in the courts and in Congress, the broadcast flag mandate would have become the law of the land.

Conclusion

By retelling the political history of each of these four policy proposals, this chapter describes in a qualitative way the story suggested by the quantitative results that are discussed in the next four chapters. In particular, as the strong fair use coalition has grown, it has grown harder to pass strong copyright DRM regulations, and it has become more conceivable that previous DRM regulations would be moderated. The Audio Home Recording Act became law with little substantial resistance, and while the Digital Millennium Copyright Act saw more substantial resistance, the lack of a longstanding strong fair use coalition made its passage far easier than it might have otherwise been. By the 108th and 109th Congresses, from 2003 to 2006, the strong fair use coalition had grown powerful enough that they had to be accounted for by the strong copyright coalition. They helped slow down the broadcast flag proposal, and they actually made some headway in their effort to reduce the reach of the DMCA. While this study only

examines the results of four specific policy debates—too small a number to provide a definitive verdict as to their efficacy—these outcomes suggest that the strong fair use coalition has at least achieved the goal of slowing the speed with which copyright comes to govern digital media technologies.

While valuable, this narrative political history does not adequately demonstrate the interplay between different communication strategies across multiple forums. It closely resembles the quantitative picture of the congressional debate painted in the next chapter; as strong fair use groups, especially NGOs, participated more heavily in the congressional debate, copyright policy became deadlocked. The strong fair use coalition became strong enough to slow the growth of copyright, but they were still faced with a powerful strong copyright coalition that was capable of forestalling any attempt to roll back DRM regulation. The coverage in the major newspapers was of such a low volume as to represent little substantial political benefit to the strong fair use coalition in their attempt to make the DRM policy debate into a larger political issue. The online space, however, was so thoroughly dominated by the strong fair use coalition as to represent a profoundly different view of the same debate.

Combined with the political history described in this chapter, this difference between media suggests that the internet represents a substantial source of political strength for the strong fair use coalition. The political history shows a DRM policy debate that changed very substantially in the relatively short period between 1998 and 2003. Using exclusively the expensive and hard-to-penetrate offline media of congressional hearings and newspaper coverage, the strong fair use coalition may never

have generated much momentum in such a short period. Knowing nothing about the changes in media technology that took place in the ten years leading up to the 108th Congress, such a rapid ascent would seem virtually impossible outside a major public relations catastrophe for the strong copyright coalition. This ascent is thus an important and noteworthy event, warranting further investigation. The results of the following chapters suggest some of the results of such an investigation.

CHAPTER SIX: COMMUNICATING IN CONGRESS

For most national issues of concern to policy coalitions, the US Congress is the

most important source of potential policy change. In the US system, it is exceptionally

difficult to move an idea from proposal to enacted statute; the system is characterized by

a high degree of policy friction, or gridlock (Jones & Baumgartner, 2005, p. 145). Once

the necessary momentum is built to move an idea from bill to law, the final statute that

gets passed tends to represent an over-reaction to the problem, and modifying or

overturning it with a new statute is just as hard. In principle—and, to a large extent, in

practice—the executive and judicial branches are primarily dedicated to implementing

and interpreting these statutes (see, e.g., Rosenbloom, 1983). Thus, in debates over policy

topics where the federal legislature has jurisdiction, policy coalitions expend a large share

of their resources in Congress, hoping to pass—or stop—legislative proposals. Debates

over the future of digital copyright regulation fall into this category. Since copyright is

exclusively the realm of federal law (17 USC § 301(a)),[41] and since the Copyright Office

reports to the Librarian of Congress (17 USC § 701) and is thus ultimately accountable to

Congress, those calling for stronger copyright law (the strong copyright coalition) and

---

[41] While this is true of copyright *per se*, state law also plays a role. For instance, many
copyright holders have used contracts, "signed" by consumer actions such as opening a
labeled plastic wrapping or clicking to indicate agreement, to carve out exclusive rights
that exceed those defined in federal copyright law. See, for instance, Lemley (1995,
1999), Nimmer, Brown, and Frischling (1999), and Samuelson (1999). Importantly for
the purpose of this study, these contracts are often enforced via digital rights management
technologies such as encryption and watermarking, backed by Title I of the DMCA
(Gillespie, 2004).

those calling for weaker copyright law and wider exceptions such as fair use (the strong

fair use coalition), both concentrate their lobbying efforts on the US Congress.

Congressional hearings are a vital source of data for studying the legislative

process, serving a number of purposes for both legislators and scholars. They are reliable

indicators of congressional interest in policy issues (see, e.g., Jones & Baumgartner,

2005, p. 21), and they also serve to measure which groups have access to policymakers

(Leyden, 1995). Hearings also serve as vehicles by which committees choose what

information is sent to the larger chambers; Diermeier and Feddersen (2000) discuss the

strategic import of this sort of information subsidy. Finally, committees use hearings

strategically to redefine issues and claim jurisdiction away from other committees

(Talbert, Jones, & Baumgartner, 1995). This study therefore includes a comprehensive

study of relevant congressional hearings in the three time periods under consideration; in

particular, it follows Leyden's (1995) lead, using hearings as a measure for determining

which groups have congressional access.

Based on less formal research methods, earlier studies have already concluded

that those who support stronger copyright protection generally, and strong prohibitions on

circumventing digital rights management specifically, have long enjoyed greater access

to policymakers (see, e.g., Herman & Gandy Jr., 2006; Litman, 2000). The results

reported in this chapter provide more formal reinforcement for these observations; among

the total population of congressional across all periods under study, a solid majority

called for stronger copyright laws. Yet this advantage eroded over time, and in the latest

period of interest—2003 to 2006—those calling for expanding fair use even enjoyed a

slight advantage; this evolution in rhetoric, as well as changes in the groups represented at the witness table, suggest that the copyright policy subsystem has undergone a substantial shift over the past two decades.

In exploring the specifics of these findings, I first describe the basic characteristics of the congressional data. Next, I examine the rhetorical valence over the entire period under study, followed by a section reporting the shift in rhetorical valence between 1989 and 2006. Then I discuss the shift in congressional attitudes and committee coverage that have played a central role in the direction of and change in the rhetorical valence of congressional documents. Next, I draw on other variables, including the types of organizations represented in hearings, to explain the overall advantage for the strong copyright coalition and the shift toward the strong fair use coalition.

Congressional Data

This study identifies 17 relevant congressional hearings within the three time frames under study. Table 6.1 outlines these hearings and highlights pertinent details such as year held, committee, number of relevant documents, and total documents.

Table 6.1: Congressional Hearings Studied

Year	Hearing Title	Chamber; Committee; Subcommittee	Number Relevant Documents / Total
1990	Digital Audio Tape Recorder Act of 1990	Senate; Commerce, Science, and Transportation; Communications	30 / 35
1991	Audio Home Recording Act of 1991	Senate; Judiciary; Patents, Copyrights, and Trademarks	34 / 50
1992	Audio Home Recording Act of 1991	House; Judiciary; Intellectual Property and Judicial Administration	31 / 41
1992	Digital Audio Recording	House; Energy and Commerce; Commerce, Consumer Protection, and Competitiveness	23 / 26
1995	NII Copyright Protection Act of 1995 (Part 1)	Senate; Judiciary; Courts and Intellectual Property	9 / 22
1996	NII Copyright Protection Act of 1995 (Part 2)	House; Judiciary; Courts and Intellectual Property	43 / 90
1996	National Information Infrastructure Copyright Protection Act of 1995	Senate; Judiciary (full committee)	12 / 34
1997	WIPO Copyright Treaties Implementation Act; and Online Copyright Liability Limitation Act	House; Judiciary; Courts and Intellectual Property	43/78
1998	Intellectual Property Rights: The Music and Film Industry	House; International Relations; International Economic Policy and Trade	9 / 12
1998	WIPO Copyright Treaties Implementation Act	House; Commerce; Telecommunications, Trade, and Consumer Protection	38 / 39
2003	Piracy Prevention and the Broadcast Flag	House; Judiciary; Courts, the Internet, and Intellectual Property	20 / 22
2004	Digital Media Consumers' Rights Act of 2003	House; Energy and Commerce; Commerce, Trade, and Consumer Protection	42 / 42
2005	Content Protection in the Digital Age: The Broadcast Flag, High-Definition Radio, and the Analog Hole	House; Judiciary; Courts, the Internet, and Intellectual Property	14 / 23

Table 6.1, Continued: Congressional Hearings Studied

Year	Hearing Title	Chamber; Committee; Subcommittee	Number Relevant Documents / Total
2005	Fair Use: Its Effects On Consumers and Industry	House; Energy and Commerce; Commerce, Trade, and Consumer Protection	22 / 37
2006	Broadcast and Audio Flag	Senate; Commerce, Science, and Transportation (full committee)	32 / 33
2006	Digital Content and Enabling Technology: Satisfying the 21st Century Consumer	House; Energy and Commerce; Commerce, Trade, and Consumer Protection	9 / 49
2006	Audio and Video Flags: Can Content Protection and Technological Innovation Coexist?	House; Energy and Commerce; Telecommunications and the Internet	24 / 27

Across these 17 hearings, there were a total of 660 documents, 435 of them relevant to this study. This illustrates a reasonably high congressional interest in DRM regulation. The hearings were spread across two committees, Judiciary and Commerce, suggesting a competition over jurisdiction. Despite the traditional jurisdiction of the judiciary committees over copyright law, the commerce committees held the same number of relevant hearings (8), and a few more of their documents were relevant (220 versus 206). Only one other committee, the House Committee on International Relations, held a relevant hearing. See Table 6.2.

Table 6.2: Relevant Documents by Committee

Number of Relevant Documents (Hearings)	Commerce	Judiciary	International Relations	Total
House	158 (6)	160 (5)	9 (1)	327 (12)
Senate	62 (2)	46 (3)	-	108 (5)
Total	220 (8)	206 (8)	9 (1)	435 (17)

Further buttressing the view that hearings reflect strategic competition between Commerce and Judiciary, the change in hearing distribution over time suggests that the commerce committees have pushed their way into the copyright debate. Table 6.3 illustrates how the commerce committees had a much more active schedule of relevant hearings in the most recent period under study.

Table 6.3: Committee Involvement Over Time

Number of Relevant Documents (Hearings)	Commerce	Judiciary	International Relations	Total
1989-1992	53 (2)	65 (2)	-	118 (4)
1995-1998	38 (1)	107 (4)	9 (1)	154 (6)
2003-2006	129 (5)	34 (2)	-	163 (7)
Total	220 (8)	206 (8)	9 (1)	435 (17)

The House Committee on Energy and Commerce had a particularly important role in this shift away from Judiciary toward Commerce, holding 4 relevant hearings in the latest period. This move helps explain the shift in DRM policymaking. As discussed below, Commerce and Judiciary tend to disagree regarding DRM regulation, with Commerce taking an increasingly strong fair use stand over time; thus, this shift suggests that Commerce has taken an aggressive interest in copyright with an eye toward reshaping the debate.

Rhetorical Valence in Congress

Taken together, the documents across all three periods under study were

significantly biased in the direction of calling for stronger copyright law. Figure 6.1

highlights this advantage.

Figure 6.1: Rhetorical Categories, Congressional Documents
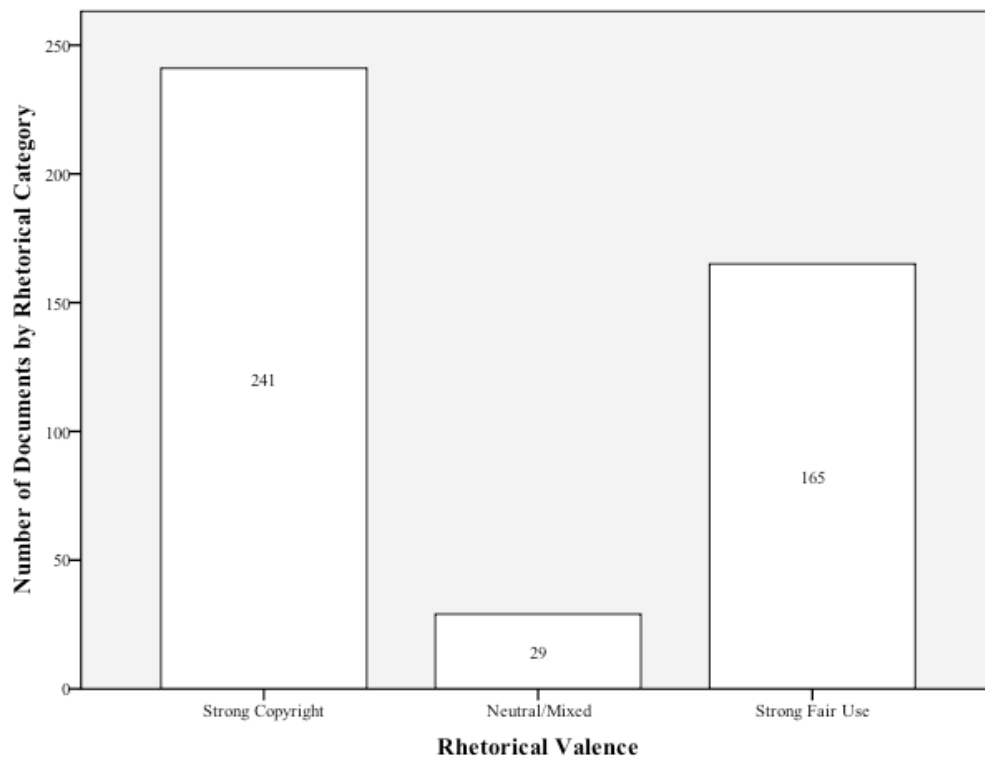
Documents calling for stronger copyright outnumbered those calling for stronger

fair use by 241 to 165—a net advantage of 76 more documents calling for strong

copyright than for strong fair use, for a ratio of 1.45 to 1. The overall advantage for

strong copyright proponents is also reflected in the mean valence score. Where 1

represents strong copyright and 3 represents strong fair use, the mean score for all congressional documents is 1.83. There were just 29 neutral or mixed documents, or fewer than 7 percent of the total. Among all 435 documents, 55 percent of documents called for strong copyright, and strong fair use documents accounted for 38 percent. If this were an election, one would say that the strong copyright side had won in a landslide.

Government voices accounted for most documents that took neither side. Of 29 neutral or mixed documents, congresspersons (22) and appointed government officials (3), such as officials from the Copyright Office and the Patent and Trademark Office, authored all but 4. Among the government-authored documents that did take sides, the advantage greatly favored the call for stronger copyright protection. Congresspersons made strong copyright arguments (43 documents) more than twice as often as strong fair use arguments (20). Documents from appointed federal officials were even more reliable in their calls for stronger copyright, with 28 such documents compared to just 2 calling for stronger fair use. This represents a net total of 49 more strong copyright documents than strong fair use documents, a solid majority (64.5 percent) of the 76-document advantage for the strong copyright coalition across all witness types.

Figure 6.2 highlights this tendency to favor the strong copyright side; the graphed area of each column represents all documents authored by elected (left column) or appointed (right) government officials, and each is divided to illustrate the share of strong copyright, neutral, and strong fair use authored by each witness type.

Figure 6.2: Rhetorical Categories, Congressional Documents by Elected and Appointed Officials
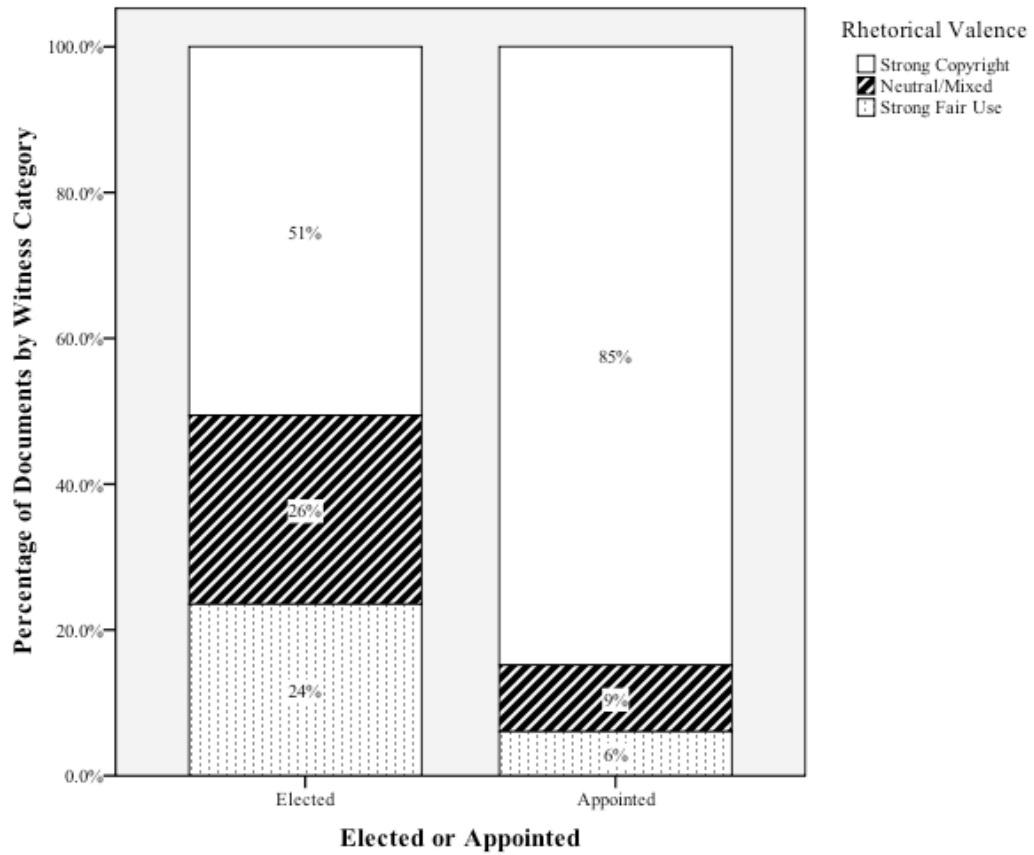
Over the period studied, most congresspersons and nearly every federal bureaucrat in pertinent offices were reliable members of the strong copyright coalition, at least as regards DRM policy. With such an explicit statement of allegiance to the strong copyright coalition, it is unsurprising that members of Congress arranged hearings that tended to favor greater regulation of DRM.

Rhetorical Changes in Congress

    The strong fair use coalition began to make substantial inroads into the

congressional record over the course of the study. By the 2003-2006 period, strong fair

use documents even outnumbered strong copyright documents. This is shown in Figure

6.3.

Figure 6.3: Rhetorical Categories by Period, Congressional Documents
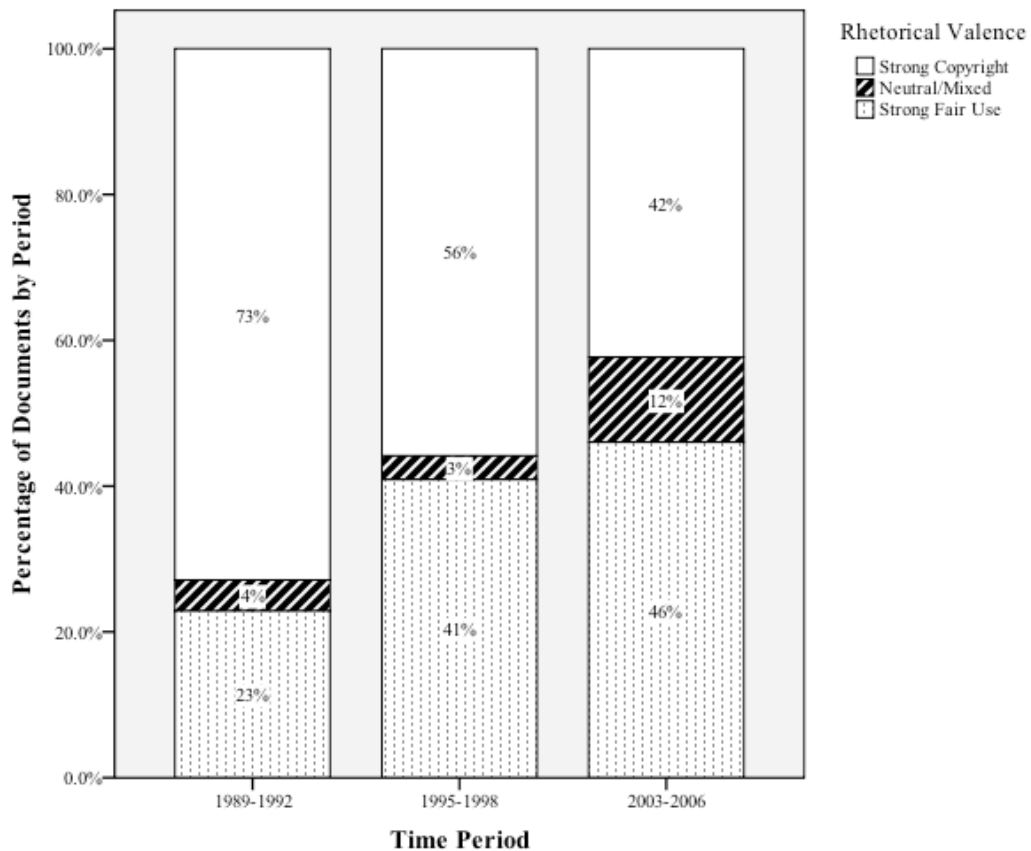
This is a seismic shift in the congressional debate over DRM regulation over time. In the

first period, strong copyright documents outnumbered strong fair use documents by 86 to

27—for a ratio of 3.19 to 1—and the mean rhetorical score was 1.50, remarkably close to the strong copyright score of 1. In the 1995-1998 period, this advantage was 86 documents to 63, a much-reduced ratio of 1.37 to 1, with a rhetorical score of 1.85. By the 2003-2006 period, documents calling for stronger fair use just barely outnumbered those calling for stronger copyright by 75 to 69, for a ratio of .92 strong copyright documents per strong fair use document, with a mean valence score of 2.05. Figure 6.4 highlights the change in rhetorical valence over time.

Figure 6.4: Rhetorical Valence by Period, Congressional Documents

1= Strong Copyright, 3 = Strong Fair Use
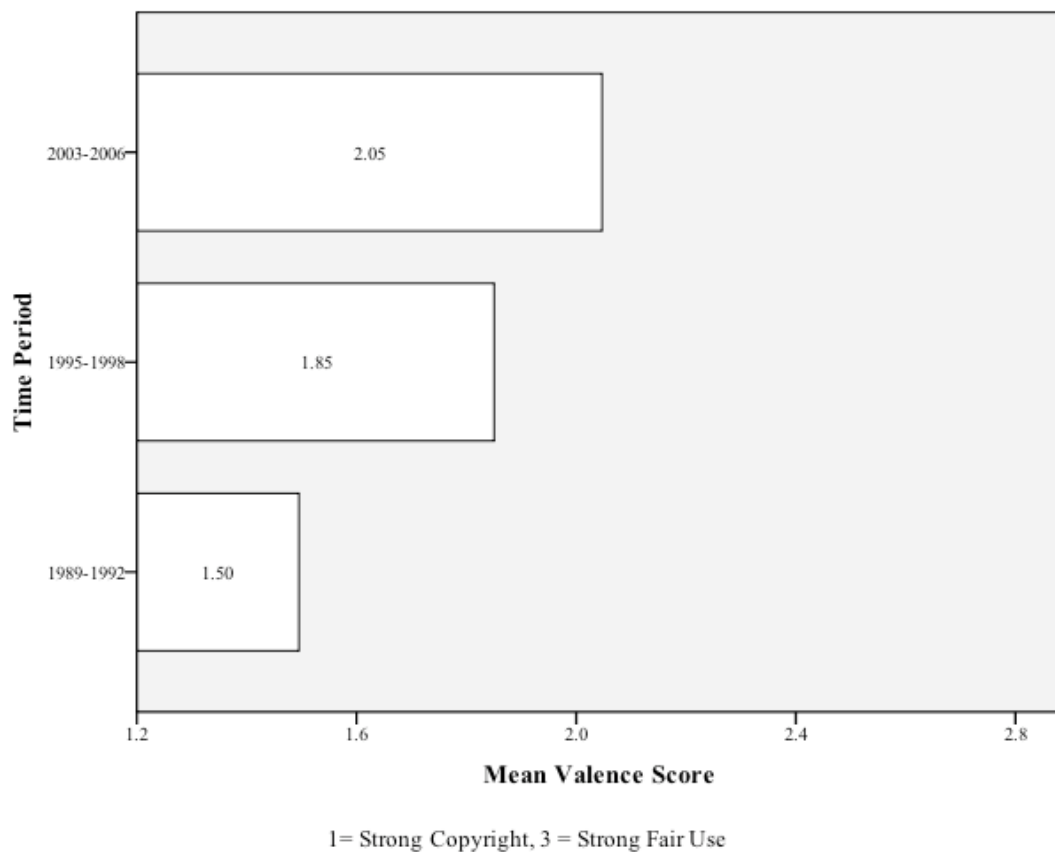
This represents a substantial change over time. Among those documents taking sides—again, a vast majority—a document was 2.33 times more likely to call for strong fair use in the second period than in the first. A given document in the third period was 1.48 times more likely to support the strong fair use position relative to the second period and 3.46 times more likely than a document in the first period.[42] Effect size estimates based on interval valence scores suggest similarly remarkable change over time.[43]

An even more dramatic change happened among the oral communication in these hearings. Hearing documents can be given orally for one of two reasons; either a witness was invited to testify at the hearing, or a member of Congress spoke in person. For witnesses, an invitation to testify is a unique opportunity to communicate directly with policymakers. Members of Congress who attend hear the witnesses' views directly and have a chance to interact, even asking for clarification or further elaboration on points that matter to them. For a member of Congress, speaking at a hearing is also more significant than submitting a written statement; it suggests that the hearing is on a topic of

---

[42] Quasi-odds ratios calculated factoring in neutral/mixed documents produce almost identical results. To calculate these ratios, I needed a formula that would translate the interval-level valence score into a ratio. I did so using the formula:
    Ratio (Strong Fair Use / Strong Copyright) = (Valence Score – 1) / (3 – Valence Score)
This has facial validity; if one begins with a valence score calculated using only strong fair use and strong copyright documents, it produces the correct ratio.
  Calculating odds ratios using these figures yields the following: T2/T1 = 2.52, T3/T2 = 1.48, and T3/T1 = 3.34. These figures are virtually identical to those calculated ignoring neutral documents, reflecting their very small share (6.7%) of the population.
[43] For the difference between the first and second period, $d = 0.39$, which Cohen (1992) describes as a small to medium sized effect. The difference between the second period and the third, $d = .21$, is "noticeably smaller than medium but not so small as to be trivial" (Cohen, 1992, p. 156). The difference between the first and third, $d = .61$, represents a medium to large effect size.

enough importance to justify attending, and it signals to their peers that the policy beliefs expressed are worth taking seriously.

Compared to the valence for all types of submissions, the strong fair use coalition was even less well represented by in-person communication in the first and second periods, and did even better in the third period. From 1989 to 1992, there were only 8 speeches—whether from the witness table or the Congressperson's chair—calling for strong fair use, compared to 37 strong copyright speeches. Setting aside the 2 neutral or mixed speeches, this means a ratio of 4.62 strong copyright documents per strong fair use document. From 1995 to 1998, there were 31 strong copyright speeches, 4 neutral, and 20 strong fair use, or 1.55 strong copyright speeches per strong fair use speech. In contrast, the period from 2003 to 2006 saw 26 strong copyright speeches, 8 neutral, and 32 strong fair use—just .81 strong copyright speeches per strong fair use speech. Relative to the first period, a speech in the second period was 2.98 times more likely to call for stronger fair use. In the third period, a speech was 1.91 times more likely to call for stronger fair use than in the second period and 5.70 times more likely than in the first period.[44] While it has taken the better part of 2 decades, the strong fair use coalition has reason to be encouraged by this degree of progress.

---

[44] As with the calculations all congressional documents, effect size calculations for congressional speeches suggest results similar to those implied by the odds ratios, and in comparing the two, they also suggest an even more significant time effect for speeches. For the difference between the first and second period, the effect was $d = 0.49$, a medium sized effect, or one "likely to be visible to the naked eye of a careful observer" (Cohen, 1992, p. 156). The difference between the second and the third, $d = .31$, was small to medium sized, and the difference between the first and third, $d = .83$, was a remarkably large effect.

Allies In Congress

Members of Congress are potentially potent allies for policy coalitions. At the

start of this study, the only side of the debate with congressional allies speaking during

relevant hearings was the strong copyright coalition, but each successive period saw the

strong fair use coalition add allies in Congress. As discussed below, this shift is partly

reflected in a shift in the types of witnesses invited to participate in these hearings and in

the committees holding them. For more direct evidence, however, consider the substantial

shift in amount and rhetorical valence of documents authored by members of Congress.

Over time, they have taken an increasingly active role, and their rhetoric has moved

toward the strong fair use side. Table 6.4 provides the exact numbers.

Table 6.4: Valence Categories of Congresspersons' Documents Over Time

Valence of Congresspersons' Documents	Strong Copyright	Neutral/Mixed	Strong Fair Use	Total Documents	Mean Valence
1989-1992	12	2	-	14	1.10
1995-1998	10	5	6	21	1.81
2003-2006	21	15	14	50	1.87
Total	43	22	20	85	1.73

The "Total Documents" column of Table 6.4 highlights how congresspersons'

contributions to the debate have risen dramatically over time. Relative to either of the

first two periods, they contributed more than twice as many documents in the period from

2003 to 2006. As the total number of relevant documents per period was rather stable

(118, 154, and 163 documents, respectively), this rise in congressionally authored

documents also increased their relative share of documents, which started at 11.9 percent,

rose to 13.6 percent, and jumped to 30.7 percent.

This impressive growth in congressional expression of opinion reflects the increasingly contentious politics surrounding DRM policy. The Audio Home Recording Act was a relatively easy move for Congress; once the industries agreed on the bill's specifics, it only required congressional ratification. Thus, no congresspersons spoke against this expansion of copyright. The debate has become far more contested, and congresspersons jumped into the fray. As the debate opened up, congressional support for stronger DRM regulation began to erode. Figure 6.5 illustrates this change.

Figure 6.5: Rhetorical Categories by Period, Documents by Congresspersons
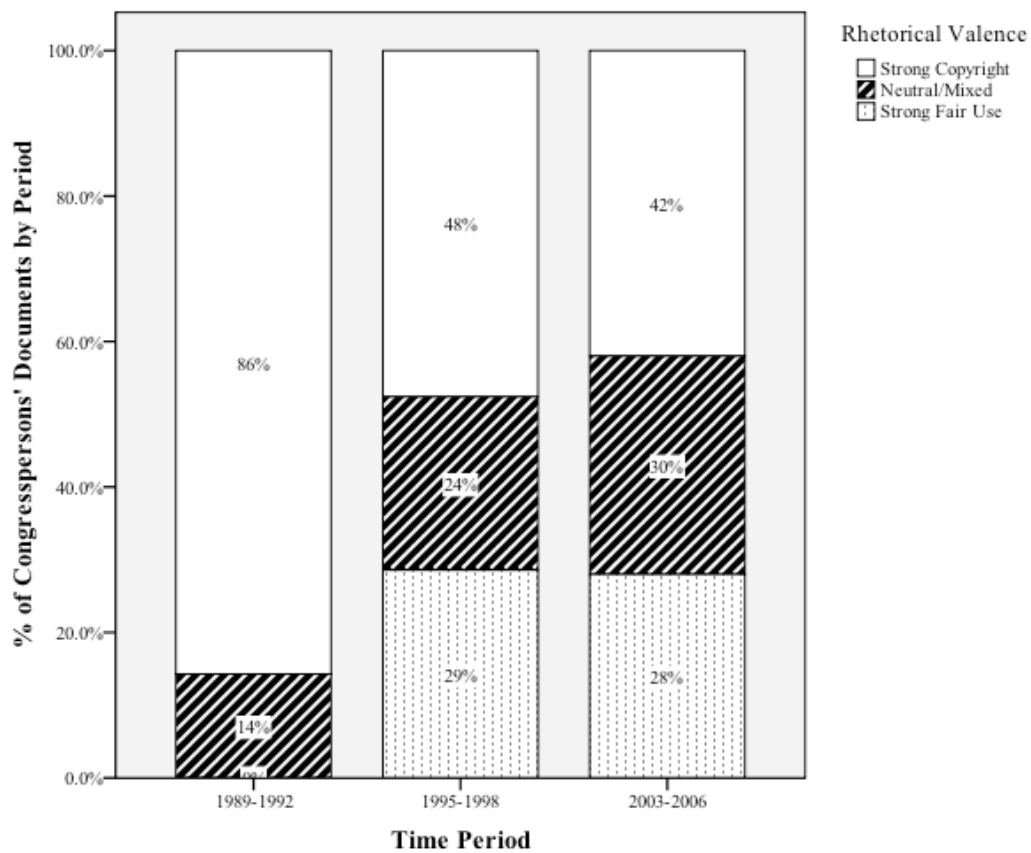
Relative to the period from 1989 to 1992, the proportion of congresspersons' documents advancing the strong copyright position decreased in each succeeding period, from 86 percent to 48 percent to 42 percent. The second period represents a watershed on this count, in which members of Congress for the first time began to express doubts about the wisdom of strong-copyright DRM regulations. The increasing number of neutral documents also suggests that members came to see both sides of the debate as having some merit—or, at least, as having some political clout. This trend is also reflected in each period's mean valence score. In the first period, congressionally authored documents had a mean valence score of 1.10, indistinguishable from the strong copyright score of 1; the scores got much closer to the neutral score of 2 in the second and third period—1.81 and 1.87, respectively.

Relative to earlier periods, congresspersons' documents in the latest period also showed a more sophisticated understanding of specific policy debates.[45] For instance, of the 15 neutral documents, five expressed support for a video flag mandate and opposition to an audio flag mandate. Congresspersons expressed detailed opinions on the relatively more advanced state of inter-industry dialogue regarding the technological specifics of the video flag to be deployed and contrasted this with the less advanced state of dialogue around the audio flag. This increasing level of sophistication suggests that the strong fair use coalition has, over time, helped raise congressional awareness that DRM mandates do

---

[45] While this study does not engage in systematic analysis of specific arguments, the act of coding for overall rhetorical valence did lend itself to observing some patterns; this is one that stuck out.

come with important costs—that the strong copyright coalition's mantra of protecting the creative industries presents an incomplete picture.

The relevant congressional committees becoming increasingly friendly to the strong fair use coalition was the natural result as that coalition picked up a number of key allies in Congress. In particular, the middle time period saw Representative Rick Boucher, Democrat of Virginia, become a tireless advocate for the strong fair use side in the debate over the regulation of DRM technologies. During the DMCA hearings, he was the only Representative on both the Commerce and Judiciary committees. Along with Representative Scott Klug (R-WI), Boucher fought to limit the reach of the DMCA's anticircumvention provisions as the bill was moving through the 105th Congress (Herman & Gandy, 2006, pp. 146-147). He has been the primary legislator behind the push to reform these provisions since they became law. These efforts have drawn commendation from others in the strong fair use coalition; the *Library Journal* recognized Boucher as its 2006 "Politician of the Year" (Berry III, 2006), and the advocacy group Public Knowledge honored him with a 2004 IP3 award, its highest honor (Public Knowledge, 2008).

Klug is no longer in Congress, but other congresspersons have joined Boucher on the strong fair use side, including John T. Doolittle (R-CA), Zoe Lofgren (D-CA), and Joe Barton (R-TX). Barton was an especially important pick-up in the latest time period, as he chaired the House Committee on Energy and Commerce from February 2004 through the end of 2006. In the 105th Congress, the bill that became Title I of the DMCA, HR 2281, passed through both the Judiciary and Commerce committees, creating a

precedent that DRM regulation is under the jurisdiction of both and opening the door for inter-committee struggles over issue definition.

Barton exploited this dual jurisdiction, holding 4 relevant committee hearings from 2004 to 2006. Due to Commerce oversight of the FCC, it might seem that this was merely a natural result of the flag issue coming to the fore. But the hearings show exactly the opposite pattern; Barton's committee led the charge in calling for DMCA reform, while other committees held most of the hearings related to the broadcast flag. Of the 4 hearings dedicated to the flag, 2 were held in the House Judiciary Committee's Subcommittee on Courts, the Internet, and Intellectual Property (34 total relevant documents), and 1 was held in the Senate Committee on Commerce, Science, and Transportation (32 relevant documents). Only 1 was held in the House Energy and Commerce committee (24 relevant documents, plus 6 flag-relevant documents from other hearings), and the remaining 3 hearings held in that committee were called to discuss proposed reforms to Title I of the DMCA. Thus, the House Energy and Commerce Committee held hearings that led to 74 of the 76 DMCA reform-related documents; without Boucher's vocal support and Barton's leadership, this proposed reform would have been dead on arrival. Instead, it had at least a fighting chance, even if the bill's day in the sun never quite materialized. Of particular note, Barton explicitly pushed a compromise between the coalitions: bundling broadcast flag legislation with Boucher's DMCA reform bill (Barton, 2006, pp. 44-45). The content industry never took the deal, but it was the closest the reform bill ever came to becoming legislation.

In the DRM debate, it matters a great deal which committee holds a hearing; in

particular, the differences between Judiciary and Commerce were substantial in the second and third periods. In addition to the number of hearings and hearing documents dedicated to each side's agenda, consider the overall valence of all documents from the hearings these committees held. By the second time period, the commerce committees had become friendly venues for the strong fair use coalition; in contrast, the judiciary committees remained steadfast in producing hearings supporting the strong copyright coalition. Figure 6.6 highlights each committee's change over time.[46]

Figure 6.6: Rhetorical Valence by Period, Within Congressional Committees

1= Strong Copyright, 3 = Strong Fair Use

---
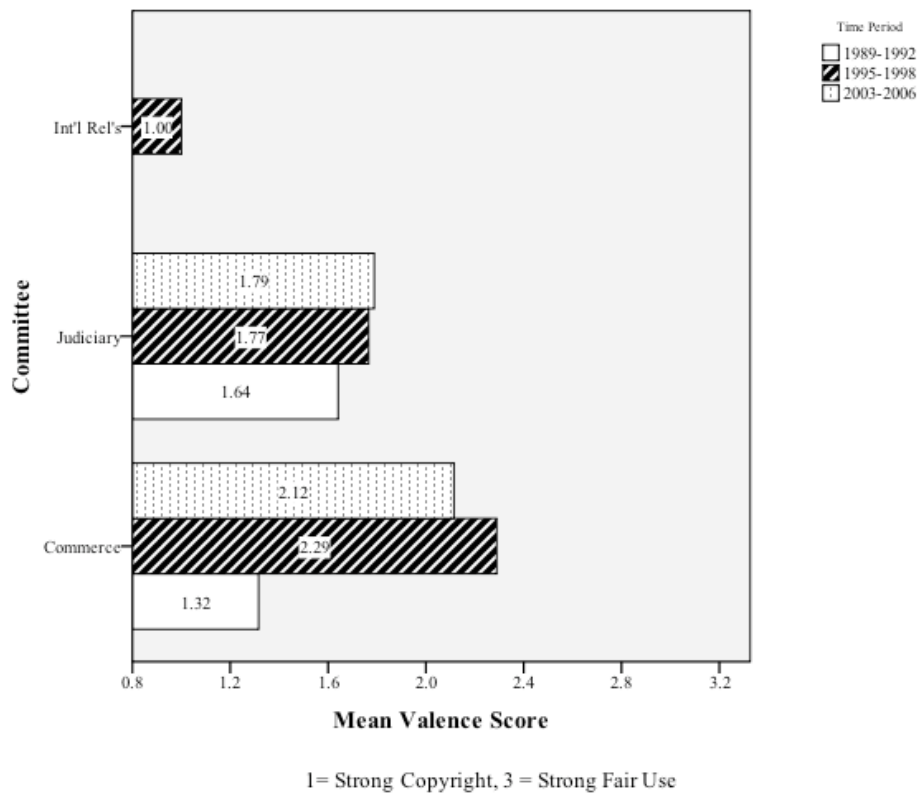
[46] The graph's scale exceeds the range of possible scores; otherwise, the International Relations Committee's score, a perfect 1, would not have shown up on the graph.

In hearing documents from the first time period, the mean valence score for Commerce is very much on the side of stronger copyright (mean valence score = 1.32). However, this shifts to an overall mean in support of strong fair use in the second and third time periods (2.29 and 2.12, respectively). Judiciary documents also move in the same direction between the first and second period, but the shift is much smaller, and the overall mean valence stays clearly on the side of stronger copyright for each period (1.64, 1.77, and 1.79, respectively).

The commerce committees, especially the House Energy and Commerce Committee, also play an important role in shifting the overall congressional record into a more favorably strong fair use position simply by playing a more active role in the most recent period. There are just 2 relevant hearings in Commerce from 1989 to 1992 and 1 from 1995 to 1998. In contrast, the period from 2003 to 2006 featured 5 relevant hearings, overshadowing the 2 in the House Judiciary Committee. Having moved closer to those calling for expanded fair use, their increased attention to this issue helped shift the overall valence in Congress slightly toward the strong fair use side.

The House Commerce Committee's increasing role was particularly helpful for the strong fair use coalition. They held just 1 hearing in each of the first 2 periods (with 23 and 38 relevant documents, respectively), but this grew to 4 hearings (97 relevant documents) in the third time period. Since most Commerce hearings were in the House, the rhetorical valence mirrored the numbers for all Commerce documents—that is, for the set including both House and Senate hearings—moving from 1.35 in the first time period

to 2.29 in the second period and 2.14 in the third. While the third period's valence was less strongly pro fair use relative to the second period, the much higher volume of documents was enough to move the mean valence of all documents into neutral territory. Under Barton's leadership and with Boucher doing his best to make political hay out of the issue, the House Commerce Committee provided an important venue for the strong fair use coalition in the latest period.

The shift in congressionally authored documents was part of the broader erosion of the strong copyright coalition's rhetorical edge in Congress, but the shift of venue, from Judiciary to Commerce, played an even greater role in amplifying the rhetoric of fair use. A vehicle for this shift was the change in faces at the witness table, the subject of the next section.

Representation in Congress

Government voices—Congresspersons and appointed bureaucrats—have automatic or near-automatic access to participation in congressional hearings. When these policymakers take a stand—as most have done in favor of strong DRM regulation—that is a substantial part of the explanation for the policy outcomes that result. Yet other groups also play a role in shaping the debate, and thus policy outcomes.

Some of the groups that participate in the copyright debate are quite at home in the halls of Congress, while other groups are rarely heard there. Simply counting participation from various sectors provides a fairly reliable indicator of whether a set of documents will lean toward stronger copyright or stronger fair use, because most groups are fairly reliable in their coalition membership. This breaks down largely along the lines

of copyright holders versus users of copyrighted works. Music, movie, publishing, and software companies generally support strong copyright and strong regulation of DRM circumvention, while librarians, educators, consumer electronics companies, and consumers generally support strong fair use and weak or nonexistent DRM regulation. Group affiliation is the single most important intervening variable in this study. Differences in rhetorical valence—over time, between committees, or between media—can be explained largely by changes in the relative share of the groups who are providing information. While rhetorical changes also occur within groups—shifts among members of Congress, discussed above, are incredibly important, both to policy outcomes and to determining the content of hearings—the story of the evolving debate over DRM policy is largely a story of evolving access to the attention of policymakers.

To better explain the patterns in rhetorical valence described above, this section explores the frequency with which groups participated and how this participation changed over time. First, I briefly describe the 11 sectors into which I divided all participants, providing the basic document count for each. Then I discuss the rhetorical valence for the groups with significant participation. Finally, I explore how changes over time help explain changes in rhetorical valence.

Sectors and Head Counts

Using the witness categories identified in Herman and Gandy, Jr. (2006) as a starting point, this study identifies 11 categories of policy actors. Figure 6.7 documents the representation of all 11 categories.

Figure 6.7: Document Count by Sector, Congressional Documents
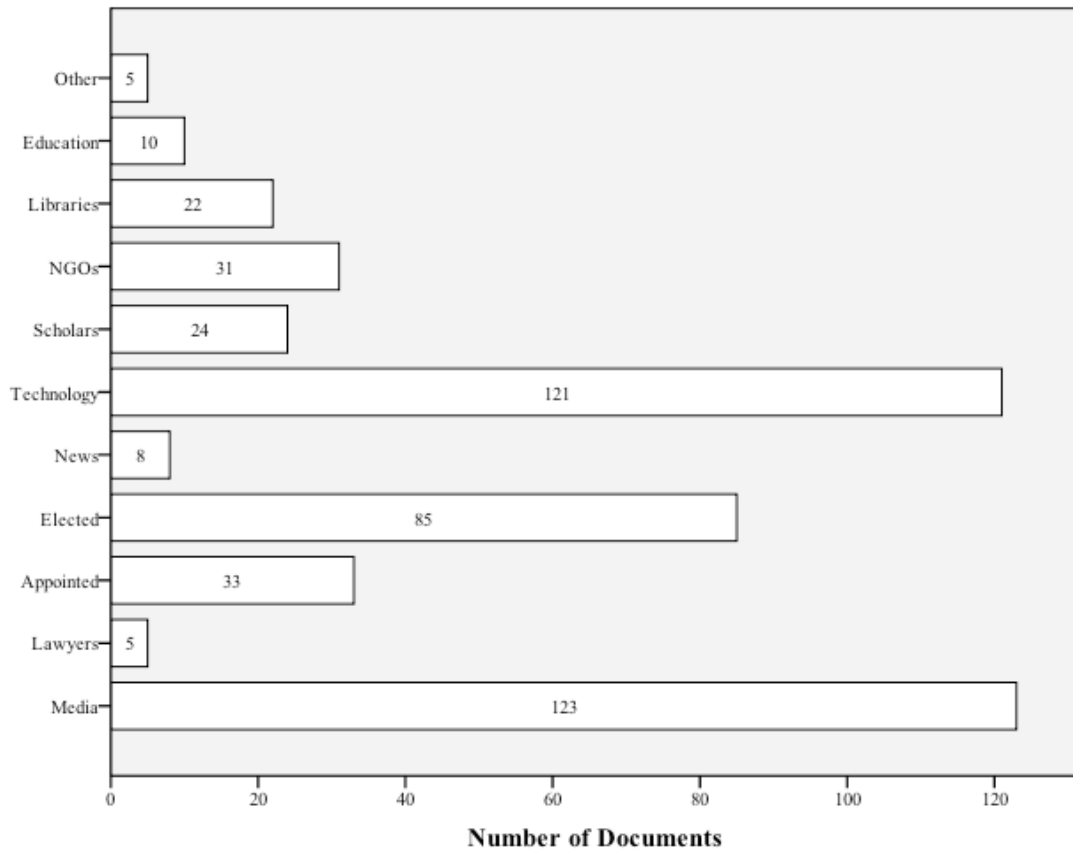
Among all categories, the media sector (123 documents) and the technology sector (121) were far and away the most frequently represented groups. The media sector includes industries such as movies, recorded music, radio and television broadcasting, newspapers and other periodicals (except trade publications dedicated to technology), and books. Otherwise unaffiliated persons with media jobs, such as musicians, producers, directors, writers, and journalists are also included here. The most commonly represented among these are movies and recorded music.

205

The technology sector includes industries such as consumer electronics, software (including video games), and computer hardware, as well as periodicals (generally magazines) dedicated to any of the above. Firms that develop and sell DRM technologies, such as Macrovision and Digimarc, are also included here, as are groups that develop and sell DRM circumvention technologies, such as 321 Studios.

As the media and technology sectors include the for-profit industries with the most at stake and the most capital to invest, it is unsurprising that they have made the hefty financial investment to appear with such high frequency (see, e.g., Leyden, 1995). Removing members of Congress (85 documents) and appointed government officials (33) brings this domination of hearings into even more stark relief: the media and technology sectors authored 244 of 317 unique documents[47] offered by non-government sources, or 77 percent of the total. This domination by for-profit sectors exerts substantial force in the direction of stronger copyright law, as discussed below.

Other sectors also contributed substantially to the record. Elected members of Congress (85) contributed documents with very high frequency, and appointed federal officials (33) were the next most common participants. Nongovernmental Organizations, or NGOs (31 documents), also contributed quite frequently. These groups are dedicated to representing the public, including consumers. Examples include Public Knowledge, the Electronic Frontier Foundation, and the Progress and Freedom Foundation.[48]

---

[47] Sector codes were applied nonexclusively; thus, there were 467 sector codes applied to just 435 documents.
[48] Unsurprisingly, many participants described their policy positions as being in the public's best interest, but most participants have direct ties to one or more of the other

Scholars speaking on their own behalf (24), libraries and librarians (22), and representatives of educational institutions such as universities (10) round out those with at least 10 appearances. Along with technology and NGO totals, these numbers are a bit inflated. Due to nonexclusive coding, the Digital Future Coalition (DFC) was counted as representing each of these 5 sectors. With 8 total documents, this created 32 extra sector codes. Grouping all nonprofit sectors—NGOs, scholars, libraries, and education—as one, and including DFC documents, there were a total of 63 documents. Even discounting DFC-authored documents, the technology sector still appeared 113 times.

Three groups appeared more rarely: representatives of legal associations and other unaffiliated lawyers (5 documents), news articles submitted for the hearing record (8), and participants who fall into none of the above categories (5). With such small numbers, these groups are excluded from further analysis here.

Coalition Allegiances of Groups

It matters who appears at the witness table because different sectors have different loyalties and agendas. This can be quantified by examining the share of documents from

sectors described here. With just two exceptions, only groups that had no ties to other sectors were coded as NGOs.

This policy subsystem has few faux public interest groups; the industry groups that participate generally identify themselves as such. Two groups in this study are coded as representing both nonprofit and other sectors: the Digital Future Coalition (representing nonprofit, technology, scholars, libraries, and education) and the Chilling Effects Clearinghouse (nonprofit and scholars). The former is a coalition of multiple groups, including several true NGOs, and the latter has a track record of soliciting input and participation from the general public; this is not true, for instance, of the consumer electronics-driven Home Recording Rights Coalition. Of these two, only the Digital Future Coalition appears in the congressional record, with 8 documents. This accounts entirely for the 32 extra sector codes: 8 documents with 4 extra sectors apiece.

each sector that support a given rhetorical position. Figure 6.8 illustrates the share of each

sector's documents within each of the three valence categories: strong copyright, neutral,

and strong fair use. For simplicity's sake, all nonprofit sectors—scholars, NGOs,

libraries, and educational institutions—have been collapsed into one category, and the 8

Digital Future Coalition documents have been included in the nonprofit totals and

subtracted from the technology sector's totals. The graph also excludes sectors with less

than 10 documents.

Figure 6.8: Rhetorical Categories by Regularly Participating Sectors
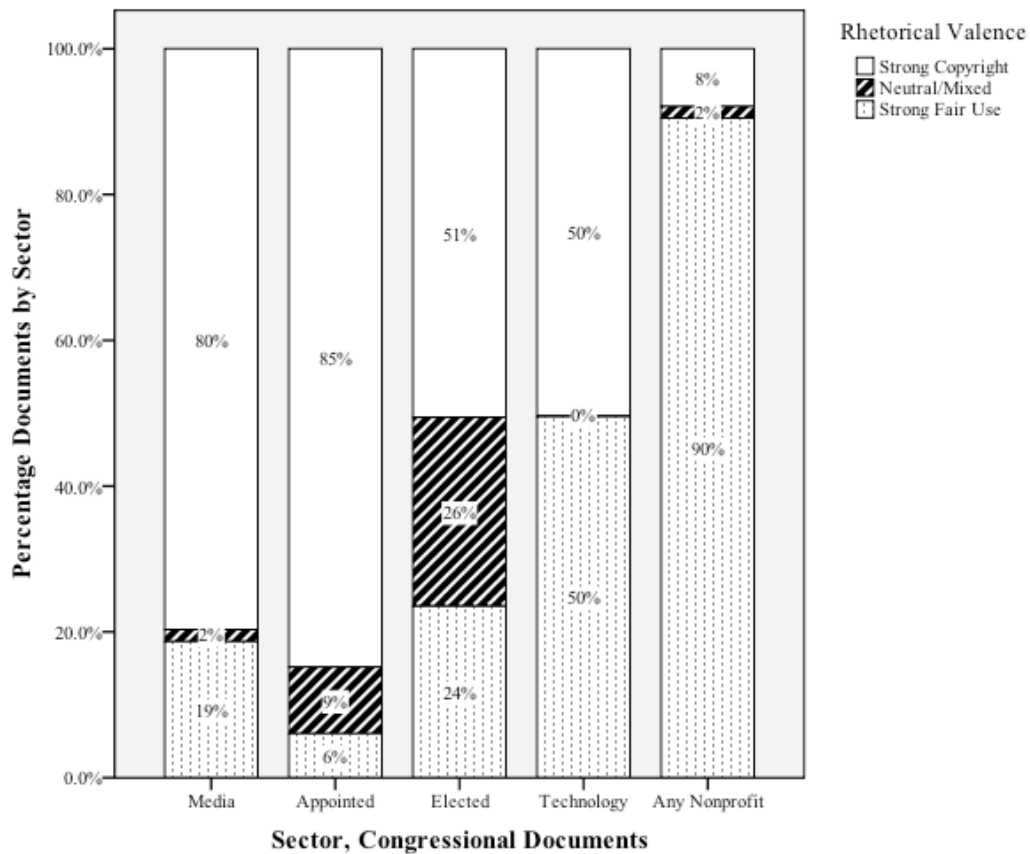
This is a remarkably clear illustration of which sectors are in which coalitions. The media sector and appointed officials were quite reliably in the strong copyright coalition, with 80 percent and 85 percent of their respective documents calling for strong DRM regulation. In contrast, 90 percent of the nonprofit groups' documents called for less DRM regulation. All three groups had less than 10 percent of their documents in the neutral category. In the debate over DRM policy, these groups are the rhetorical anchors at the far ends of the spectrum.

As discussed above, members of Congress were more divided, with 51 percent supporting the strong copyright coalition, 26 percent neutral, and 24 percent in the strong fair use coalition.[49] As this changed over time, it reflected an environment in which a different type of hearing became possible. The technology sector was almost perfectly divided. Among 113 documents—excluding the 8 authored by the Digital Future Coalition, all of which supported stronger fair use—57 were in the strong copyright camp and 56 were in the strong fair use camp. This changed over time, as discussed below.

Looking at the combination of sector representation and allegiance, the distribution of witnesses suggests the strong copyright coalition's preferred level of congressional access. In turn, this helps explain how the AHRA and DMCA became law, as well as why the broadcast flags came close to becoming mandatory and the DMCA reform bill—the only proposal that would have reduced the reach of DRM regulation— had less of a chance of passing. Again, the media and technology sectors dominated the hearings, framing the debate as being primarily of concern to for-profit industries. This

---

[49] Totals equal 101% due to rounding error.

framing benefits the strong copyright coalition in two ways. First, as long as the technology industry gets the exemptions it needs to do business, regulations of DRM technology seem less objectionable. The resulting legislative language thus reads like a contract between industries. Litman (2000) describes this process in some detail, in particular as it occurred during negotiations over the Digital Millennium Copyright Act (see, e.g., pp. 122-145). If left up to these sectors, DRM regulation can continue to expand, leaving space only for the exceptions for which the technology sector has successfully fought.

The second way this framing helps accelerate the growth of copyright springs from diversity within the technology sector. As a heterogeneous group of companies, the sector is of two minds about copyright law. Some companies—and some divisions within larger firms—are primarily in the business of selling copyrighted software, such as operating systems, productivity suites, entertainment software, and high-end creative software. For many of them, DRM and regulations against hacking it are important business tools. Obviously, DRM vendors are also quite happy to see laws requiring the implementation of their products or making it illegal to tamper with them; the former automatically increases sales, while the latter increases these products' perceived efficacy[50] and reducing the need to out-engineer every would-be circumventor with a broadband connection. Other companies and divisions sell mostly hardware; they generally see DRM regulations unfavorably. For obvious reasons, companies selling

---

[50] There is some considerable debate as to the degree to which various DRM deployments actually do reduce infringement, a difficult thing to measure; this study takes no position on this empirical question.

DRM circumvention software are also in favor of loosening or eliminating the rules governing circumvention and the marketing of circumvention devices. Those within the technology sector who push against the growth of copyright have thus been fighting the war on two fronts, both within their own ranks and against strongly driven interests in the media sector. Combined with the strong copyright coalition's ability to appease various technology interests with specific exemptions and concessions, this is an exceedingly difficult battle to win. If the battle is between the media and technology sectors, it is a recipe for the unending expansion of copyright.

Changes in Representation and Shifts in Allegiance

Over the time periods studied, some groups had an increasing presence in the hearing record, and the technology sector moved toward becoming a more reliable part of the strong fair use coalition. Most remarkably, the nonprofit sectors' participation rose substantially in each successive period, both in absolute numbers and in relative share. In the first time period, nonprofit actors authored just 8 documents, compared to 48 documents by the media sector and 24 by the technology sector. By the second period, there were 19 nonprofit documents, 37 media sector documents, and 58 technology sector documents. By the period from 2003 to 2006, nonprofit groups had reached virtual parity, contributing 36 documents—compared to 38 for the media sector and 31 for the technology sector. Figure 6.9 depicts the evolution of each group's relative share over time.

Figure 6.9: Sector Participation by Time Period, Regularly Appearing Nongovernment Sectors
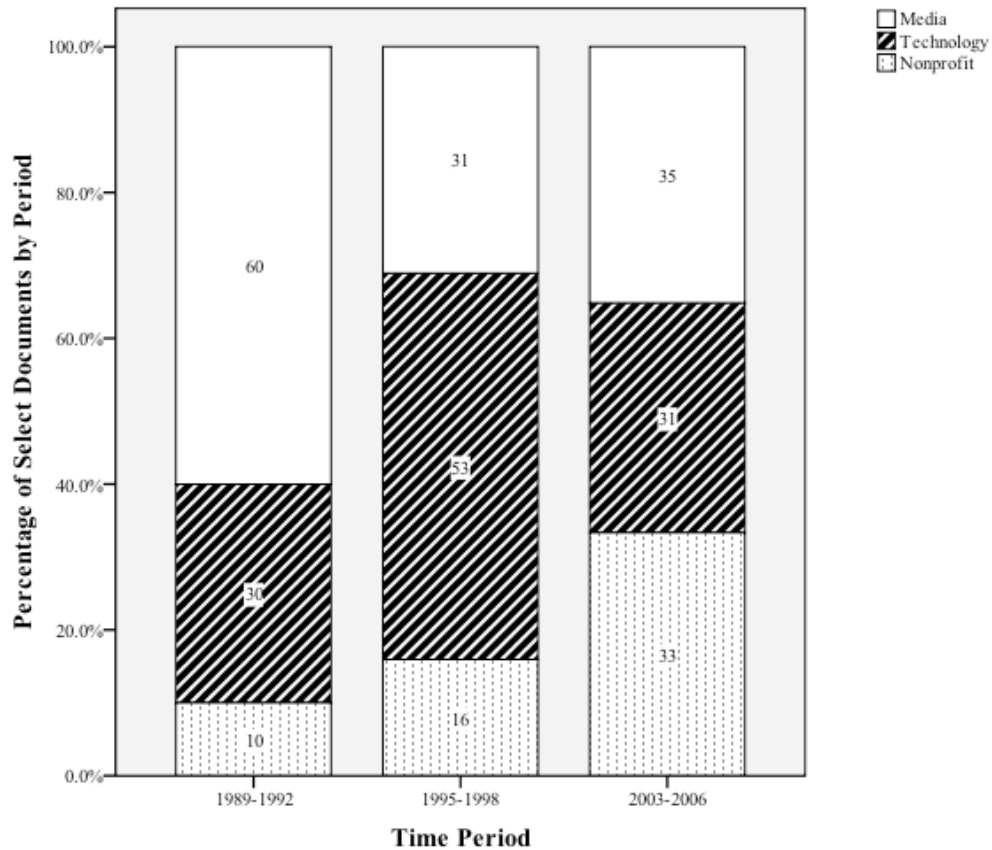
This is a substantial growth in visibility for the nonprofit sectors. A document authored by a person or group in any of these sectors was 1.67 times more likely to be authored by a nonprofit sector in the second period relative to the first, 2.06 times more likely in the third than the second, and 3.43 times more likely in the third than the first.

This growth was due in no small part to the deliberate creation of NGOs designed in whole or substantial part to voice the public interest on exactly this issue. First came the Digital Future Coalition, or DFC, "forged in 1995 in response to the release of the Clinton administration's White Paper on Intellectual Property and the National

Information Infrastructure" (Digital Future Coalition, n.d.). The ideas in the White Paper were the starting point for Title I of the DMCA, so the DFC was essentially formed to retard the lurch toward sweeping regulation of DRM. The DFC authored 5 documents in the middle period and 3 in the latest period—an important contribution to the growth of public sector presence. Public Knowledge, created in 2001 "to defend citizens' rights in the emerging digital culture" (Public Knowledge, 2008), has had even more success gaining access to the hearing record.[51] From 2003 to 2006, Public Knowledge authored or coauthored 13 documents—1 coauthored with the DFC—which was a vital part of the growth in nonprofit presence in the latest period.

    A similar story occurred at the witness table; nonprofit participation in this most exclusive part of the hearing record[52] was slight until the third period. Nonprofit sectors accounted for just 3 in-person witnesses in the first period, far fewer than the media sector (18) and technology sector (10). From 1995 to 1998, nonprofit witnesses testified in person just 4 times, compared to 10 media witnesses and 20 technology witnesses. By the latest period, there were 12 nonprofit sector witnesses, 12 technology witnesses, and 13 media witnesses, an almost perfect three-way split. A witness from one of these three

---

[51] In the interest of full disclosure, I interned with Public Knowledge in 2006, and I sought out Peter Jaszi, the driving force behind the DFC, for consultation on this research project. In both cases, I did so having already come to the conclusion that they were important players in the DRM debate; these numbers justify that conclusion.

[52] A citizen or group with something to contribute to a specific debate need not have been invited to testify in order to submit written items to be added to the hearing record. In order to follow hearing schedules adequately to time such submissions for inclusion, however, one still must have a nontrivial amount of financial or human capital devoted to such an outcome; thus, inclusion in the hearing record suggests a certain degree of resource mobilization by a statement's author(s).

categories was 2.76 times more likely to be from a nonprofit group in the third period than in the second period and 3.35 times more likely than in the first period.

Another contributor in the shift toward strong fair use was the movement of the technology sector toward the strong fair use camp. They were strongly in favor of the 1992 AHRA, viewing it as a compromise that would allow them to sell digital audio recording devices, and 83 percent of their documents from the first period (20 out of 24) were in support of this expansion of DRM regulation. In the debate over Title I of the DMCA (from 1995 to 1998), however, 57 percent of technology sector documents (33 of 58) were in the strong fair use camp, and 61 percent of documents (19 out of 31) from 2003 to 2006 supported the strong fair use position. Since resource mobilization is such a vital part of policy advocacy (Leyden, 1995), and since the technology sector has access to a much larger pool of capital than the nonprofit sectors—especially NGOs, who do the lion's share of work in this area—their movement toward stronger fair use is of vital importance to that coalition.

<div align="center">Conclusion</div>

Over the course of this study, Congress shifted away from very strong support for ever-stronger copyright-based regulation of DRM technologies, moving to a position of rhetorical neutrality. This neutrality was reflected in the policy outcomes of the period from 2003 to 2006. After the DC Circuit overturned the FCC's broadcast flag order in 2005, the motion picture industry pushed hard for video broadcast flag legislation. Backed by the recording industry in 2006, Representative Mike Ferguson also pushed for his bill to mandate an audio broadcast flag; his bill and the recording industry testimony

attempted to piggyback on the video flag effort. Both bills had a realistic shot at passage, yet both stalled due in part to stiff opposition from the strong fair use coalition[53]—opposition that got real traction, thanks largely to congressional allies such as Representative Boucher and Chairman Barton. This same coalition also saw DMCA reform move from near invisibility to an idea that got a genuine airing and a little movement. The policy environment did not change so much in Congress that this coalition had an easy time passing their agenda, but they had at least fought the strong copyright coalition to a draw.

By the period from 2003 to 2006, Congress had come to a deadlock on DRM policy. The strong copyright coalition and the strong fair use coalition each had the means to forestall attempts by the other coalition to advance major new legislation. For

---

[53] For instance, video and audio flag proposals were folded into the omnibus Senate telecommunications bill spearheaded by Senator Ted Stevens, Republican of Alaska (S.2686, 2006, §§ 451-454). Since this bill was the companion to legislation that passed the House (H.R. 5252), these flags were close to becoming law. The Stevens bill was doubly objectionable to Public Knowledge, for it would have cemented in statutory law the principle that broadband service providers have no real obligation to provide a neutral internet; as a strong advocate for network neutrality and a leading member of the strong fair use coalition, Public Knowledge put every effort into stalling the Stevens bill. Had the bill passed and gone to House-Senate conference committee, the strong fair use coalition would have been grateful to have an ally such as Chairman Barton with a good shot at removing the flag provision.

While this project is focused on copyright policy, the rancor over network neutrality had much more to do with stalling Stevens' bill. Yet the attempt to pass the flag proposals via inclusion in a much larger bill was an end-around, suggesting the strong copyright coalition knew they had less of a chance of passing it as standalone legislation. For example, the House telecommunications bill did not include a flag provision. While the broadcast flag debate had little to do with the collapse of the Stevens bill, the fact that a broadcast flag mandate hinged on the outcome of a bill that had nothing to do with copyright suggests this chapter's overall conclusion that the strong copyright coalition's position in Congress is less solid than it once was.

the strong fair use coalition, merely getting to this form of deadlock was a victory. In 1992 and 1998, major bills passed, expanding the role of copyright as a vehicle for governing digital media technology. In both instances, the bills represented compromises between for-profit industries, even though the minor changes introduced during negotiations over the DMCA suggest that the strong fair use coalition was beginning to have some success in the fight for those outside the commercial sector, including librarians and academic researchers.

By the 2003 to 2006 period, it was no longer possible for the strong copyright coalition to negotiate DRM regulations with a divided technology sector and present the compromise for congressional ratification. Not-for-profit groups had become too substantial a counterbalance, the technology sector had become stronger allies of the fair use coalition, members of Congress had begun to voice support stronger fair use, and the commerce committees of both chambers—and the House in particular—had taken an increasingly pro-fair use position and become more active in shaping the congressional record on DRM policy. Thanks to these changes, the overall rhetorical tone in the relevant congressional hearings moved from strongly supportive of the strong copyright position to relatively neutral. This is a recipe for gridlock, and that is exactly what transpired.

CHAPTER SEVEN: COMMUNICATING IN NEWSPAPERS

Major newspapers are important vehicles for policy actors to gain the public's attention. The broader public may generally get their news from other sources, such as television news and local newspapers. Nonetheless, major national papers—especially the *New York Times* and *Washington Post*—are vitally important for a number of reasons. Elite papers and the *Times* in particular exert the greatest inter-media agenda setting effect (McCombs, 2004, p. 113). The media exert a substantial third-person effect on policymakers, who use the amount and tone of coverage to gauge public opinion (Mutz, 1998, pp. 51-53), and policymakers themselves are particularly likely to read the *Times* and the *Post* to decide which issues and frames are growing in importance (Kingdon, 2003, p. 60). Elite papers are more likely to be read by the part of the public that follows and cares about national policy issues—exactly the part of the public whose opinion is most important on policy issues. Both newspapers make large investments in the coverage of national policy issues generally and science and technology specifically, "with a large and prestigious staff of science writers and editors. ... Given their influence, both papers are primary targets of media lobbying by various political actors" (Nisbet & Huge, 2006, p. 19). In tracking mainstream news coverage of national technology policy, then, the *Times* and *Post* are natural choices.

If one or more policy coalitions seek public attention on an issue, and if the issue fits newspapers' definition of newsworthiness, substantial coverage may ensue. The public is regularly inundated with coverage of various policy debates, and the public

perception of highly visible issues can shape policy outcomes. As is the case in this

particular study, it may be the case that one policy coalition has every incentive to seek

public attention on a set of policy issues (in this instance, the strong fair use coalition

seeking attention regarding DRM policy), while another policy coalition seeks no such

attention to those issues (here, the strong copyright coalition). In this circumstance, even

the very amount of coverage is a politically contested outcome, and as noted by

Baumgartner and Jones (1993), a change in the amount and tenor of coverage suggests an

increased chance for substantial policy change. Thus, any study of the communication of

specific policy issues must at least acknowledge the potentially influential role of

newspaper coverage and, if possible, examine such coverage. For the reasons discussed

above, the *Washington Post* and *New York Times* are particularly important forums for

policy debate coverage. This chapter discusses these two newspapers' coverage of DRM

policy in three 4-year windows from 1989 to 2006.

    First, I describe the basic distributions of the newspaper articles included in this

study. Then, I discuss the overall valence of their coverage. Next, I look at how this

valence has changed over time. Finally, I consider how the sectors quoted in newspaper

stories may provide a partial explanation of the overall rhetorical valence.

<center>Amount and Distribution of Newspaper Coverage</center>

    The *Post* and *Times* each featured a modest number of relevant articles in each

period. The *Times* had a total of 36 across all three periods, and the *Post* had just 22. See

Table 7.1.

<center>218</center>

Table 7.1: Newspaper Coverage by Period

Number of Relevant Articles (Total copyright-related articles)	New York Times	Washington Post	Total
1989-1992	16 (177)	5 (107)	21 (284)
1995-1998	6 (530)	7 (366)	13 (896)
2003-2006	14 (1000)	10 (431)	24 (1431)
Total	36 (1707)	22 (904)	58 (2611)

Especially compared to the wealth of data generated by congressional hearings, this is a fairly sparse amount of coverage. Spread across the time period studied, these numbers suggest the newspapers' editors do not believe DRM policy is of particular public interest. This is not to say they think the same of copyright in general or copyright concerns related to digital media specifically; quite the contrary, copyright as a subject appears very frequently. These relevant articles were identified from a total population of 2,611 documents for which LexisNexis identified "copyright" in the headline, lead paragraph, or indexing terms. While this study does not identify the exact number, the cursory glance at each article required to identify the relevant articles convinced this researcher that there are hundreds dedicated to concerns about the scope of digital copyright infringement—the very concerns that led to the proposal and adoption of the DRM regulations studied. For instance, in the periods from 1995-1998 and 2003-2006, the unauthorized transmission of copyrighted content via the internet is a topic of what must be hundreds of stories. From 2003 to 2006, the *Times* featured 193 stories with "copyright" in the headline, lead paragraphs, and indexing terms and one of the following terms in the body of the story: "bittorrent," "bit torrent," "grokster," "napster," "peer to peer," "peer-to-peer," and "p2p." In the same period, the *Post* had 88 such stories.

Both newspapers treated concerns about the digital transmission of copyrighted content as a subject of substantial concern, but both failed to provide extended coverage of some of the most important policies that resulted from these concerns. This shortfall is most apparent in the context of the 1998 Digital Millennium Copyright Act. As early as 1994, the Patent and Trademark Office publicly signaled that the administration would push for something like the DMCA. The negotiations leading up to the 1996 signature of the 2 WIPO treaties featured an extended international debate over the future of copyright law in the digital age. The DMCA itself represents the most profound change in copyright law since 1976 (Litman, 2000). Title I of the DMCA has profoundly reshaped copyright law. In the 4 years leading up to its passage, the nation's flagship newspapers featured just 13 stories combined. One *Post* article, excluded because it had too little relevant content, passes on detailed coverage of the anticircumvention provision, describing it as "relatively noncontroversial" (Mills, 1996, p. C01).

If history is any guide, major daily newspapers are too thin on detailed policy information for any citizen who wants to follow the debate over digital rights management policy. In light of the content of this debate, this should be unsurprising. "Media agendas are shaped far more by the news values of immediate events and situations than by the social value of deliberation" (McCombs, 2004, pp. 101-102). There are many criteria for what determines the newsworthiness of a given event; one author identifies the orthodox set of news values as including timeliness, relevance to the audience, the audience's potential identification with an event, the degree of conflict involved, and sensation (Schultz, 2007, pp. 196-198).

Under the criteria identified by Schultz or virtually any similar set of criteria, the DRM policy debate rarely involves newsworthy events. The general public has little perception that encryption is relevant to their day-to-day lives; this study is conducted in the belief that it plays an important role in shaping the media environment, though even I must concede that this is an indirect impact for most people. Few DRM-related events offer chances for audience identification; for instance, most people would have a hard time sympathizing with encryption researchers who do not follow specific lines of research due to vague legal threats. In the context of policy debates, the newsworthiness of a conflict is generally indexed to visible political figures (Bennett, 1990), and until the recent past, the degree of high-level political conflict has been fairly low; Congresspersons on the strong fair use side of the debate have never included the top congressional leadership of either party. Finally, DRM regulation rarely involves sensational events.

This relatively low newsworthiness contrasts sharply with a well-covered issue: the legal saber rattling of the music industry against peer-to-peer users. These legal machinations often create discrete, timely news events that are sensational, create a potential sense of identification in many readers, and foreshadow potential direct effects on the millions of internet users who continue to download music illegally. These peer-to-peer lawsuit events—attorneys issue sternly worded letters by the hundreds, lawsuits are filed, high-profile cases are settled, juries reach findings, and so on—have more general-interest news value and are thus more frequently covered in the news.

DRM regulation can sometimes become fodder for high-profile news events. A few topical news stories in the recent past have come out of sensational, timely events involving high-stakes courtroom conflict and a moderate degree of potential audience identification. The stories of Professor Edward W. Felten and Dmitri Sklyarov present such examples.

Felten, a professor of computer science at Princeton, was part of a research team that in 2000 successfully disabled several watermarking schemes created by the recording industry's Secure Digital Music Initiative (SDMI). SDMI had posted an online challenge, complete with a cash prize of $10,000, but the researchers eschewed the prize and the nondisclosure agreement that came with it. The researchers were scheduled to present their findings at the Fourth International Information Hiding Workshop in April of 2001, but the presentation was cancelled in light of a cease-and-desist letter from Matthew Oppenheim, the litigation chief of the Recording Industry Association of America (RIAA) and secretary of SDMI. Invoking Title I of the DMCA, Oppenheim threatened the researchers and conference organizers with legal action if the research results were presented (Markoff, 2001). Outraged, Felten went public, creating a public relations nightmare for the recording industry. Felten used a David and Goliath narrative, portraying himself and his colleagues as researchers trying to do their jobs, only to be confronted by corporations' legal threats based on an obscure law explicitly aimed at preserving these corporations' profits. This was a newsworthy event and led to headlines such as "Record Panel Threatens Researcher With Lawsuit" (Markoff, 2001) and "Group

Says It Beat Music Security but Can't Reveal How" (Harmon, 2001b). In 2000 and 2001, the *Times* ran 7 articles covering this story and the *Post* ran 5.

In 2001, a similar story began when federal agents arrested Russian programmer Dmitri Sklyarov, the first criminal arrest under Title I of the DMCA. Sklyarov, then an employee of Russian software company Elcomsoft and Ph.D. student at Moscow State Technical University, was part of a team that had hacked the encryption on Adobe's eBooks. Sklyarov presented his team's findings at the DEF CON convention in Las Vegas, and Adobe filed a complaint with the FBI, leading to Sklyarov's arrest on July 16 (Lee, 2001). The following days saw such vocal public protests that, by July 23, Adobe reversed course and publicly asked the Justice Department not to prosecute Sklyarov (Harmon, 2001a). The Justice Department released Sklyarov on the condition that he testify against Elcomsoft; the company was ultimately acquitted in 2002 (Richtel, 2002). In 2001 and 2002, this story merited 19 articles in the *Times* (including 2 that also discussed Felten's story) and 4 in the *Post*.

For the strong fair use coalition, these are valuable opportunities to communicate with the public about the perils of DRM policy, but these opportunities are rare for exactly that reason; they must be initiated by copyright holder behavior that can reasonably be cast as bullying, so paradoxically, copyright holders have a long-term strategic interest in pursuing few legal cases. The current situation favors the strong copyright coalition: the threat of legal action prevents the development and sale of DRM circumvention devices in the mainstream commercial market, and it scares many researchers and tinkerers away from reverse-engineering research, but it does so without

223

garnering widespread public attention. Thus, copyright holders can do a lot to keep this issue out of the newspapers simply by not giving the strong fair use coalition a sensational story line to pitch. From 2000 to 2002, the Felten and Sklyarov stories merited 33 total stories in the *Times* and *Post*, 38% more than all DRM policy-relevant stories combined from 2003 to 2006.[54] The lack of similarly newsworthy storylines during the 2003-2006 period helped lead to the low story count and suggests that copyright holders have decided to avoid DRM-related legal actions that garner public attention.

Unlike in the other results chapters, the amount of data available is itself an important finding. The small number of relevant articles suggests that substantial policy change via public attention is unlikely—assuming a world in which newspapers serve as the gatekeepers to the public's awareness of policy issues. Despite the low number of articles, though, it is still important to consider the rhetorical valence and the representation of various sectors in newspaper coverage. These are covered in the following sections.

---

[54] One might ask whether this undermines the methodological choice to focus on DMCA reform debate beginning in 2003, but the Felten and Sklyarov incidents were hardly the spark for an immediate congressional push to reform the law. Of 21 copyright-related hearings between the start of coverage of Felten's story (November 2, 2000) and the end of 2002, only 1 meets this study's standards for relevance to the DMCA reform debate; its title was, "Consumer Benefits of Today's Digital Rights Management (DRM) Solutions" (2002), hardly suggesting an impending revolt against the DMCA.

Rhetorical Valence in Newspapers

Relative to the highly contentious debate in Congress, the newspapers provided a reasonably balanced view of the DRM policy debate, with a much smaller proportion of documents falling cleanly on either policy coalition's side. A slim majority of all documents were neutral or mixed (30), compared with 16 supporting the strong fair use position and 12 supporting the strong copyright position. See Figure 7.1.

Figure 7.1: Rhetorical Categories, Newspaper Articles
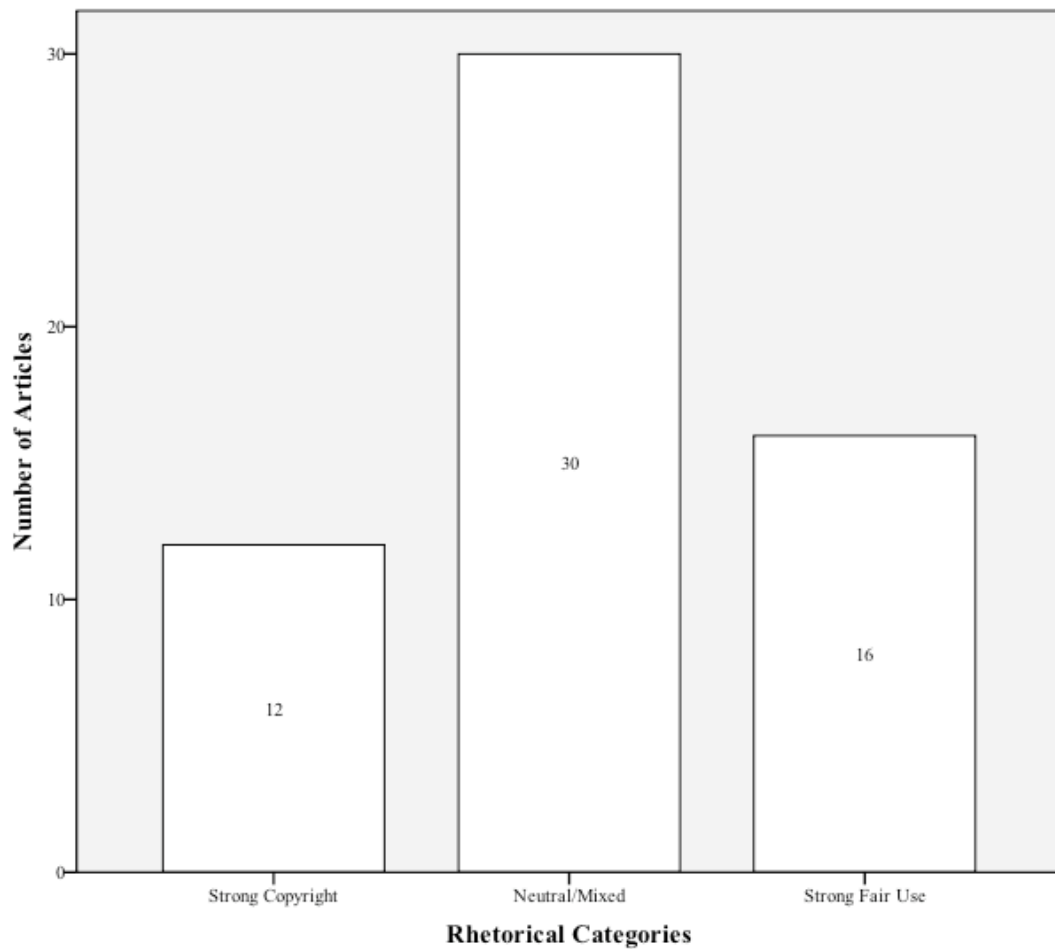
Unlike the highly polarized set of congressional documents, in which the distribution of valence is U-shaped, the distribution of valence in newspaper articles is closer to normal. This reflects the journalistic practice of presenting both sides of political conflicts, which is caused by "the fear of appearing biased, [and] which leads to a formulaic 'he said, she said' reporting style that can stymie citizens in their search for understanding" (Jamieson & Waldman, 2002, p. 168). Nearly a century of journalists' self-identification as objective—itself a strategic reaction to skepticism of the press (Schudson, 1981, p. 122)—has pushed reporters toward the middle on all conflicts; as long as there are two or more well-represented sides to a subject, the story is reliably presented as a political conflict between two groups. This leads to a relatively normal curve on most issues; some articles will present one side or the other, but most will try to present both sides of a conflict.

This contrasts sharply with congressional documents, in which the vast majority of speakers and authors participate because they hold quite strong beliefs about a policy issue. An industry lobbyist who strived for objectivity would soon be looking for work. A scholar or NGO representative who is undecided on a given policy conflict will not spend the time and expense to publicly declare their impartiality; they will wait for a fight in which they have strong beliefs.

In the case of newspaper articles covering the DRM policy debate, not only were the majority of articles neutral, so was the overall valence of coverage. Among those articles taking sides, there were a slightly higher number of strong fair use articles relative to those favoring stronger copyright (16 to 12). This very minor advantage is also

reflected in the mean valence scores. Where a score of 1 represents strong copyright and 3 represents strong fair use, the mean valence score for all articles is 2.111 (SD = .721). The same is true within the 30 neutral or mixed articles, which have a mean score of 2.083 (SD = .272). Especially with such a small set of articles, these advantages are not meaningfully different from 2, the score reflecting perfect neutrality. An alternative outcome featuring a change in just a few articles could make the entire set skew, to a similarly inconsequential degree, in the opposite direction. Even if these scores represented hundreds of articles and the confidence intervals did not include 2, their small variance from 2—the mean is .154 standard deviations above 2 for all articles, .305 standard deviations for the neutral articles—is tame enough to suggest that any bias remains quite small. Again, this is assuming a much larger data set; these data are just too thin to suggest any definitive results.

Breaking the small sample into even smaller subsets to analyze each paper stretches the thin data even thinner. Of 36 articles, the *Times* ran 8 strong copyright, 8 strong fair use, and 20 neutral articles, a perfectly balanced distribution among the three categories. Of 22 articles, the *Post* ran 4 strong copyright, 8 strong fair use, and 10 neutral or mixed pieces. As discussed below, even the *Post's* distribution is indistinguishable from perfect neutrality. Valence scores are also quite close to neutral. The *Times'* mean score is 2.034 (SD = .713), and the *Post's* mean score is 2.239 (SD = .733).

With the *Post* running twice as many strong fair use articles as strong copyright, and running a smaller proportion of neutral articles, one could argue that the *Post* in

227

particular has a bias toward the strong fair use camp. This is a difficult argument based on such a small set of articles, however, and as with all newspaper articles, the mean score is a small divergence from neutral. The mean valence score is less than a third of a standard deviation (.326) higher than 2. Especially with such a small dataset, that small divergence from neutrality means little, even in the context of a census of relevant articles. Statistical tests suggest as much. Comparing the *Times* (again, a perfectly balanced 8-20-8 split) versus the *Post* on the distribution of articles within the three valence categories, the Chi-Square (1.367, p = .505) is very small and far from significant.[55] Likewise, the confidence interval for both mean valence scores includes the perfectly neutral 2, and the difference between the means (d = .285) is relatively small as well. Both papers fail to provide substantial coverage of the DRM policy debate, and the overall coverage is too close to perfect balance to demonstrate any meaningful bias.

These data are not consistent with Hypothesis 2, that strong copyright messages will be communicated in elite newspapers more often than will strong fair use arguments. If anything, they are consistent with the opposite—that strong fair use arguments appear more regularly—but the dataset is so small and the variation from perfect neutrality so slim that we could not reject the null hypothesis of neutrality in either direction. This may

---

[55] As these articles represent a census of relevant articles from the chosen newspapers, tests of statistical significance are not formally necessary. In the context of these results, they are still a useful tool for examining whether differences represent a reliable pattern. Taking the present example, if these findings or something like them could happen by random chance over 50% of the time, that is a strong indicator that the difference between the two papers is not meaningful; a few editorial decisions could swing the results the other way. Because these results are based on a census, we know that these newspapers really are different, but the results of this particular significance test provide a warning not to take this difference seriously.

simply reflect the journalistic norm of objectivity; given the historical and institutional strength of this norm, it is reasonable to treat this finding of relative neutrality as unremarkable.

Rhetorical Changes in Newspaper

As with comparing the *Times* and *Post*, one must be wary of tracking rhetorical changes in the articles of both papers across three separate time periods, a process that divides the already-thin data into even smaller piles. That proviso in mind, the differences are large enough to justify meaningful conclusions. In short, the newspapers have become friendlier to the strong fair use side over time.

Consider the distribution of articles among the three rhetorical categories in each period. This is shown in Table 7.2.

Table 7.2: Rhetorical Categories by Period

Number of Articles per Category (Expected Count)	Strong Copyright	Neutral / Mixed	Strong Fair Use	Total
1989-1992	9 (4.3)	6 (10.9)	6 (5.8)	21
1995-1998	2 (2.7)	7 (6.7)	4 (3.6)	13
2003-2006	1 (5.0)	17 (12.4)	6 (6.6)	24
Total	12	30	16	58

This distribution is significantly different from a random distribution based on the marginal totals. Four cells in particular stand out. From 1989 to 1992, newspapers ran more strong copyright articles than expected (9 articles, 4.3 expected); this number was higher than even the actual number of neutral articles (6 articles, 10.9 expected). In

contrast, from 2003 to 2006, they ran just 1 strong copyright article (5 expected), running a particularly high number of neutral articles (17, with 12.4 expected). Again, these are based on very small numbers, but a statistical test for significance[56] (Cramer's V = .326, p = .015) suggests that this change over time represents a reliable finding rather than a fluke outcome of a few editorial decisions. Especially considering that these articles are an attempted census, this test (which assumes a margin for sampling error) highlights what can safely be described as a meaningful change over time.

Comparing mean valence scores over the three periods also suggests a meaningful difference, and in particular, between the first and last periods. From 1989 to 1992, the mean valence score was 1.806 (SD = .853), while the second and third periods had substantially higher valence scores of 2.292 (SD = .700) and 2.282 (SD = .515). See Figure 7.2.

---

[56] Cramer's V is more appropriate here than Chi-Square because 4 of the 9 cells have an expected count of less than 5.

Figure 7.2: Mean Valence by Period, Newspaper Articles

1 = Strong Copyright, 3 = Strong Fair Use
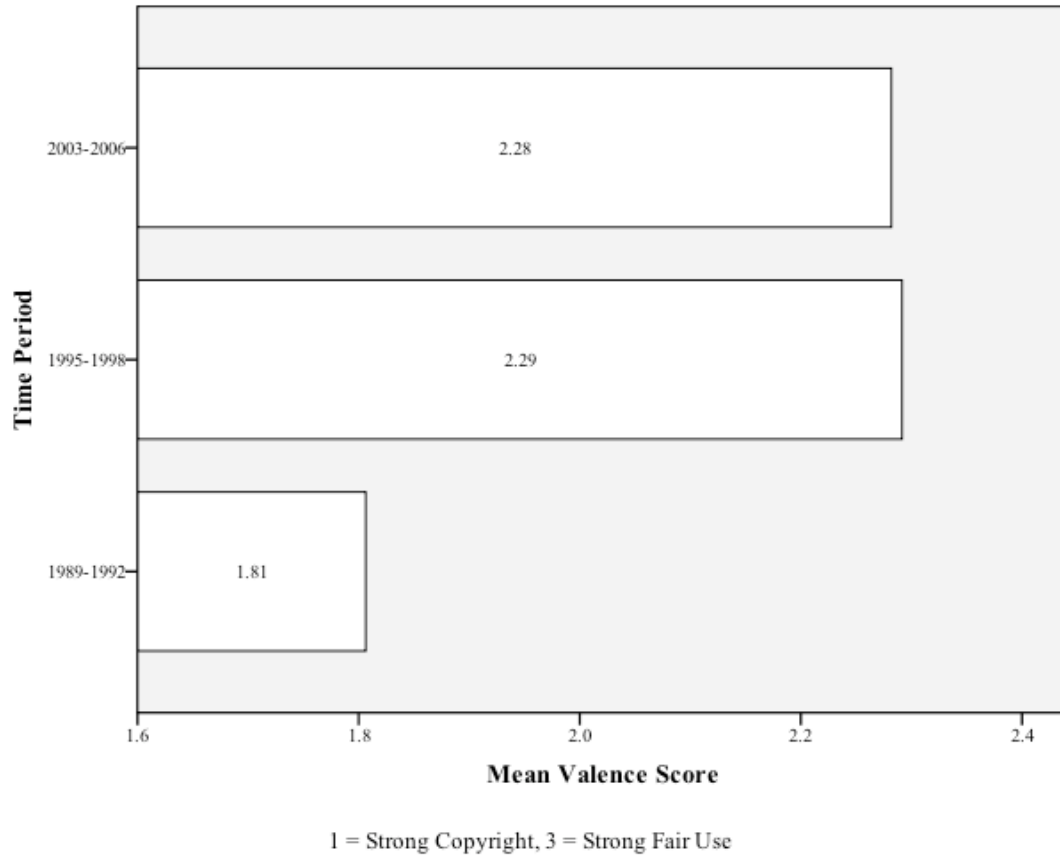
As in the previous section, one might reasonably hesitate to describe any one period as suggesting a bias away from neutrality due to the small number of documents, but the difference over time is substantial enough to overcome this concern. The second period has just 13 articles, so comparisons between it and either the first or third period are fairly tenuous. This is less of a problem for the first period (21 articles) and the third period (24), and comparing these two means suggests a difference that is significant and substantial. Cohen's *d* is .687, which is a medium to relatively large effect size.

231

This study's hypotheses examining change over time predict a difference between the first two periods (lumped together) and the latest period. Apropos of this chapter, Hypothesis 5 predicts: The ratio of strong fair use messages to strong copyright messages in elite newspapers is higher from 2003 to 2006 than from 1989 to 1998. Phrased this way, there is still an identifiable change, but it is less substantial. Consider the numbers for categorical valences, as presented in Table 7.3.[57]

Table 7.3: Rhetorical Categories, Earlier Periods versus Latest

Number of Articles per Category (Expected Count)	Strong Copyright	Neutral / Mixed	Strong Fair Use	Total
1989-1992, 1995-1998	11 (7)	13 (17.6)	10 (9.4)	34
2003-2006	1 (5.0)	17 (12.4)	6 (6.6)	24
Total	12	30	16	58

While they are not strictly necessary for a census such as this, tests for significance suggest that it is highly unlikely that this outcome would result from a few stray articles (Chi Square = 8.39, p = .015). However, comparing the mean valence for the two groups is less consistent with this conclusion.[58] If this study were a random sample, one would not be able to exclude sampling error as an explanation for the variance between the groups. Further, unlike the medium-to-large effect size when

---

[57] This table is the same as Table 7.2, with the first and second lines added together. Even the expected counts for the two grouped time periods for a given cell is the sum of the two groups separate expected cells.

[58] While these tests are not strictly necessary, I ran a t-test and, because the variances are unequal and the dependent variable is arguably ordinal, a Mann-Whitney U test. Were this a sample, one would not be able to reject the null hypothesis in either case: t = -1.647, p = .105, and U = 339, p = .232.

comparing Period 1 and Period 3 (d = .687), the effect size in comparing Periods 1 and 2 versus Period 3 is modest (d = .407). This is undoubtedly due to the mean score for the second period being virtually the same as that for the third period—namely, a time in which the strong fair use group has a minor advantage—and yet being grouped with the first. Phrasing the question in terms of categorical rhetorical valence, the results are loud and clear: the change over time is not likely a fluke. Phrased in terms of mean valence score, however, the results are only consistent with the change-over-time hypothesis because the dataset is based on a census rather than a sample.

In sum, these data are consistent with but only provide modest support for Hypothesis 5: Relative to the period from 1989 to 1998, the rhetorical valence in elite newspapers is more favorable to the strong fair use coalition in the period from 2003 to 2006. As discussed above, the comparison of 1989 to 1992 versus 2003 to 2006 suggests a more reliably identifiable change in newspaper coverage, but the research hypothesis was framed in a way that incorrectly assumed a later inflection point. With just 13 articles in the middle period, one is left wondering if the inflection point could even be correctly spotted. A larger dataset would inspire greater confidence in spotting the exact moment of change.

This study began with the assumption that the latest time period would see a shift toward strong fair use rhetoric in Congress and in newspapers due to the influence of the internet on the policy debate. To the extent that such a small number of articles could be relied upon, these results suggest that the strong fair use coalition began to gain traction in the press before the widespread adoption of online political strategies. This is a

reflection of that coalition's interests; as the group seeking a substantial change in the current political order, they have an interest in expanding the scope of conflict to include a broader political discussion (Baumgartner & Jones, 1993; Schattschneider, 1960). If this mobilization is a key to explaining the shift in the valence of newspaper coverage, one can ask which sectors have been included as quoted sources in newspaper coverage; over time, these sources should include an increasing number of spokespersons from sectors traditionally aligned with the strong fair use coalition.

<div align="center">Sector Representation in Newspaper Coverage</div>

The representation of sectors in newspaper coverage of the DRM debate is similar to that in Congress. The best-represented sectors are the same in both venues: the media sector and the technology sector. As in congressional hearings, the large industries with the most money on the line are the most frequently represented in newspaper quotations. The change over time is also similar to that in Congress: the relative representation of not-for-profit sectors such as NGOs increases over time, suggesting a change in narrative from one of inter-industry feuding to one of broader debate about the public good.

Thanks to substantive differences between the media, these comparisons come with a small caveat. In the other media studied—congressional hearings and web documents—the coding strategy used makes it easy to line up industry sectors with coding units. Most documents are written to advance the goals of one of the two coalitions, and most are written in a single voice. Newspaper articles meet neither of these conditions. They often seek to present both sides to an issue, and they often represent an amalgam of quoted authorities from several sectors who may or may not

agree. Therefore, one cannot make a direct apples-to-apples comparison between the

sector representation in Congress and online on one hand, and the sector representation in

newspaper articles on the other. Thankfully, such direct comparisons are not necessary;

one can track the relative representation of each sector within a medium. Further, this

unique trait of news articles—that multiple sectors may be quoted in any one article—

also opens the chance to ask a unique question: Has the mean number of sectors quoted

per article changed over time? If the number of sectors quoted has increased over time,

this suggests an increasing politicization of the issue of DRM policy, which would be

consistent with the last chapter's findings. I begin this section there, asking whether mean

sectors quoted have indeed increased. Next, I consider which groups have appeared most

frequently across all time periods. Finally, I examine changes over time in the frequency

with which different sectors are quoted.

Change in Numbers of Sectors Represented

Over time, newspapers cite more sectors per article, a finding that comports with

this project's overall finding that DRM policy has become more politically contentious

over time. From 1989 to 1992, the average article cited just 1.24 sectors (21 articles, SD

= 1.14), with a minor increase to a mean of 1.46 in the period from 1995 to 1998 (13

articles, SD = 1.33).[59] Many of these articles contain no quotations from any person

---

[59] Absent from these calculations are news outlets; except for the occasional op-ed or letter to the editor, nearly all news articles are coded as news. Thus, if an ordinary news article quotes no representatives of any other sectors, it is calculated as representing 0 sectors. Additionally, note that each sector is counted only once. If an article quotes two representatives of the technology sector, for instance—say, a spokesperson for a software

representing any sector—no recording industry spokespersons, no electronics retailers, no NGOs. By the period from 2003 to 2006, however, the mean sectors quoted per article rose to 2.33 (24 articles, SD = 1.34). This is a statistically significant effect for time period on mean sectors quoted, F (2, 55) = 4.59, p = .014. Post hoc Tukey HSD Tests show that articles from the first period and the last period differed significantly in the number of sectors cited (p = .015, d = .87). Over time, the increasing rancor over DRM policy is reflected in a ramped-up effort by competing coalitions to gain access to these major newspapers.

Comparing Access Across Sectors

While the overall change in sectors cited is itself interesting, it is also important to ask which industries are regularly quoted in these newspaper articles. As in Congress, the voices of some sectors appeared more often than others. See Figure 7.3.

---

industry trade association, and an electronics retailer—then the article is coded as representing only the technology sector.

Figure 7.3: Number of Newspaper Articles Quoting Each Sector
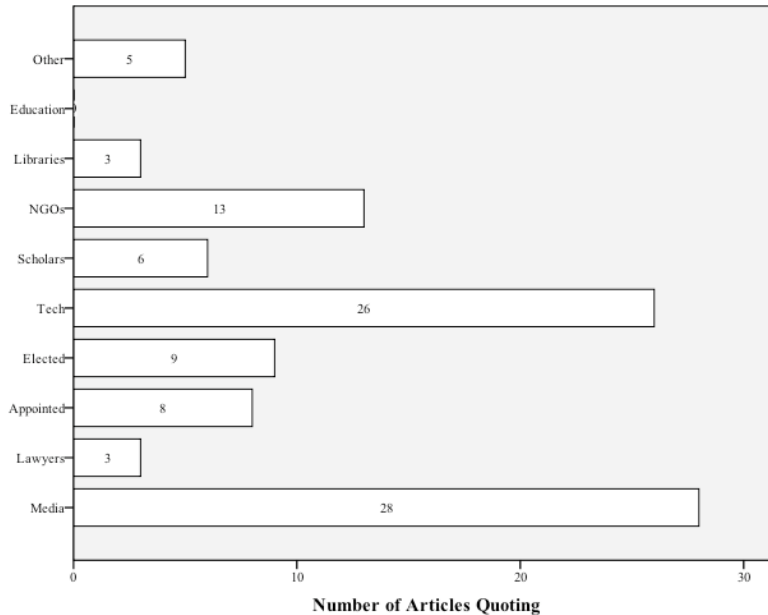
The 2 sectors that appeared most often in Congress were quoted most often in newspaper articles. The media sector, including interests such as the recorded music and motion picture industries, were quoted in 28 articles. Likewise, technology industry sources, representing firms such as consumer electronics manufacturers, computer manufacturers, and software firms, appeared in 26 news articles.

As in Congress, elected and appointed government voices also appeared with relative frequency—9 and 8 articles, respectively—but nongovernmental organizations (NGOs) appeared in 13 articles, good enough for third place. Scholarly voices (6) and others (5) such as ordinary consumers appeared just a bit more often than lawyers (3) and librarians (3). Educational institutions such as colleges are absent, but relative to the much smaller population of documents, this is not too different from their presence in

Congress; there, educational institutions contributed just 10 out of 435 documents, or 2.3 percent.

Changing Sector Representation

Sectors' presence in newspaper coverage has not remained static over time. As DRM policy has become more politically contentious, some groups have enjoyed increasing access to major newspapers. See Figure 7.4, which highlights the changes over time for sectors with at least 6 total articles.

Figure 7.4: Newspaper Articles Quoting Select Sectors by Time Period
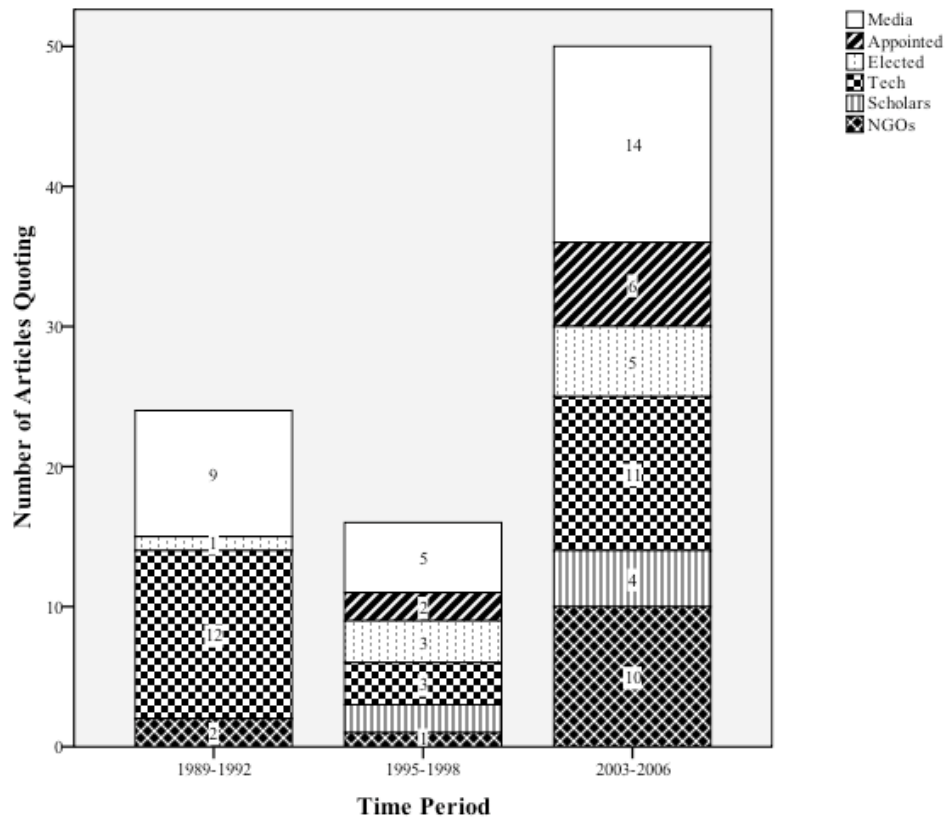
As noted above, the latest time period saw a substantial growth in the number of sectors quoted per article. This figure highlights which sectors saw increasing newspaper presence over time. The presence of the technology and media sectors is quite comparable between the first and last periods. Technology sources appeared in 12 out of 21 articles (57%) in the first period and 11 of 24 in the latest period (46%). In the same periods, media sources were quoted in 9 of 21 (43%) and 14 of 24 (58%). Especially within such a small number of articles, these changes are unremarkable.

In contrast, appointed officials, members of Congress, and NGOs saw their representation increase substantially. Appointed officials were not quoted in any articles from 1989 to 1992, but their voices appeared in 6 of 24 articles (25%) from 2003-2006. Congresspersons were quoted just once in the first period, but this rose to 5 articles (21%) in the latest. Likewise, NGOs appeared in just 2 articles in 1989-1992, a figure that jumped to 10 articles (42%) from 2003 to 2006. See Figure 7.5, which presents a visual representation of this upswing in comparative visibility of select sectors.

Figure 7.5: Proportional Sector Representation in Newspaper Articles by Time Period
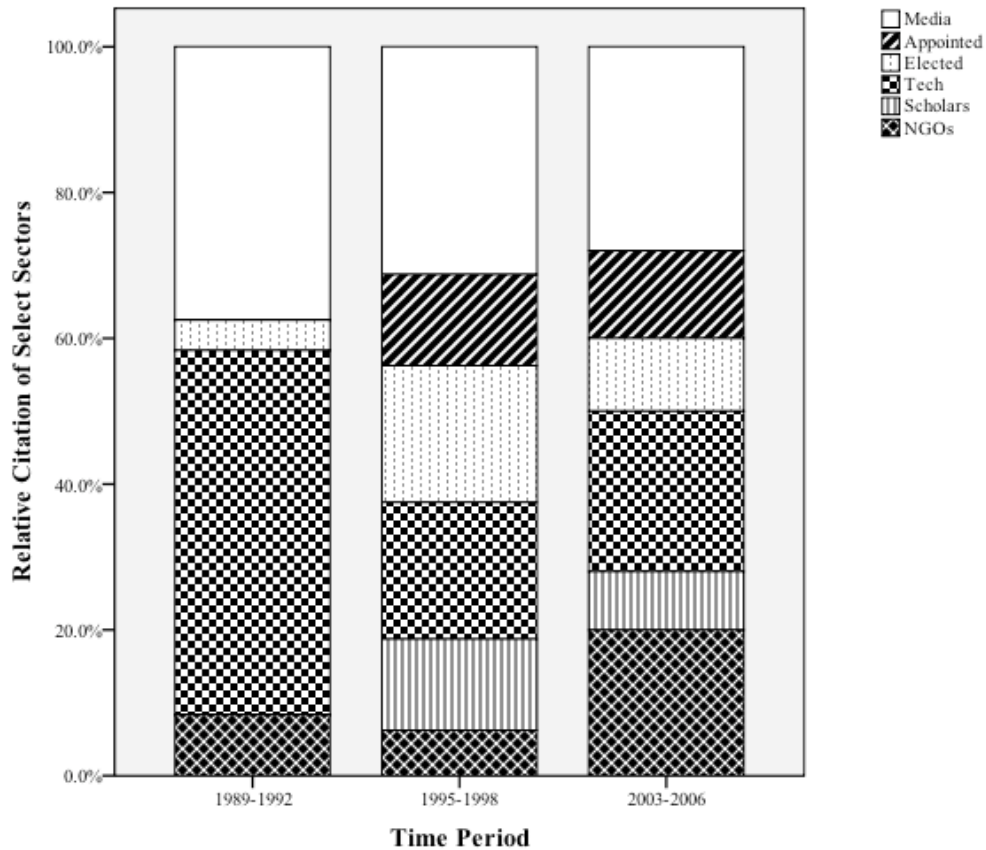
This graph presents the same data as Figure 7.4, but instead of a simple count of the number of articles citing each sector, it changes the scale for each period to 100%. This makes it easier to assess each sector's representation relative to the others in a given period—and to track these relative changes over time independent of the substantial increase in sectors quoted per article.

Note that the media and technology sectors are represented as having lower relative visibility over time. This is because of the increase in other groups' visibility. As discussed above, the media and technology sectors have maintained relatively steady

participation over time in absolute terms. But NGOs, congresspersons, and appointed

government officials have all increased from near silence to meaningful participation. As

measured by the affiliations of quoted sources, newspapers first created the impression

that DRM policy is primarily a private discussion between the media and technology

sectors; later, however, DRM has been portrayed as a policy issue that matters to the

broader public, as represented (if imperfectly) by people such as congresspersons,

appointees, and civil society groups.

The upswing in the appearance of appointed officials is in part due to the topic of

the broadcast flag. Many voices from the Federal Communications Commission chimed

in, both debating the merits of the Commission's flag mandate and defending the

agency's jurisdiction to impose it. In contrast, the upswings in congressional and NGO

participation are the result of people who have deliberately publicized the issue of DRM

policy. The congressperson at the forefront of the effort to reform Title I of the DMCA is

Representative Rick Boucher (D-VA), though the 2003 to 2006 period also saw

important support—including support in the media—from House colleagues including

John T. Doolittle (R-CA) and Zoe Lofgren (D-CA). Likewise, Public Knowledge, a DC

nonprofit advocacy group founded in 2001, has contributed substantially by serving as a

regular source in news stories. [60] The Electronic Frontier Foundation, formed in 1990, has

also served an increasingly visible advocacy role on DRM policy since the passage of the

---

[60] In the interest of disclosure, I must note that I interned for Public Knowledge in the
summer of 2006, but I did so as a volunteer, and the organization has never paid me for
any services.

DMCA. Without quotes from Public Knowledge and EFF, NGO visibility in newspaper coverage would still be very low.

Conclusion

The overall valence of newspaper coverage across all three time periods cannot be confidently distinguished from perfectly neutral, balanced coverage. Over time, however, this study identifies a meaningful change in the direction of favoring the strong fair use coalition. This is best understood as stemming in large part from the concerted efforts of a few policy actors. Sympathetic congresspersons, especially Representative Rick Boucher, and NGOs such as Public Knowledge and EFF, have helped reshape the newspaper dialog around DRM policy. They have contributed to the overall rise in the number of sectors quoted per article, adding to the perception that DRM policy is the subject of meaningful political debate in which the broader public has a meaningful interest. To the extent that these elite newspapers set the parameters for legitimate policy debate, this development is a necessary precursor for the strong fair use crowd's attempts to stop the expansion of DRM regulation and reduce the reach of extant DRM regulations, especially the DMCA.

This reflects the strategies expected by groups who seek to alter the dynamics of a policy subsystem. Historically, those who seek to roll back DRM policy and to halt its further expansion have been outgunned in the relevant congressional committees, especially the judiciary committees. Thus, they have every incentive to seek new venues, including the court of public opinion (see Baumgartner & Jones, 1993; Schattschneider, 1960). As described in the previous chapter, this has meant an escalating involvement by

the commerce committees, where technology interests have more traction. In the context of the public media, it implies exactly what this chapter's results suggest: the strong fair use coalition is seeking to expand the conflict to include the broader public in the discussion.

While these changes over time are important, the most significant finding of this chapter is that newspapers do not provide much coverage of DRM policy debates. This paucity of coverage—an average of just over 7 articles per topic per paper—is not enough to enable even daily readers to follow the intricacies of the policy debates at hand. This is an understandable editorial decision for a general-interest newspaper. In the contest for scarce column inches and ever-scarcer newsgathering resources, this is a story that has too little general-interest newsworthiness; it is not perceived as having much direct relevance to much of the audience, and it rarely involves the kind of sensation and potential audience identification that could overcome this hurdle. Yet if broader public opinion is to be mobilized primarily through newspaper coverage, those who seek to involve the public in debates over DRM policy face a very steep uphill battle. Historically, papers have not provided much coverage, and if actors in the strong fair use coalition seek to change this, they must convince newspapers that the copyright story is about much more than peer-to-peer trading and industry lawsuits.

The various industry sectors have still managed to be informed about these developments without newspaper coverage. Each sector identified here (except "Other") has long had its own print media outlets, from musicians and movie executives to librarians and copyright attorneys. NGOs have a long tradition of circulating their own

studies, proposals, and other literature. These come with disadvantages, however; it costs substantial money to print and circulate a paper periodical, and because they have to be physically delivered to specific people, the reach of such niche publications is relatively limited. These constraints limit the efficacy of these outlets as tools of reaching a broader public. Thus, the earliest period of this study is a world where newspapers serve as the gateway to broader knowledge among the citizenry of specific policy issues. This is a media environment in which limited newspaper coverage greatly reduces the odds that an existing policy dynamic can be altered.

In terms of the media environment, at least, this contrasts sharply with the latest policy period, a time in which every successive year saw millions more US households adding broadband internet connections (Organisation for Economic Co-Operation and Development, 2008). By the mid-2000's, political activists of all stripes were emboldened by the newfound power of the internet to sidestep old media gateways and shape policy outcomes (Hewitt, 2005; Trippi, 2004). The limits of the general-purpose newspaper and the limited-audience print publication were beginning to dissolve as obstacles to reaching a broader issue public. This is due to the widespread adoption of the internet, which has become an important vehicle for expanding the scope of the conflict over DRM policy; this is the subject of the next chapter.

CHAPTER EIGHT: COMMUNICATING ONLINE

The internet is a new tool for the strong fair use coalition to use in their efforts to upset the decades-long dominance by those calling for ever-stronger copyright law. As predicted, this study found that the strong copyright coalition historically did quite well in Congress. The historical balance in Congress substantially favored the strong copyright coalition, and while this balance shifted toward the strong fair use coalition in the latest time period, this primarily meant that each side had the opportunity to slow down the other's legislative agenda; the sides fought to a draw. While strong copyright advocates did not dominate the pages of the *Times* and *Post*, they generally held their own. The historical balance in the newspapers was quite close to neutral, and while the trajectory was also in the direction of the strong fair use coalition, the low volume of coverage gave them slim hope for successfully shifting the public dialogue via these outlets alone.

The story up to this point is consistent with the general impression that well-resourced policy advocates tend to win the day over less-well-funded groups. Even under the most optimistically pro-fair use reading, the strong copyright coalition's remarkable policy strength of the past two decades was met by a burgeoning strong fair use coalition with, at most, the political strength to defeat attempts to pass legislation imposing additional affirmative requirements to implement specific DRM technologies (Benkler,

2006, pp. 411-412).[61] Thus, the fair use coalition had every reason to seek a new vehicle in their efforts to re-order the policy subsystem.

The internet is potentially a tool for the fair use coalition to increase its political leverage despite being vastly outspent by the strong copyright coalition. As the coalition that is struggling to upset the current policy order, the strong fair use side has every incentive to expand the scope of conflict to include the broader public in the discussion (Baumgartner & Jones, 1993; Schattschneider, 1960). The internet gives them the chance to do so with relatively small capital expenditures (Bimber, 2003).

The role the internet plays in political conversation is quite distinct from that of the major newspapers. As noted in Chapter 7, policymakers and their staffs often read the *Times* and *Post*; the same cannot be said of even the most highly visible copyright advocacy websites, but the internet still potentially serves several important roles. First, online communication is a cost-effective tool for under-resourced civil society groups, interested scholars, libraries, and other interested citizens to combine their efforts in building an issue network (Marres, 2006). This includes information sharing, message development, and agenda setting. This part of policy advocacy is often under-valued, but it represents the infrastructural backbone of more public forms of advocacy.

---

[61] Even the capacity to defeat such proposals is not automatic and must be directed with full force to stop any proposal with serious legislative support. At least one bill that contains such a requirement, more or less, has passed since the end of this study's time frame. H.R. 4137, the College Opportunity and Affordability Act of 2007, was signed into law on August 14, 2008. In it, § 493(a) adds 20 U.S.C. § 1094(a)(29), which requires in part that a federally funded institution of higher education certify that it "has developed plans to effectively combat the unauthorized distribution of copyrighted material, including through the use of a variety of technology-based deterrents."

As a second important role, online tools offer highly cost-effective ways to mobilize an issue public, at least under certain circumstances (Bennett & Manheim, 2001). Congressional staff may not read issue advocates' blogs and email listservs, but they do read emails and take calls from constituents. If there is an issue public to be mobilized around an issue, and if that population has the access and skills to use online communication tools, they can serve as a vehicle for getting a coalition's views heard in Congress, both directly via constituent contact and indirectly by legitimizing an ascendant coalition's message. The internet also allows that same issue public to do a good deal of one's research and writing (see, e.g., Benkler, 2006), blurring the line between intra-coalition message development and public mobilization in a way that builds issue ownership among those mobilized.

Viral online media can accelerate mobilization. A policy actor need not have a website with intrinsically high visibility; end users who are mobilized can forward emails, spread the message via social networking sites, and recommend a coalition's messages on media hub sites such as YouTube and aggregator sites like Slashdot and Digg. Once a coalition reaches a high enough number of sympathetic cyber-activists, these activists can spread the message, reaching many of those most likely to care about an issue with a wealth of specific information. This method of information circulation contrasts sharply with newspapers, which reach a larger audience but provide far less information, reach target audiences with almost zero precision, offer topical information much less often, and offer much less control over message to a policy coalition.

As a third important role for the internet, online advocacy can give some coalitions disproportionate power to shape the general public's perception of an issue. Search engine results and the nature of web browsing both favor websites with a high number of inlinks (Rogers, 2004). A link from one website to another is literally an invitation to visit the other site. From the perspective of the linked-to site, this is an inlink—an incoming link from another website. If one follows hyperlinks via browsing, one is more likely to wind up at sites with a high number of inlinks; compared to sites with few inlinks, those with many inlinks are easier to stumble into because more sites point visitors toward them. Following the pioneering PageRank technology developed by Google (Google, 2008d), most search engines now use inlinks to measure a site's relative authority. A site with more inlinks has more authority, and if 2 sites are otherwise equally relevant results for a given search term, the site with higher inlink authority appears higher in search results. I use the term "inlink authority" throughout this chapter to refer to a site's number of inlinks, as compared with other sites.

To the extent that citizens gain or would seek out an understanding of an issue via the internet, the importance of hyperlinks greatly favors the coalition whose hubs have the most inlinks. Many who create hyperlinks online are aware of this vote-by-link system and deliberately game the system to create search results that favor their worldview (Tatum, 2005). For relatively obscure policy topics, such as the regulation of digital rights management (DRM) technologies, the internet is likely to be the preferred means for most citizens to learn more due to its low cost for end users (Hindman, Tsioutsiouliklis, and Johnson, 2003). Thus, sites that have high inlink authority will have

248

a disproportionate chance to shape fresh opinions on an issue. Because this authority can be altered by the linking behavior of the thousands of interested people who have related websites, it gives an edge to those coalitions with large numbers of motivated web denizens relative to those coalitions with more financing but fewer sympathetic online voices. On the internet, an advantage in human capital is generally more important than an advantage in financial capital.

Fourth, online communication has the potential to shape offline news coverage. For instance, the agenda-setting relationship between political blogs and print media is complex and often bidirectional, suggesting that on at least some issues, newspapers rely in part on the web to determine when an issue merits coverage (Wallsten, 2007). Additionally, information-rich online policy advocacy provides a powerful, accessible information subsidy. Thanks to the recent, dramatic increase in journalists' workloads, journalists are increasingly dependent on external news sources for their content (Davis, 2002). As one journalist explains, "people are increasingly reliant on the wire services and Internet and other information coming to [them]" (Davis, 2002, p. 37)[62] rather than investigative journalism. This opens the door for policy actors who can deliver information to fill this vacuum.

---

[62] Davis conducted this interview in May of 1999. This means that, almost a decade ago, journalists were already regularly using the internet as a technique for cheap newsgathering. The trends that Davis identifies—reductions in newsroom budgets leading to rising workloads for journalists, and increasing investment in news subsidies by outside news sources—have continued on dramatic trajectories in the years since, with the internet playing a major role on both ends.

If certain policy actors have established themselves online as reliable, expert sources on an issue—if they stand out from the online crowd—a rich online information subsidy makes it easier for offline press to include favorable information and issue frames as developed online. This study does not test the direction of inter-media influence between online and offline media, but the real possibility of influence on offline media is another important reason that groups invest in online communication.

For all of these reasons, the internet holds a great deal of potential for those who have not successfully gained the upper hand in Congress and landed regular, friendly stories in the elite press; in this study, that means the strong fair use coalition. This chapter illustrates their complete and utter victory in the online debate over DRM policy.

First, I explore the results of 13 months of web crawls, discussing one web graph as an example, and describing the distributions of 3 key variables: number of relevant documents, inlink authority, and rhetorical valence. Next, I demonstrate that the sites that provide the highest number of relevant documents and have the highest inlink authority are overwhelmingly in the strong fair use coalition. Where the first two sections treat each website as the unit of analysis, the third and fourth sections are at the document level. The third section provides a basic description of the distribution of each of the relevant variables, and the fourth section looks for relationships between these variables.

Issue Crawler Results

This project's investigation of the online debate over DRM policy begins with the results from repeated tests using the Issue Crawler developed and described by Richard Rogers (Rogers, 2004), available at www.issuecrawler.net. Chapter 4 provides a fuller

description of this method. Every month for 13 months, from October 2006 to October 2007, the crawler started with five seed URLs, following the outbound hyperlinks from those sites to other sites. It then followed the outbound links from the newly discovered sites. At each step, a site was required to have incoming links from at least 2 other websites in the issue network to remain in the pool. The issue crawler can then display these results as a list of websites ordered by number of inlinks, or as a 2-dimensional map. While the maps themselves are not central to this project's analysis—the quantitative measures of inlinks are the result of interest—they do add explanatory power, so it is worth briefly describing what they convey and including an example. I do this first. Then, I describe the population of websites uncovered by these crawls.

Mapping the DRM Debate Online

On a web graph as rendered by the Issue Crawler, each website is represented as a circle; a site's circle is also called its node. There are three key variables that determine the representation of each node. First, a site's share of inlinks determines the area of the circle; a site with more inlinks will have a larger circle. The map also depicts arrows that identify which sites are linked to which,[63] and larger nodes therefore have many incoming arrows. Second, sites are color-coded according to their top-level domain; sites that end in .org will be one color, .com another, and so on.

---

[63] The native display of the maps is in the Scalable Vector Graphics (SVG) format. Within a browser window, one can toggle this mass of links on and off; with "links off," all arrows except those going into or out from a chosen site are hidden, and one can click on any site's node to see that site's inlinks and outlinks. All pictures here include all links, as no one site is the point of analysis, even though many individual arrows are hard to distinguish.

251

Third, sites will be placed in proximity to other sites based on co-link analysis; if they are commonly linked from other sites in the issue space, they will be placed close together. For example, suppose several pages each contain links to both the Motion Picture Association of America (MPAA) website and the Recording Industry Association of America (RIAA) website; Page A links to both sites, Page B links to both, and so on. This implies a relationship between those organizations, so the software will try to place them close together. If no pages contain links to both the RIAA and the University of Pennsylvania websites—Pages A and B link to the RIAA but not the University's website, and Pages C and D link to the University's site but not the RIAA—this implies no relationship, and these sites will be placed far apart. This is exactly what happens in Figure 8.1, which depicts an example map. [64]

---

[64] These graphs are included despite their less than ideal quality. For instance, the linking arrows are almost unacceptably faint. This was almost certainly done in order to make the site labels legible. As described below, site nodes and labels are also on top of each other in tightly clustered areas. Reservations aside, the maps convey enough information to merit their inclusion. Graphs are presented with a landscape orientation to maximize their resolution.

   To view digital versions of all 13 maps, including both picture and scalable vector graphics (SVG) versions, please visit the author's home page: http://www.billyherman.com

Figure 8.1: Web Graph, Online Copyright Debate, November 2006

253
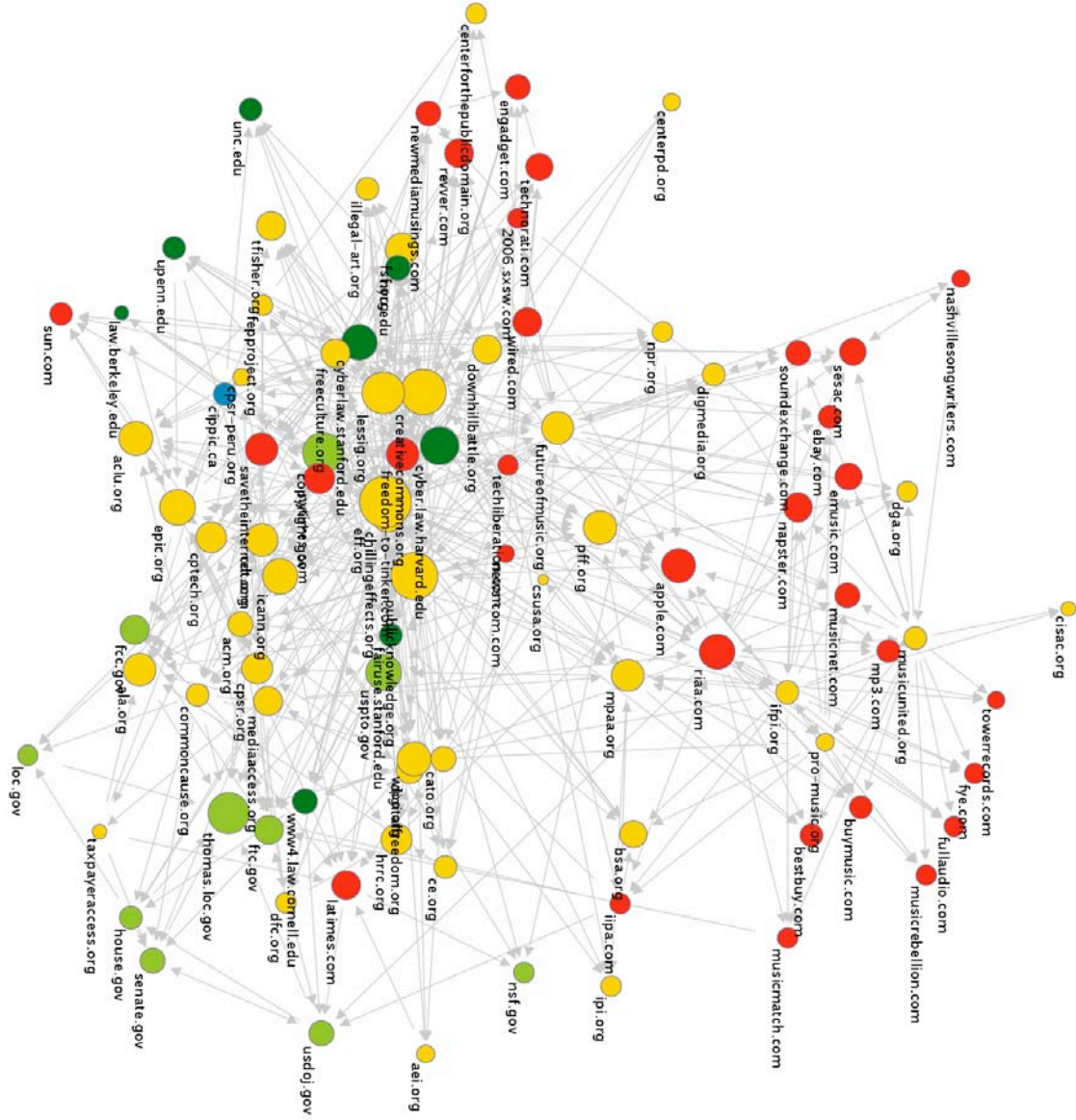
This map represents all 95 sites from the web crawl from November 2006; this is the entire issue network. It represents the strong fair use coalition and the strong copyright coalition, and it even represents the substantial separation between the two groups. Site nodes break fairly cleanly into 2 groups: the strong fair use group on the bottom, and the strong copyright group on the top.

The strong fair use coalition dominates this map, anchored by the group of closely linked large nodes that are clustered just to the bottom-left of the graph's center of the graph. The patterns of links pointing into these sites strongly suggest that they are closely related. The high number of sites and the tightness of their connections are responsible for the poor readability in that region of the map; site labels and linking arrows are on top of one another. There are several peripheral sites orbiting around this gravitational center, located to the left, right, and bottom. Sites ending in .org, roughly 30 in total, dominate this entire area. Starting with the Copyright Office site in the center of this cluster and spreading rightward, there are 10 sites ending in .gov. A total of 8 .edu sites are scattered across this area, as are 13 .com sites.

Between the strong fair use cluster in the lower portion of the graph and the strong copyright coalition in the upper portion is a relatively large gap with few sites. This suggests momentous differences of opinion between these groups of organizations; linking websites effectively declared that there are two kinds of organizations, and this is the space between them. In the upper portion of the graph, there are fewer sites with smaller nodes, and they are less densely linked. The 18 .com sites make up a majority of the nodes, and 8 to 11 .org sites (depending on which sites one includes) make up the

254

balance. The number of .com sites is proportionally much higher than in the bottom portion of the map; in the strong fair use region, .com sites represent less than 25% of sites, whereas in the strong copyright region, they represent at least 60% of sites.

While the numbers examined in full detail below illustrate this definitively, even a first-impression glance at this map correctly suggests that the strong fair use coalition has more websites in this network, and the most linked-to sites are in the strong fair use coalition. The strong fair use coalition is better represented and much more heavily interlinked.

One important site that does not fit neatly in either cluster is the Progress and Freedom Foundation (PFF) site, pff.org. This was included as one of the five seed websites that formed the basis for the scheduled crawls.[65] The Foundation, which makes rhetorical ties between copyright and right wing support for strong property rights, is the only civil society group[66] that is both an important member of this debate and a reliable supporter of strong regulation of DRM technologies. They maintain a blog that is updated frequently by many authors. Thus, while the PFF is part of the strong copyright coalition, this map correctly suggests that it is different from the other organizations in the coalition. Unlike the industry lobbying groups like the RIAA, the PFF is a deliberate and engaged participant in the online debate.

---

[65] As discussed in Chapter 4, the other 4 seed URLs were sites for Public Knowledge, the Consumer Electronics Association, the RIAA, and the Copyright Office.
[66] As discussed in Chapter 6, the NGO category does not include nonprofit groups created to advocate for one specific industry. I consider "civil society group" to be synonymous with "NGO".

Figures 8.2 and 8.3 are zoomed-in shots of the same graph. Figure 8.2 shows the densely connected core of websites at the center of the strong fair use coalition. Sites for NGOs like the Electronic Frontier Foundation (eff.org) and Public Knowledge (publicknowledge.org) make up the gravitational center of the entire graph. Linking patterns correctly suggest that these central sites are closely related.

The arrows illustrate so much linking into and between the strong fair use sites that one cannot identify a relationship between two specific sites. These sites are heavily interlinked around the topics of interest. This reflects substantial investment in building an online issue network, including information sharing, message development, and agenda setting. This also suggests a highly disproportionate power over the online representation of an issue. There are far more strong fair use sites, and they link to each other far more often than they link to strong copyright sites. Thus, an end user investigating the topic is more likely to stumble into one of these sites and, from that starting point, far more likely to follow links to the other sites in this cluster than to the sites in the strong copyright coalition. While this does not *per se* measure activist mobilization, it does strongly suggest that, to the extent that the internet serves as a vehicle for such mobilization, the strong fair use coalition is far more likely to succeed on this count.

Figure 8.3 highlights the sites in the strong copyright coalition. Moving up and to the right from the PFF website, the first cluster of sites belongs to Apple (apple.com), the Motion Picture Association of America (mpaa.org), and the Recording Industry Association of America (riaa.com). After these, the next-closest nodes are the sites for the

256

International Federation of the Phonographic Industry (ifpi.org) and the Business

Software Alliance (bsa.org). Except for Apple, these are major organizations created to

advance the interests of copyright holders. The RIAA and MPAA in particular are very

close allies, and each group's node is the closest neighbor for the other on the map. From

this cluster, the arrows point almost universally away from the strong fair use coalition.

They point at other allied groups, such as the Director's Guild of America (dga.org), but

more links point at a large number of commercial sites for the legal purchase of media,

such as Best Buy (bestbuy.com) and the new Napster (napster.com); these account for

most of the .com sites in the cluster. By pointing at sites where one can buy media, the

advocacy groups in the strong copyright coalition are effectively saying, "Stop stealing

our work and buy your music, movies, and software legally!"

Figure 8.2: Web Graph, Online Copyright Debate, November 2006, Strong Fair Use Core

258
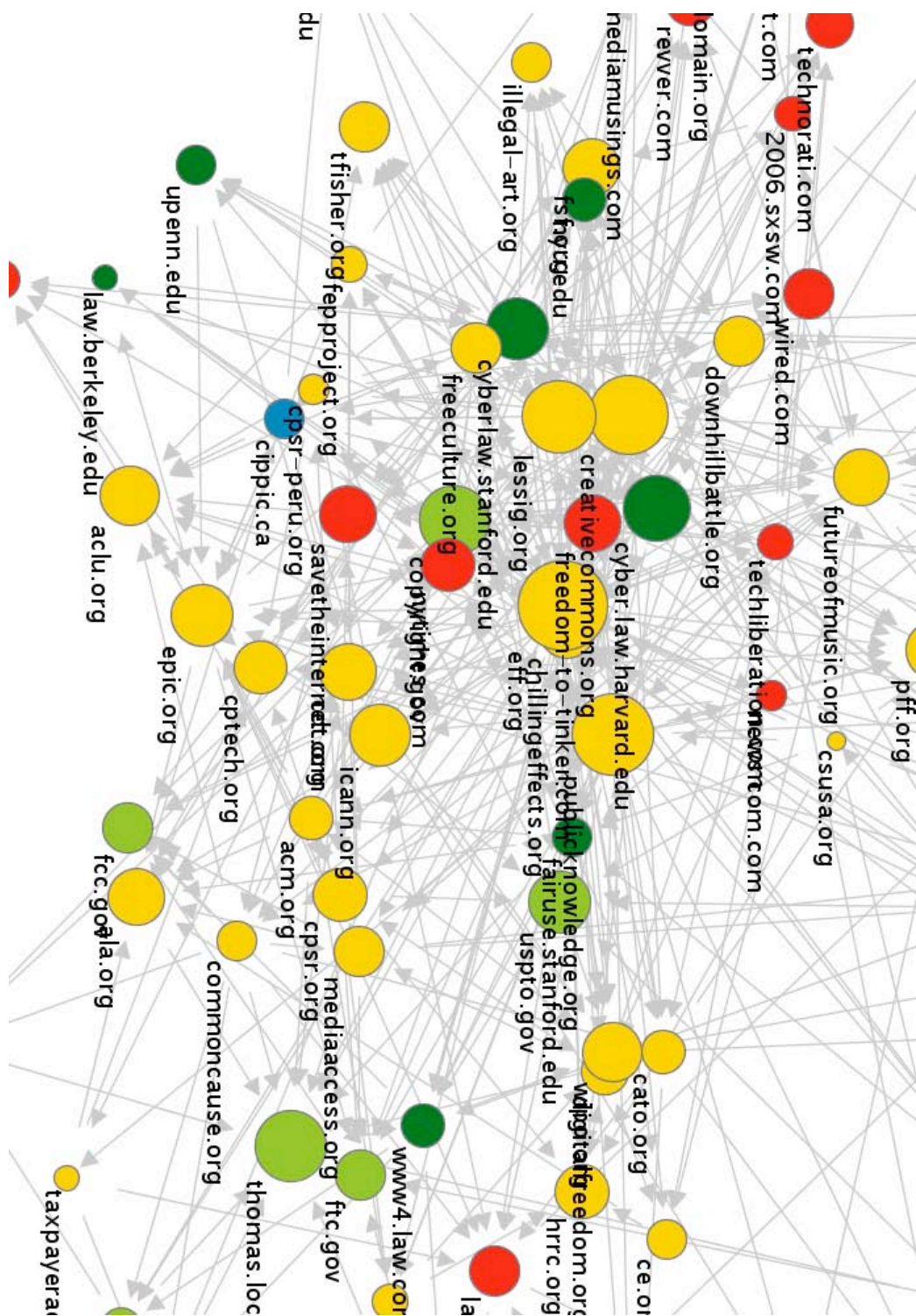
Figure 8.3: Web Graph, Online Copyright Debate, November 2006, Strong Copyright Cluster

259
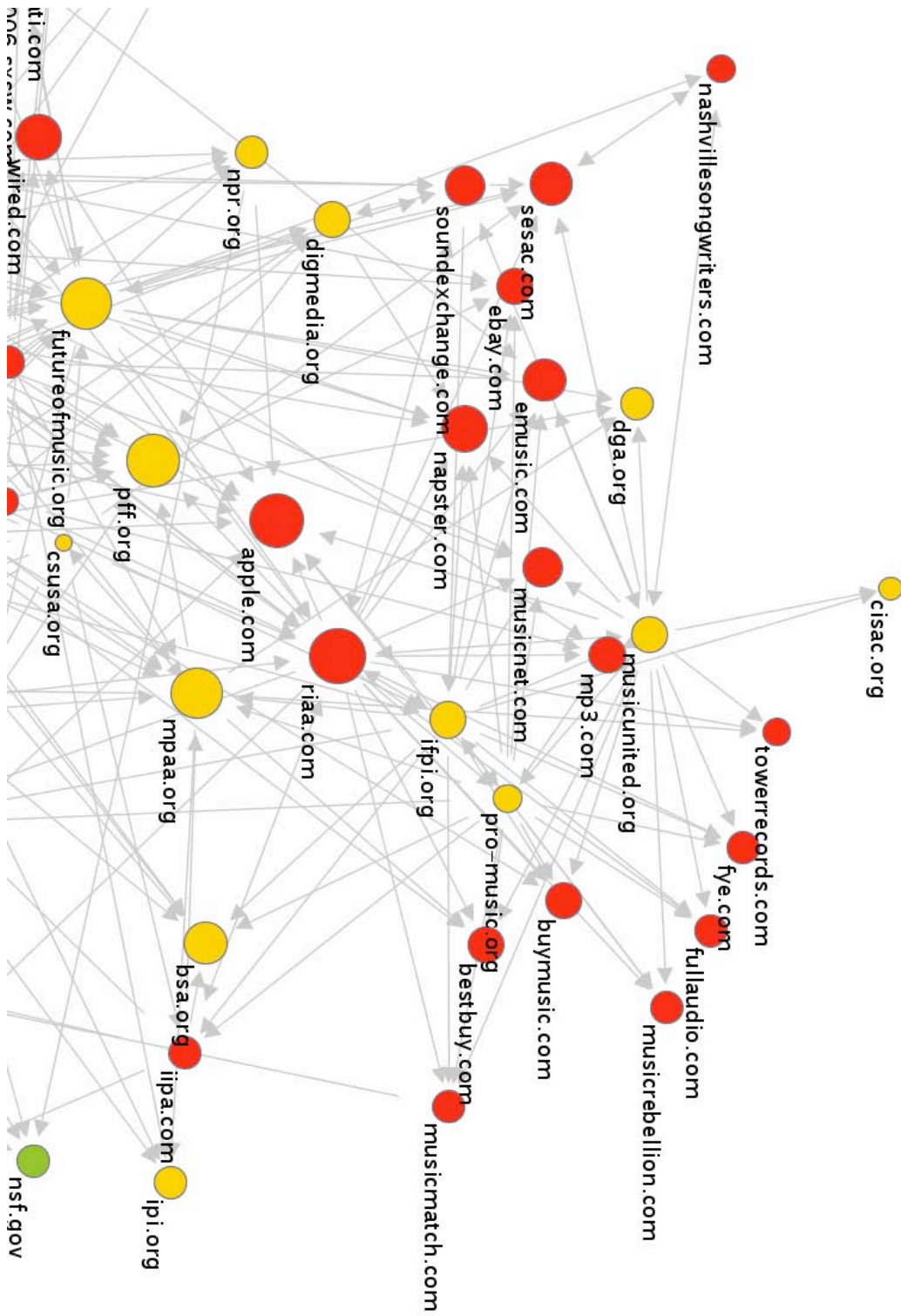
These maps provide a powerful visualization of the online debate over copyright

policy. They are useful for providing an overview of the relative status of various actors

and the relationships between them, though without further explanation, they are

primarily of use to those familiar with the policy actors involved. This familiarity can be

acquired and translated into verifiable results through the use of content analysis of

online documents. Additionally, while these maps are based on quantitative measures of

centrality and inlink authority, the underlying data are available and are more appropriate

for testing hypotheses. For this project, inlink authority is the variable of concern. The

other site-level variables of interest—number of relevant documents and mean rhetorical

valence—are aggregate results from the content analysis of online documents.

Quantifying the DRM Debate Online

In addition to the issue network map, each crawl can also be represented as a list

of websites; the number of total websites included in the results of a given crawl ranged

from 86 to 95. Over the 13 crawls, 210 sites appeared in the final results at least once,

and 78 appeared at least 7 times—that is, in a majority of crawls. Since this study is

concerned with those sites that are regular participants in the online debate over DRM,

this chapter focuses on those 78 sites that appeared in most crawls.

Using the results over multiple crawls, these sites can be ranked in a number of

ways. The crawler reports results based on the total number of inlinks from other pages in

the population, ranking each site by inlinks for that crawl. The results for any given crawl

are fairly definitive, but sorting sites based on the net results of a year of crawls requires

some judgment. My first instinct was to rank sites based on their mean rank for each

crawl. For instance, the Public Knowledge website, publicknowledge.org, was roughly

the 16th most-linked site in any given crawl; their site was as high as 10th (January 27,

2007) and as low as 21st (November 27, 2007), but the mean and median rank was 16th.

This is a reliable presence in the upper part of the rankings. Most of the other sites above

them in any given crawl were less reliably near the top. For instance, the Consumer

Project on Technology site, cptech.org, was more often than not ahead of Public

Knowledge; in the 8 crawls from January to August 2007, the site was ranked 7th, 8th, or

9th. In the other 5 crawls, however, the site's ranks were 31st, 57th, 44th (October through

December 2006), 13th, and 68th (September and October 2007). Thus, the Consumer

Project on Technology was behind Public Knowledge based on mean rank (21.31) but

ahead based on median rank (8).

Ordering sites based on their mean or median rank comes with one important

disadvantage: the loss of a tremendous amount of data. The difference between 1st and 2nd

is unlikely to be the same as the difference between the 39th and 40th. Quite the contrary,

other research has already demonstrated that the distribution of links between websites

follows "a 'winners-take-all' power-law distribution, where a few successful sites receive

the bulk of online traffic" (M. Hindman et al., 2003, p. 1).

As operationalized in the context of the Issue Crawler results, this means that one

would expect the top-ranked site to have substantially more inlinks than even the 5th or

10th site, and that the difference between the 1st and 2nd most inlinks will be more

substantial than the difference between the 9th and 10th-ranked sites, and much more

substantial than the difference between 39th and 40th. This was certainly true of the results

261

from most of the crawls. Despite the relatively small number of sites—unlike the broader web, it is realistic to imagine 90 sites linking to each other with relative equity—the linking patterns reflected a winners-take-all distribution. The median ratio of inlinks to the top-ranked versus the second-ranked site was 1.43. For the 5th- and 10th-ranked sites in a given crawl, the median ratio was 2.72 and 3.00 times as many links directed at the top site. For the 40th-ranked site, the median ratio was 49.79; in a typical crawl, the top site had almost 50 times as many inlinks as the 40th most-linked-to site. In contrast, the median ratio of the 40th-ranked site's inlinks to those for the 80th-ranked site was 4.17. The typical difference between 1st place and 40th place was over 10 times greater than the typical difference between 40th and 80th. This is to say nothing of the hundreds of sites that link into this population but were not included because they never had enough inlinks to be included in the final population. For instance, compared to the blog written by this author and several of his colleagues,[67] shoutingloudly.com, the top-ranked site in any given crawl was almost infinitely more visible.

Faced with such a clear power law distribution, it is unacceptable simply to rank sites and use these ranks for all tests, which would impose linearity on a highly non-linear distribution. Thus, I use a measure based on share of inlinks in a given crawl. Using proportional share also eliminates the problem of unbelievably large differences in the total number of inlinks in any given crawl; totals ranged from 6,810 to 130,470. Finally, I sought to eliminate what I would describe as the "USA.gov problem" in measuring inlink

---

[67] The other authors are, in alphabetical order, David Karpf (Ph.D. student, University of Pennsylvania Department of Political Science), Jason Tocci, and Lokman Tsui (Ph.D. candidates, Annenberg School for Communication, University of Pennsylvania).

counts. The site is the official web portal of the US federal government, hardly a hotbed of copyright policy activism. Yet in 2 crawls, it had incredibly high proportions of total inlinks: .589 (March 11, 2007) and .322 (October 23, 2007).[68] These results were hardly typical. Its third-highest share of inlinks was a much more modest .055. The site did not even appear in the results of 6 of 13 crawls; it was thus barely included in the final group of 78 websites. While the site included no relevant documents and thus had no impact on calculations testing the relative number of documents or inlink share of each coalition, this fluke strongly suggests the need to adjust for such chance fluctuations, even though no other site had fluctuations anywhere near this size. I therefore chose to drop the top two and bottom two figures for each site, measuring each site's inlink authority using the mean inlink share from the remaining nine crawls. For USA.gov, for instance, I dropped the figures of .589 and .322 on the high end and two of the six crawls in which the site had zero share;[69] this resulted in a mean inlink share of .021, good enough for 13th place overall.

---

[68] Without looking into the link-by-link results—an untenably labor-intensive process— or programming additional tools for analyzing such questions, I am unable to definitively state why this occurred. I can only assume that an unusually high number of government websites that link to USA.gov were measured in these crawls.

[69] This is similar to other systems for reducing the effect of unrepresentative fluctuations in ranking systems. For instance, in competitive events judged by humans, from diving to individual rankings in competitive team debate, the highest and lowest scores for any competitor may be dropped. This reduces the potential effect of bias by any given judge. Likewise, dropping each site's highest and lowest two rankings ensures that the estimate of a site's inlink share is a fair representation of their presence across the entire period and is not unduly influenced by the results of any one crawl.

The USA.gov case in particular illustrates the need to approach any given Issue Crawler result with a bit of caution. The results here suggest that one can use the results

While this did indeed solve the problem of chance fluctuations, it created an

additional minor problem: the sum of all the mean inlink shares created with this formula

for the 78 included sites was just .780.[70] I thus multiplied every site's mean inlink share

by the inverse of .780, or 1.281, to reach an adjusted mean inlink share; the sum of all

such shares among included sites is 1, rendering the results more amenable to intuitive

interpretation. Table 8.1 provides a list of the top ten websites according to this adjusted

mean inlink share—henceforth, simply referred to as inlink share. As one would predict

based on the literature on inlink share among websites generally and political websites

specifically (M. Hindman et al., 2003, p. 1), the top sites in this population earned a very

high share of the inlinks: 62.6 percent. The top site, Creative Commons, had an inlink

share of 14.3 percent.

---

of multiple crawls over time, especially when backed by validation of seed URLs by
comparison with crawls started with other URLs in the same network.

[70] This is almost purely an artifact of removing the top 2 scores for each included site. For
instance, the sum of the 2 remarkably large values for USA.gov, divided by 11, equals
.083. This is a substantial portion of the .220 gap between the sum of shares, .780, and
the expected inlink share among the dominant sites—which, in a power law system,
should approach 1. Regardless, setting the sum of all shares to 1 has no effect on the
relative ranking of sites or on the ratio between any 2 sites.

Table 8.1: Inlink Share Among Top Ten Sites

Rank	Organization/Person	URL	Adjusted Mean Inlink Share
1	Creative Commons	creativecommons.org	.143
2	Electronic Frontier Foundation	eff.org	.094
3	Free Software Foundation	fsf.org	.088
4	Lawrence Lessig (Stanford Law School) et al.	lessig.org	.058
5	Center for Democracy & Technology	cdt.org	.056
6	Computer Professionals for Social Responsibility (Peru chapter)	cpsr-peru.org	.044
7	Consumer Project on Technology	cptech.org	.041
8	Future of Music Coalition	futureofmusic.org	.036
9	Media Access Project	mediaaccess.org	.034
10	The Fair Use Network	fairusenetwork.org	.031
	Total Adjusted Mean Inlink Share		.626

Plotting these points on a scatterplot, with each site's mean inlink share plotted along the Y-axis and its rank by mean inlink share along the X-axis, highlights this power law distribution. Figure 8.4 does just this.

265

Figure 8.4: Mean Inlink Share by Mean Inlink Share Rank
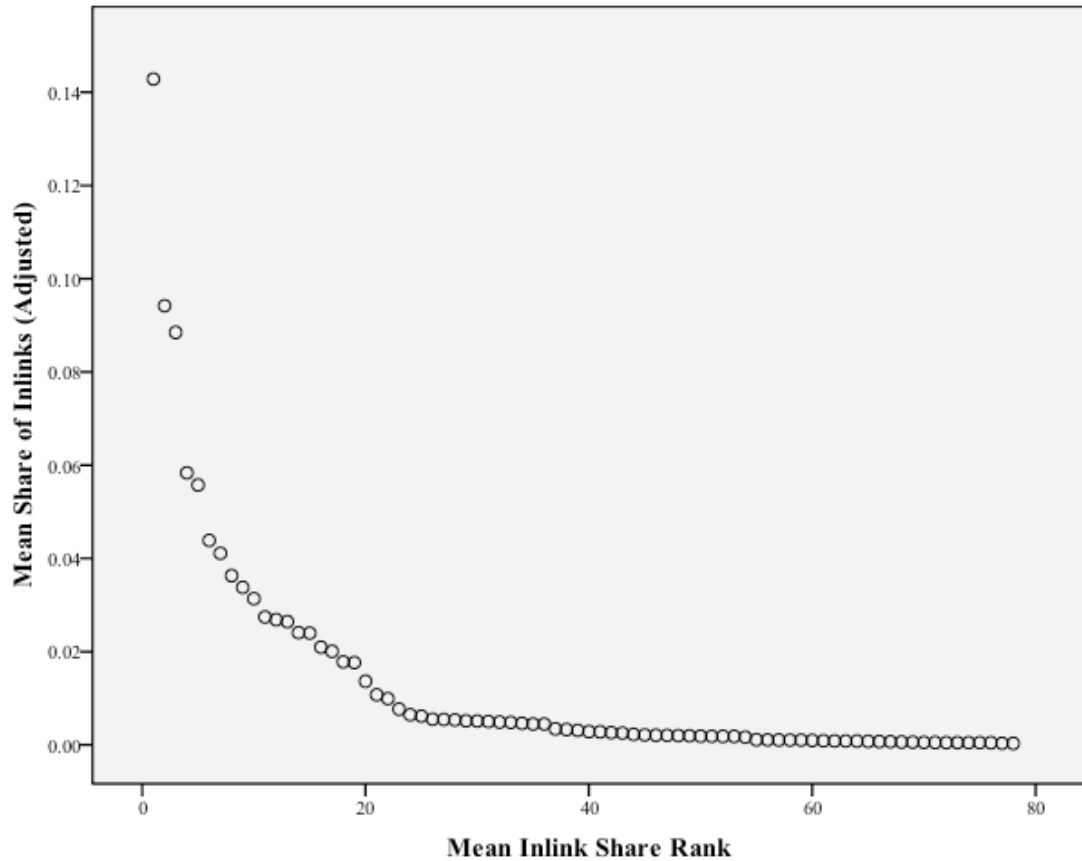
The top-ranked site, to the far left, is the only data point above .14. The 2nd- and 3rd-ranked sites hover near .09, part of a very steep drop among the top 10. The slope becomes far less steep as one approaches the bottom of the rankings; after roughly the 20th-ranked site, the difference between any two adjoining sites becomes negligible. This is a clear power law distribution.

Two other key variables, the number of relevant documents per website and each websites' mean rhetorical valence score, also featured highly non-normal distributions. First, consider the number of relevant documents per website. Of 78 that were returned in

a majority of crawls, 26 of them—exactly one third—had zero relevant documents; another 11 had just 1 document. On the opposite end of the spectrum, the Electronic Frontier Foundation site (eff.org) had 81 relevant documents,[71] Public Knowledge (publicknowledge.org) had 74, and four more sites (wired.com, berkeley.edu, freedom-to-tinker.com, and ala.org) had more than 50.[72] This distribution is reflected in Figure 8.5.[73]

---

[71] This figure would have been slightly higher were it not for the methodological constraints described in Chapter 4. The US House of Representatives website turned out to set good benchmarks for approximate maximums for both searches: 39 documents for the DMCA reform debate, and 42 for the broadcast flag debate. Just one site exceeded the combined total for both topics; thus, this was a reasonable cap to set.

[72] House.gov is not included in this list of websites with the most relevant documents. For the purpose of focusing on the online debate *as differentiated from* the congressional debate, I only included documents that were not online reproductions of hearing documents. This left 6 relevant documents, and these had a mean valence of 2.472.

[73] Relative to Figure 8.4, this histogram is inverted, with the top-ranked sites stretching rightward rather than upward. Were one to rank sites by number of relevant documents and produce a scatterplot, with number of relevant documents along the Y axis, the overall shape would be much like Figure 8.4.

Figure 8.5: Number of Relevant Documents per Site
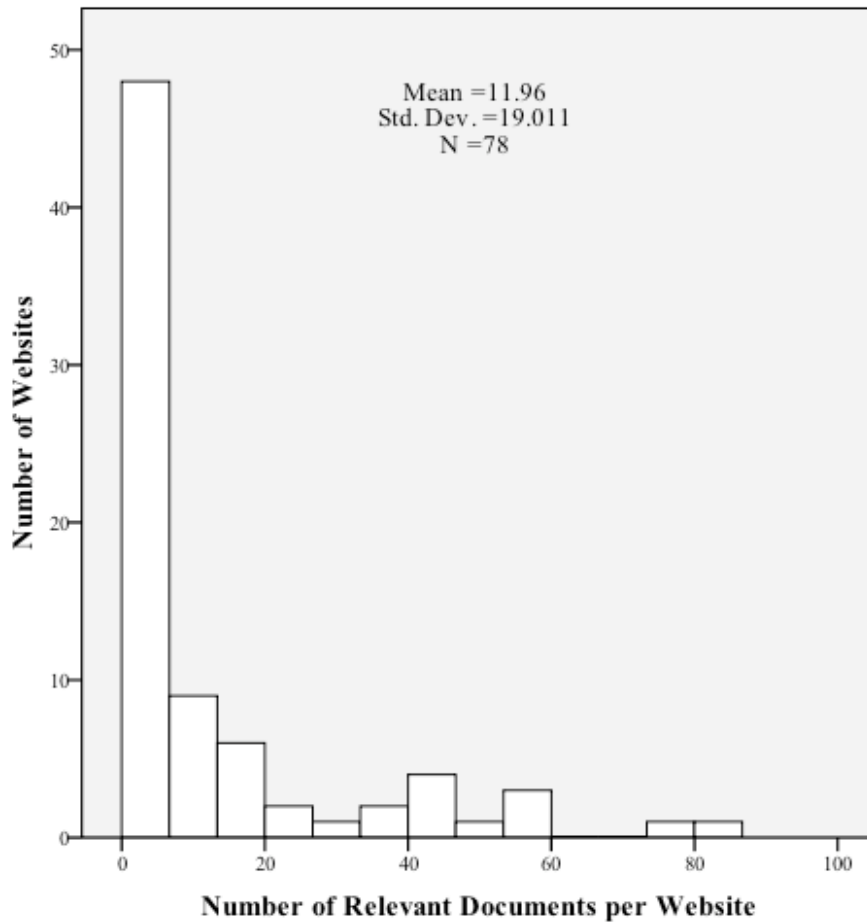
The distribution of valence scores was also highly non-normal. It was also distinctly different from the "field goals" distribution of valence in Congress, featuring a remarkably one-sided distribution favoring the strong fair use coalition. The mean (2.54), median (2.98), and mode (3) scores leaned heavily toward the strong fair use coalition. As Figure 8.6 shows, the most common mean score for a website is far and away 3, representing unconditional support for the strong fair use position; 26 websites had this score, which was exactly half of the 52 websites that featured at least one relevant

268

document. Another 10 websites had scores of at least 2.7, and 5 had scores between 2 and 2.5. That leaves just 11 websites with scores below 2, or in favor of stronger copyright. Of these sites, 6 had a score of 1, but only 2 of these (mpaa.org, with 14 documents, and ipi.org, with 7) had more than 2 relevant documents; the other 4 scores are thus not necessarily signs of strong policy views by those sites' sponsors. In stark contrast, among the 36 sites with a mean valence score of 2.8 or greater, 22 sites had at least 10 relevant documents; it is therefore much more reasonable to assume that these sites' sponsors stand solidly behind the strong fair use beliefs advanced on their sites.

Figure 8.6: Distribution of Websites' Mean Valence Scores[74]

Mean =2.54
Std. Dev. =0.712
N =52

Number of Websites (y-axis)

Mean Valence of Documents on a Website

1 = Strong Copyright, 3 = Strong Fair Use

---
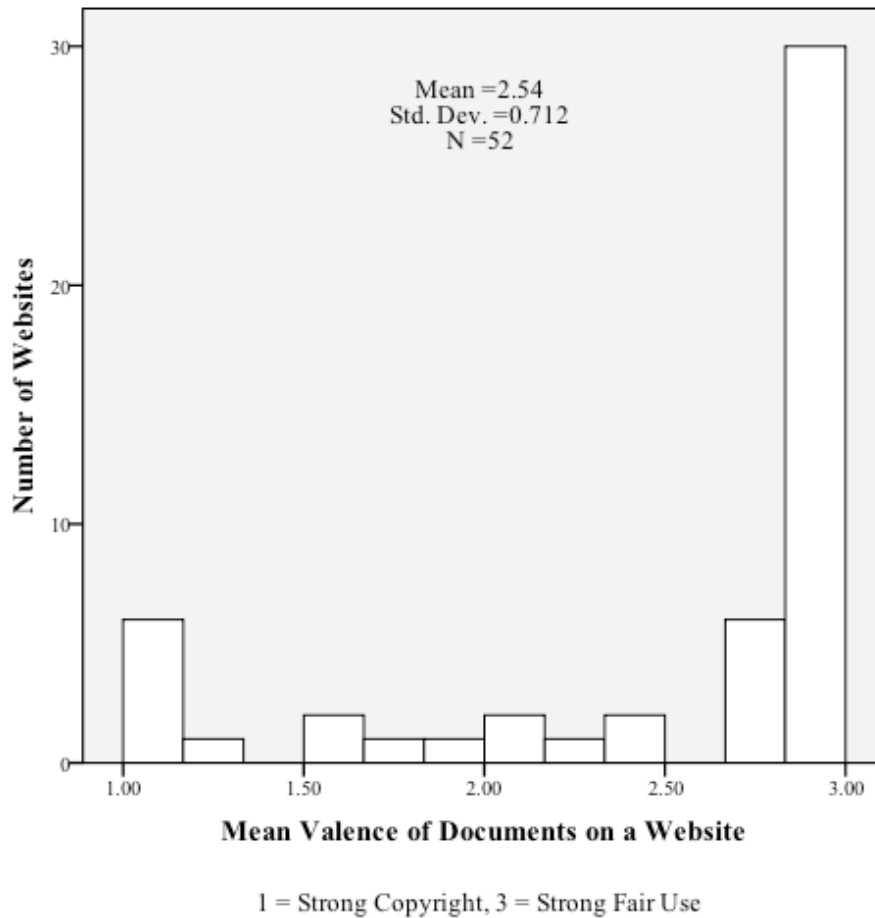
[74] The attentive reader may notice a nontrivial error in this graph; even though the maximum value for a valence score is 3, SPSS 16 for Macintosh graphs the bulk of websites as having a mean greater than 3. This is regrettable, though the graph still conveys the overall distribution with reasonable fidelity. On a personal note, I have personally found SPSS 16 to have been a terrible piece of software and the company exceptionally unhelpful.

270

This distribution of sites' mean valence strongly suggests that the strong fair use coalition was winning the online debate—out of sheer numbers, if nothing else. Nearly 79 percent of the regularly included sites that voiced an opinion supported the strong fair use side, versus 21 percent supporting the strong copyright position. To formalize this conclusion, however, one must consider the relationships between these variables.

Dominance by Strong Fair Use Sites

The sheer number of strong fair use sites suggests online dominance, but the relationships between valence and 2 other variables—number of relevant documents and inlink authority—provide quite strong evidence of this dominance. This section considers these relationships in that order.

Strong Fair Use Websites Post More Relevant Information

Hypothesis 6 predicts that, on average per site, strong fair use actors will provide more information pertinent to ongoing policy debates than will strong copyright actors. This is certainly consistent with the data in this study. On the 41 sites in the strong fair use coalition, the mean number of relevant documents was 20.27 (SD = 22.1), while the 11 strong copyright sites provided a mean of just 9.27 relevant documents (SD = 12.9).[75]

---

[75] The Levene's Test shows the variances are significantly non-equal: F = 5.372, p = .025. Without assuming that variances are equal, significance testing finds the difference between these means are unlikely to be due to chance: p = .043. As these sites represent an effective census of regularly participating websites rather than a random sample, a probability higher than .05 would not preclude one from accepting this difference as meaningful. A finding of statistical significance based on such a small number of sites, however, highlights how very different these means really are.

The average site in the strong fair use coalition provided over twice as many relevant documents as the average site in the strong copyright coalition. This is a medium effect size; *d* equals .534. The strong fair use coalition sites also had a higher mean number of documents: 10 versus 6 for the strong copyright coalition. These numbers were both relatively and absolutely closer than the difference in means because of substantial upward skewing by those sites that provided a deluge of relevant information.

The relationship between number of documents and websites' mean valence is nonlinear, but a graphical depiction helps illustrate how the strong fair use coalition website is far more likely to provide a high number of relevant documents. See Figure 8.7.

Figure 8.7: Number of Relevant Documents per Website by Site Valence

1 = Strong Copyright, 3 = Strong Fair Use
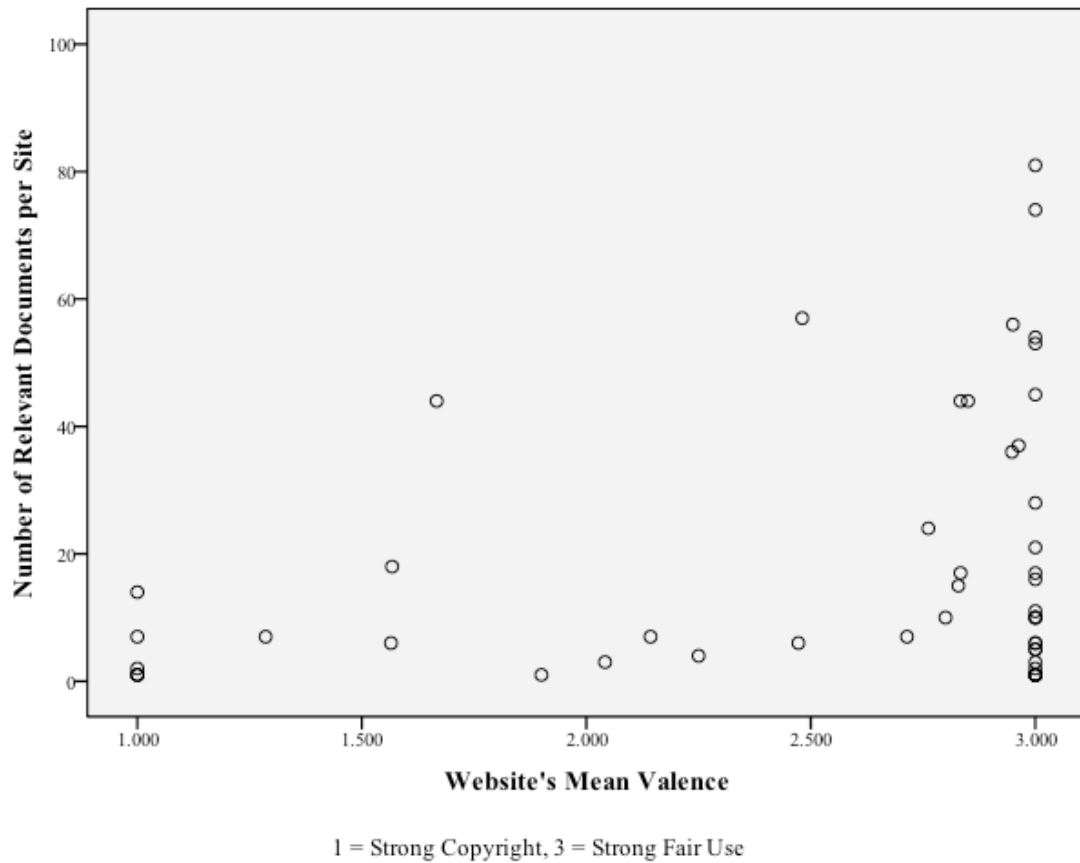
The sites in the strong fair use coalition are highly varied in the number of relevant documents, but this contrasts sharply with sites in the strong copyright coalition, which offer an almost uniformly low number of relevant documents. Except for the Copyright Office site, the graph features a distinct ramp-like shape, sloping upward once one is squarely within the strong fair use coalition.

The strong fair use coalition dominates the top 10 sites, ranked by number of relevant documents. Table 8.2 lists these sites, providing the mean rhetorical valence for each.

Table 8.2: Top Ten Sites for Number of Relevant Documents

Rank	Organization / Person	URL	Number of Relevant Documents	Mean Rhetorical Valence
1	Electronic Frontier Foundation	eff.org	81	3
2	Public Knowledge	publicknowledge.org	74	3
3	Wired	wired.com	57	2.48
4	University of Berkeley	berkeley.edu	56	2.95
5	Edward Felten (Princeton University) et al.	freedom-to-tinker.com	54	3
6	American Library Association	ala.org	53	3
7	Berkman Center for Internet & Society at Harvard University	cyber.law.harvard. edu	45	3
8	Association for Computing Machinery	acm.org	44	2.85
9	Center for Democracy and Technology	cdt.org	44	2.83
10	US Copyright Office	copyright.gov	44	1.67
	Total / Overall Mean		552	2.78

Of these sites, 9 were clearly in the strong fair use coalition, while only the Copyright Office website was in the strong copyright coalition. All 9 of the strong copyright sites were more strongly in the strong fair use coalition than the Copyright Office site was in the strong copyright coalition—a minimum divergence from 2 of .48, versus .33 for the Copyright Office. A high number of the relevant documents on the Copyright Office website that were actually authored by other people or groups and submitted as part of a rulemaking process (see Herman & Gandy, 2006), making the site

a venue for debate rather than simply a megaphone for the Copyright Office's views.[76]

These 10 sites provided 552 of the 933 total relevant documents, or 59 percent of the

total. If one is interested in learning about this debate, one could rely on any of these sites

to provide a very rich understanding, and in most cases, that wealth of information would

be sharply in favor of stronger fair use.

Considering all 78 sites, the strong fair use coalition provided many more relevant

documents than the strong copyright coalition. In total, strong fair use websites provided

831 out of 933 relevant documents, or 89 percent. This is an even higher share for the

strong fair use coalition than their share of included websites, which was 79 percent.

Strong copyright websites provided just 102 documents, or 11 percent.

---

[76] As with the EFF on both DMCA reform and the broadcast flag, the total number of DMCA reform-related documents on the Copyright Office site was higher than the 39-document cap used. As Herman and Gandy (2006) document, the majority of these documents, submitted during the triennial rulemakings to determine exemptions to the DMCA's anti-circumvention provision, were actually in favor of stronger fair use. This mean valence score is therefore almost certainly an artifact of the methodological limits of this study—specifically, the strategy for coding each website's valence based on only the first 10 documents. For nearly all sites in the population, this worked as expected; most sites do not represent as diverse an array of voices as does copyright.gov. In this case, documents authored by the Copyright Office, the Library of Congress, and allied government agencies made up 4 of the 5 broadcast flag-relevant documents and 4 of the 10 DMCA reform-related documents, giving them a slim majority (8) of the 15 relevant documents coded. Another 2 documents in the DMCA results also supported the strong copyright position, pulling the site's mean valence score a bit further in the direction of strong copyright.

A complete coding of all relevant documents on the site would likely have found it to be another site dominated by documents calling for strong fair use. Its inclusion in the strong copyright camp is thus likely a methodological artifact, albeit one that accurately suggests the Copyright Office's allegiances. Because this study concludes that most of the online documents and websites were in the strong fair use coalition, this methodological artifact has a conservative and thus acceptable effect on the conclusions drawn here.

This is an advantage that is too solid even to be subject to concerns about the relatively high error rates for the Google-based document retrieval strategy. As documented in Chapter 4, the estimated search recall was .857 for the broadcast flag and .826 for the DMCA reform proposals. One could assume the lower search recall rate for all searches, use it to estimate the number of documents not retrieved for the strong copyright coalition only, and still see a substantial advantage for the strong fair use coalition. Using this strategy, one would divide the number of relevant documents from the strong copyright coalition, 102, by the recall correction figure of .826. This suggests that a search strategy with perfect recall may have found as many as 123 relevant documents from the sites in the strong copyright coalition.[77] Relative to the slightly larger total of 954 documents, one would still estimate that the strong fair use coalition's websites had hosted 87 percent of relevant documents, versus 13 percent for the strong copyright coalition. This is still an astounding imbalance, and it is still substantially greater than the imbalance in the total number of websites. Under this estimate, the strong copyright coalition's mean number of documents per site, 11.23, is still far fewer than the original estimate of 20.27 documents per strong fair use website. This is all true even if one assumes the lowest estimate for recall for all strong copyright websites and *perfect* recall for all strong fair use websites, the most conservative means possible for ensuring against a false positive.

---

[77] Again, this substantially understates the estimate of documents not retrieved from the Copyright Office website, but most of the remaining documents supported calls for stronger fair use.

The strong fair use coalition provided many more relevant documents online, both in absolute terms and on average per site. Fully 9 of the 10 sites that provided the most relevant documents were squarely in the strong fair use camp. The per-site difference is remarkable, and this result holds up to even the most conservative possible estimates of accuracy. In short, these data could hardly be more conclusively in support of Hypothesis 6; on average per site, strong fair use actors really did provide more information pertinent to ongoing policy debates than did strong copyright actors.

Strong Fair Use Websites Have Higher Shares of Inlinks

The websites in the strong fair use coalition dominated the top of the list of websites as ranked by inlink share. Table 8.3 reproduces the list of the top 10 websites that first appeared in Table 8.1, except it also includes each site's mean valence.

The strong fair use coalition has the top 9 sites, and the 10th-place site, belonging to the Fair Use Network, likely would have been included had it hosted any relevant documents. With such domination of the top of the ranks, and with such a strong power law distribution, it is already safe to conclude that the strong fair use coalition won the quest for inlink authority; nonetheless, further tests are worth reporting, and they confirm this finding. For the strong fair use coalition, the mean inlink share is .0207 (SD = .0303), while it is just .0036 (SD = .0030) for the strong copyright coalition. This is a medium to large effect size; d equals .627. The median is also higher for the strong fair use coalition—.0062 versus .0028—if to a more understated degree than the difference between means. Summing the inlink shares of all the sites in each coalition also makes the point dramatically; 84.7 percent of links pointed to strong fair use sites, 11.3 percent

pointed at sites without relevant documents, and just 4.0 percent pointed to sites in the strong copyright coalition. Following any given link in this set is more than *forty times* more likely to lead to a strong fair use site than a strong copyright site. These findings strongly support Hypothesis 7; the websites of strong fair use actors had a higher number of total and average inlinks from within the online copyright policy space than those of strong copyright actors.

Table 8.3: Inlink Share Among Top Ten Sites, with Mean Valence

Rank	Organization / Person	URL	Adjusted Mean Inlink Share	Mean Rhetorical Valence
1	Creative Commons	creativecommons.org	.143	3
2	Electronic Frontier Foundation	eff.org	.094	3
3	Free Software Foundation	fsf.org	.088	3
4	Lawrence Lessig (Stanford Law School) et al.	lessig.org	.058	3
5	Center for Democracy & Technology	cdt.org	.056	2.83
6	Computer Professionals for Social Responsibility (Peru chapter)	cpsr-peru.org	.044	3
7	Consumer Project on Technology	cptech.org	.041	3
8	Future of Music Coalition	futureofmusic.org	.036	2.25
9	Media Access Project	mediaaccess.org	.034	3
10	The Fair Use Network	fairusenetwork.org	.031	n/a
	Total Adjusted Mean Inlink Share		.626	2.90

Plotting these data points on a graph, one again sees a ramp-shaped distribution that points sharply upward at the strong fair use end of the scale. See Figure 8.8.

Figure 8.8: Websites' Inlink Share by Site Valence

1 = Strong Copyright, 3 = Strong Fair Use
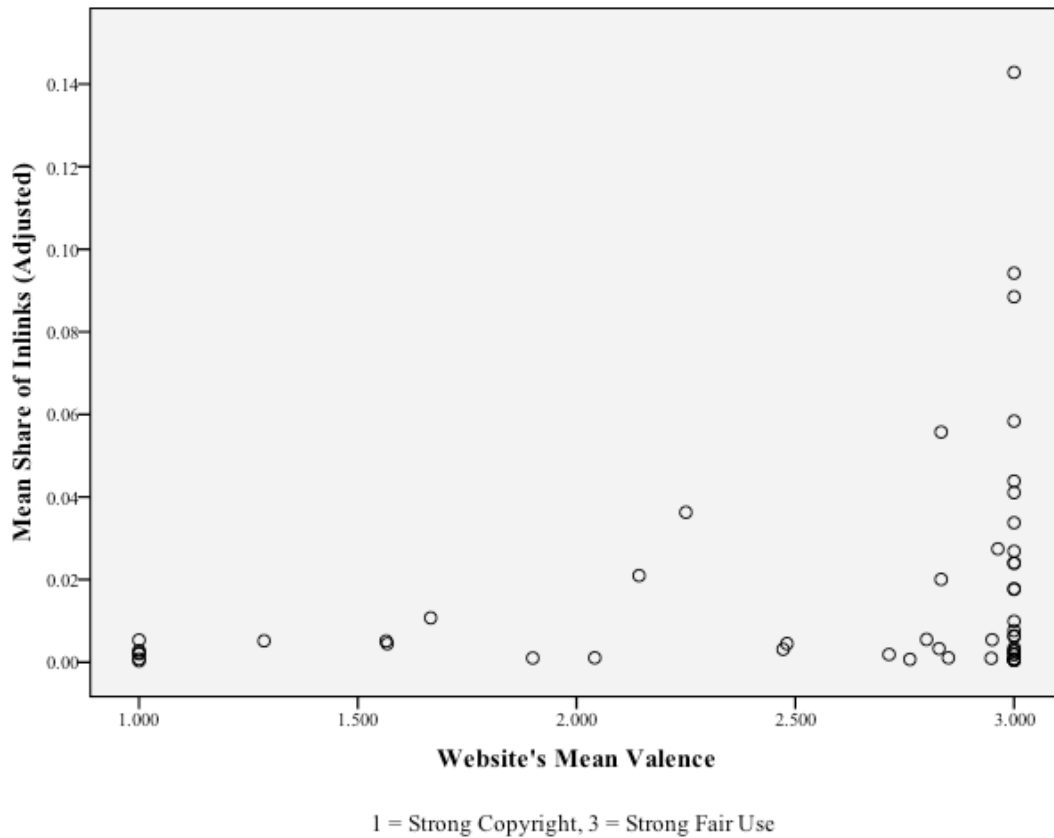
Unlike the graph of relevant documents by valence—Figure 8.7—this ramp shape has no outliers in the strong copyright coalition. Even the Copyright Office website is part of the large group of sites along the bottom of the distribution. All of the sites with a high inlink share are on the side of stronger fair use.

In terms of site-level analysis, this chapter has already shown quite decisively that the strong fair use coalition dominated the online debate in the time period studied. In anticipation of the next chapter's comparisons between online, congressional, and newspaper communication, however, it is also important to discuss the document-level

279

data; the next section starts the process by discussing the distributions of the relevant variables.

Online Documents: Basic Distributions

The site-level analysis concluded that the strong fair use coalition dominated the online debate, and the document-level analysis buttresses this conclusion. Out of 78 included websites, 52 had at least 1 relevant document; since several sites had very many relevant documents, this project identified 933 documents that were relevant to either the DMCA reform debate or the broadcast flag debate. As described in Chapter 4, this project coded for variables other than relevance for up to 10 relevant documents per website per topic, for a maximum of 20 documents per site. In all, 489 documents were coded. These documents give a sound representation of the online debate around these two topics, on variables including rhetorical valence, sector represented, year, and topic of interest.

These documents came in a number of forms; while coding did not systematically classify web documents, a brief overview of some of the more common features is in order. In terms of format, most were ordinary web pages—that is, they were primarily written in Hypertext Markup Language (HTML), the basic language of the web. Quite a few were in Portable Document Format (PDF), and a small number were in Microsoft Word (.DOC) format.

Documents' content and style were much more widely varied. A good number of pages were static web pages that put forth a group's position on a given policy issue, some briefly and some in excruciating detail. Some of these were complete white papers, of the type that in years past would have only been available if they had been issued and

circulated as glossy print products. Others were primers, such as lists of Frequently Asked Questions. Another sizable portion were entries from blogs or issues of online periodicals that contained facts, commentary, and hyperlinks related to recent policy developments. Filling a third major category were reproductions of policy documents submitted to or produced by policymakers. Several websites posted congressional testimony, letters to Congress, court briefs, court decisions, submissions to government agencies such as the Federal Communications Commission (FCC) and the Copyright Office, and agency rulings. The set also included a number of press releases. Also included were online reproductions of print media, generally periodicals such as magazines, newspapers, and law reviews. Finally, a few sites posted reproductions of conference presentations or transcripts of panels at which relevant issues were discussed.

In terms of rhetorical valence, the vast majority of online documents supported the strong fair use coalition. This is unsurprising given the distributions of the site-level variables described above; most sites supported stronger fair use, and the average strong fair use site put more than twice as many relevant documents online as the average strong copyright site. Out of 489 documents coded for valence, 381 documents—77.9 percent—supported the strong fair use position. Only 63 documents, or 12.9 percent, supported strong copyright, leaving 45 documents—9.2 percent—that took a mixed or neutral position. On a scale from one (strong copyright) to three (strong fair use), the mean valence score was 2.66 ($SD = .69$). Figure 8.9 depicts this distribution.

Figure 8.9: Distribution of Web Documents' Valence

Mean =2.66
Std. Dev. =0.691
N =489

Number of Documents (y-axis)

Valence of Web Documents (x-axis)

1 = Strong Copyright, 3 = Strong Fair Use
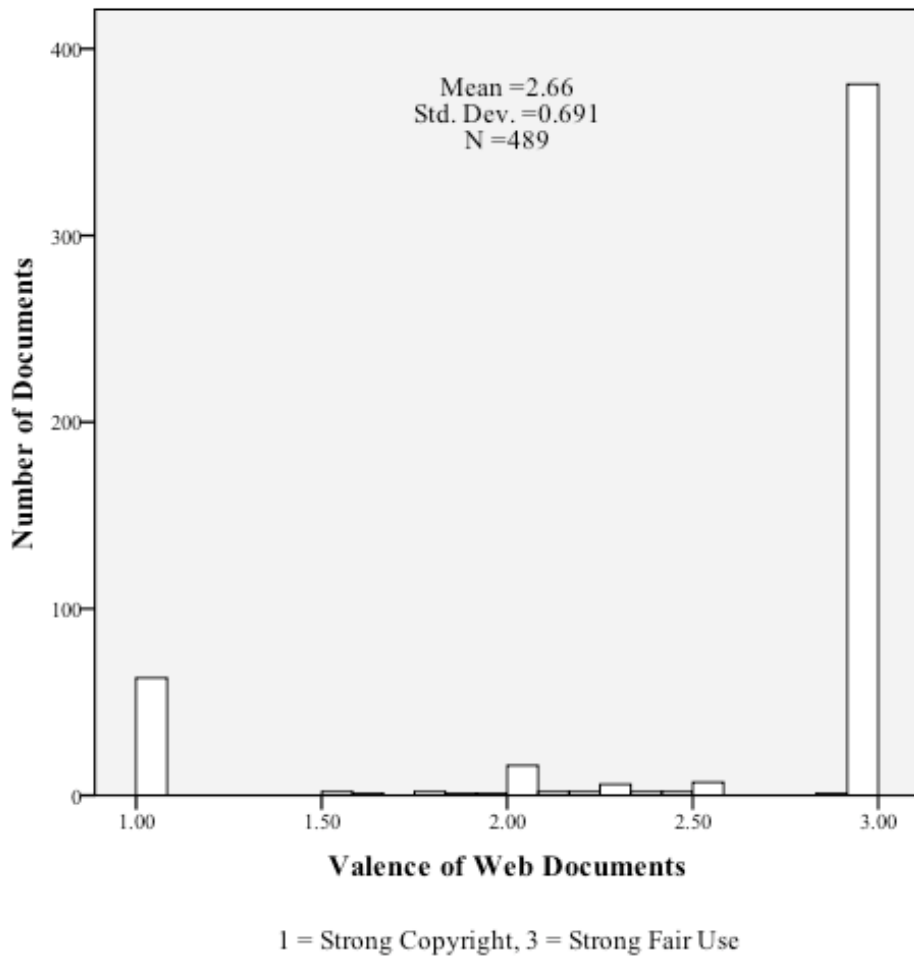
The strong fair use coalition used the web to present a great deal of information on the DRM policy debate, couched in terms that are heavily favorable to them. For every document calling for strong copyright, there were 6 documents pushing for strong fair use. With such a commanding share, from such a large number of documents accessible from nearly any computer in the country, this is a powerful information subsidy with an intended audience of anyone who will listen.

282

Those sectors traditionally associated with the strong fair use coalition provided most of the relevant web documents. Scholars and scholarly organizations such as university institutes authored or co-authored 186 documents and NGOs authored 149. Though somewhat divided in their loyalties, the technology sector remains an important source of fair use support, and they authored or co-authored 102 online documents. In comparison, bastions of support for stronger copyright authored many fewer documents. Media sector voices authored or co-authored just 41 documents, and appointed government officials added just 30. Libraries and librarians, with 47 documents, were just the 4th most-vocal group traditionally aligned with the strong fair use coalition, but they still outnumbered all sectors typically aligned with stronger copyright.

As was done across all media, documents' sectors were coded nonexclusively, leading to 604 sector codes on just 489 documents, for a difference of 111 "extra" sector codes. The website for *Wired* magazine accounted for the largest single share of extra codes; every article was coded as "News", but across 20 coded articles, the authors also quoted an additional 41 sectors. Other notable sources of extra sector codes include the Association for Computing Machinery (20 documents; coded as both "technology" and "scholars"), Chilling Effects Clearinghouse (11 documents; coded as both "scholars" and "NGOs"), and the Digital Future Coalition (4 documents; each coded as 5 sectors). Table 8.4 presents the count and share of each sector's codes when calculated using this nonexclusive coding.

Table 8.4 also presents the results of exclusive coding, using the following priorities for determining sectors: First, any news article was coded as "News", crediting

the sector that has obvious control over the final product. Second, any document coded as both "NGOs" and another sector was coded as representing NGOs. This reflects the fact that people in other sectors have started many of the relevant NGOs, such as the Chilling Effects Clearinghouse and the Digital Future Coalition, but they function more like NGOs than groups in other sectors. Finally, any document coded as representing the technology sector and any other sector was coded as representing the other sector. This had the sharpest effect on the Association for Computing Machinery, which is more scholarly organization than technology sector interest group.[78] The overall distribution between groups is roughly the same in both columns.

---

[78] In any case, the technology sector *per se* is generally composed of for-profit companies whose policy advocacy efforts are directed at increasing profits, while a "purposive organization" such as an NGO is created to advance a given set of core policy beliefs; this leaves technology sector representatives the chance "to be more fluid than those of purposive groups and more conducive to the formation of 'coalitions of convenience' containing members with very different beliefs" (Sabatier & Jenkins-Smith, 1993, p. 225). Thus, it makes more sense to say that the for-profit technology sector is joining with an NGO than vice versa. Other sectors with which the technology sector joined, such as libraries, scholars, and education, are also more purposive than bottom-line driven—at least, they are more purposive than for-profit technology companies.

Table 8.4: Count and Share of Sector Codes for Online Documents

Sector	Exclusive Coding		Nonexclusive Coding	
	Frequency	%	Number	% of Codes
Media	28	5.7	41	6.8
Lawyers	1	0.2	2	0.3
Appointed	22	4.5	30	5.0
Elected[79]	9	1.8	19	3.1
News	20	4.1	20	3.3
Technology	67	13.7	102	16.9
Scholars	160	32.7	186	30.8
NGOs	140	28.6	149	24.7
Libraries	39	8.0	47	7.8
Education	0	0.0	4	0.7
Other	3	0.6	4	0.7
Total	489	99.9[80]	604	100.1[81]

The best-represented groups, scholars and NGOs, have already been established as allied with the strong fair use coalition, and the technology sector is often on the strong fair use side (Herman & Gandy, 2006). As detailed below, their heavy online presence is what drives the strong fair use coalition's online dominance.

The distribution of web documents by year is reasonably normal. Despite the strong bias toward the recent past exhibited by Google and the internet generally (Hellsten, Leydesdorff, & Wouters, 2006), the documents were almost as likely to be

---

[79] This total excludes online reproductions of congressional hearings and documents from hearings on the House and Senate websites. Other documents from official congressional sites were included—that is, documents that are online but not resulting from or included in the printed hearing record. In contrast, hearing documents from other websites were included—as, for instance, when a group posted a copy of its own prepared remarks.
[80] Totals do not sum to 100% due to rounding error.
[81] Totals do not sum to 100% due to rounding error.

dated 2003 (96 documents) as 2006 (108 documents). The mean year was closest to 2004 (2004.22, SD = 2.051), with 47 percent of documents dated 2004 or earlier. Table 8.5 provides the specifics.

Table 8.5: Web Documents by Year

Year	Documents	%	Cum. %
1997	1	.2	.2
1998	4	.9	1.1
1999	4	.9	2.0
2000	19	4.3	6.3
2001	11	2.5	8.8
2002	38	8.7	17.5
2003	96	21.9	39.4
2004	33	7.5	46.9
2005	79	18.0	64.9
2006	108	24.7	89.6
2007	45	10.3	99.9
Total	438	99.9[82]	

This distribution buttresses the belief that the debates about DMCA reform and the broadcast flag mandate were hottest between 2003 and 2006. Out of 438 documents for which a date was assigned, 316 documents, or 72.1 percent, were in this time frame. Only 45 documents were dated 2007, a far smaller share than 3 of the 4 years in the 2003-2006 policy window. Final Google searches of individual sites were conducted in November of 2007, so if the sites had continued to host a wealth of information in that year, 2007 would have had a much higher number of relevant documents. The high totals

---

[82] Totals do not sum to 100% due to rounding error.

from 2003 to 2006 support the belief that the important part of these policy debates occurred during the 108th and 109th Congresses.

This distribution also suggests that the 2-step method used here—web crawling, followed by targeted Google searches of individual sites—provided an acceptably well-archived representation of web documents from years past.[83] Nearly as many documents were dated 2002 (38) as 2007 (45). While Hellsten, Leydesdorff, & Wouters (2006) describe a power-law distribution heavily favoring the recent past for Google search results, Figure 8.10 depicts a comparatively normal-shaped curve.

---

[83] This is not necessarily the case for any possible uses of this search strategy; the success on this count may not hold for any population of sites. It may be that information policy advocates are particularly adept at creating and maintaining a useful online archive of their views, and other populations of online communicators may do a much worse job preserving their historical record online. Future uses of this method must remain sensitive to the possibility of a distorting bias toward the recent past.

Figure 8.10: Number of Web Documents by Year of Publication
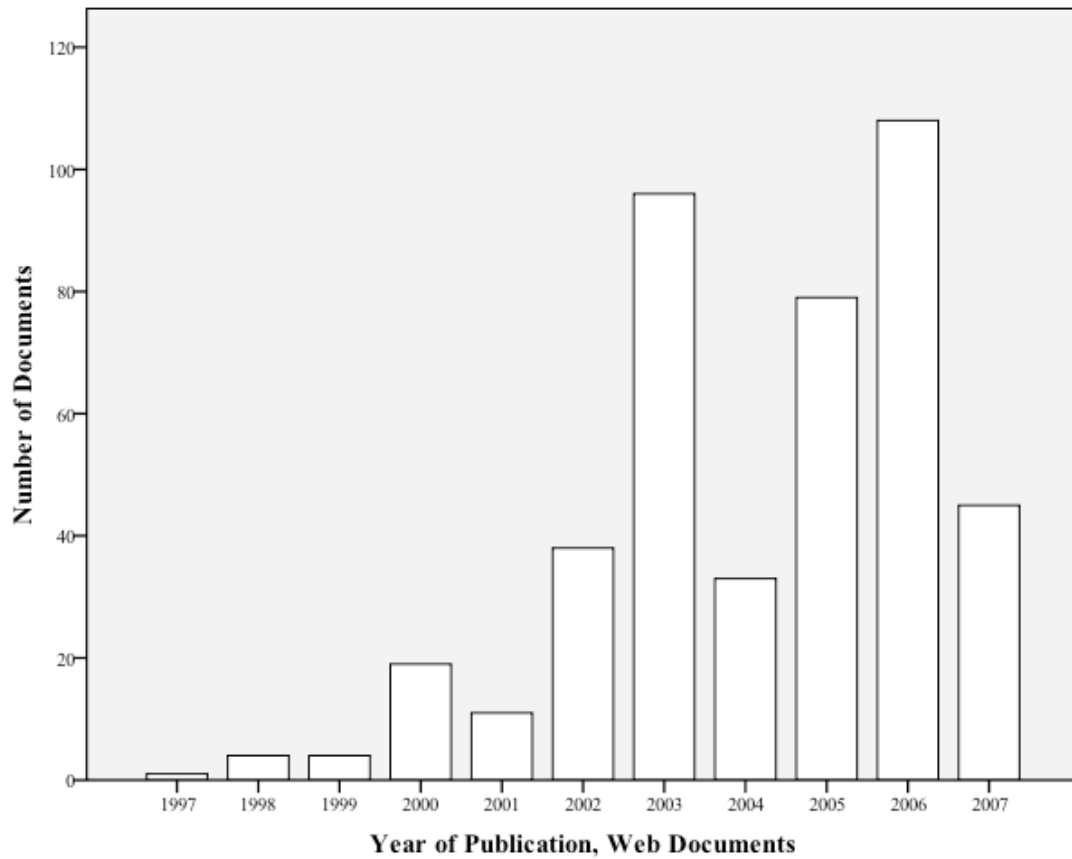
This distribution may have 2 separate peaks—2003 and 2006—but overall, the data cluster toward the middle of the distribution, roughly the line between 2004 and 2005. In short, there is little cause for concern that the data overall are biased toward the recent past. Tests below explore whether this surprisingly robust and searchable archive favors one coalition or the other.

Both of the recent policy debates were well represented, but compared to the broadcast flag, DMCA reform was the more common topic of discussion. Broadcast flag searches yielded 198 relevant topics, or 40.5 percent of the online total, while DMCA

reform searches yielded 291, or 59.5 percent. Since the strong fair use coalition supports

DMCA reform and opposes broadcast flag mandates, and since they dominated the online

debate, this suggests they used the internet to go on the legislative offensive more than

they used it to play defense.

<div align="center">Relating Online Documents' Valence to Other Variables</div>

With such clear numerical dominance by the strong fair use coalition and the

sectors that traditionally ally under the banner of fair use, the overall view of copyright as

portrayed online is fairly clear. Even knowing that, the relationships between valence and

other variables are worth exploring, each combination yielding potential additional

insights. Here, I consider the relationships between valence and sector represented,

valence and year, and valence and topic of interest—DMCA reform or broadcast flag.

As in congressional documents, a document author's sector is an excellent

predictor of a document's valence. Table 8.6 provides the distribution for each sector,

coded exclusively as described above.

Table 8.6: Rhetorical Valence by Sector

	Strong Copyright	Neutral / Mixed	Strong Fair Use	Total
Media	23	4	4	31
Lawyers	0	1	0	1
Appointed	15	7	1	23
Elected	2	1	6	9
News	1	15	4	20
Technology	4	1	62	67
Scholars	2	5	153	160
NGOs	16	10	110	136
Libraries	0	1	38	39
Other	0	0	3	3
Total	63	45	381	489

The sectors that are generally members of the strong copyright coalition who participated online to a meaningful degree exhibited reliable support for the strong copyright position. The media sector supported strong copyright in 23 of 31 online documents, and appointed government officials did so in 15 of 23 documents. The sectors that are generally in support of strong fair use also lined up quite reliably. Scholars (153 out of 160 documents), NGOs (110 of 136), and libraries (38 of 39) presented a highly unified front. Of the 26 NGO documents that were not clearly in the strong fair use camp, 10 were mixed or neutral, many of these being transcripts of forums or other chances for both sides to engage in a debate. The Progress and Freedom Foundation accounted for 10 of the 16 NGO documents in the strong copyright camp.

Remarkably, the technology sector's online documents were almost all in the strong fair use camp: 62 of 67. This is a marked difference from their sharp divisions in Congress, where the sector was split almost evenly over the prudence of strong DRM regulations. While a fuller comparison between these media is the purpose of the next

chapter, this is a remarkable divergence from the very sharp divisions of the sector in

Congress. As presented online, the technology sector is almost 100% behind the strong

fair use position on DRM regulation.

The relationship between online valence and document date is of interest for

reasons both methodological and substantive. Methodologically, if the mean rhetorical

valence trended toward the strong copyright side over time, this would open the

possibility that the strong fair use side's apparently dramatic online advantage could in

part be an artifact of systematic differences in the coalitions' internet archival practices.

That is, one would have to consider the possibility that the strong fair use coalition did a

better job of preserving online documents from years past, creating more of an

impression of online domination than one would see had this study been conducted

repeatedly over the years.

As it turns out, the fear of a valence trend toward the strong copyright side over

time was unfounded; quite the contrary, there was a notable trend in the direction of

stronger fair use, an unexpected and substantively interesting finding. Figure 8.11 shows

this trend.[84]

---

[84] This figure excludes 1997 (1 document, valence = 3), 1998 (4 documents, mean valence = 2.50), and 1999 (4 documents, mean valence = 1.96). All other years had at least 11 documents.

Figure 8.11: Mean Rhetorical Valence by Year: Web Documents

Year of Publication, Web Documents

1 = Strong Copyright, 3 = Strong Fair Use
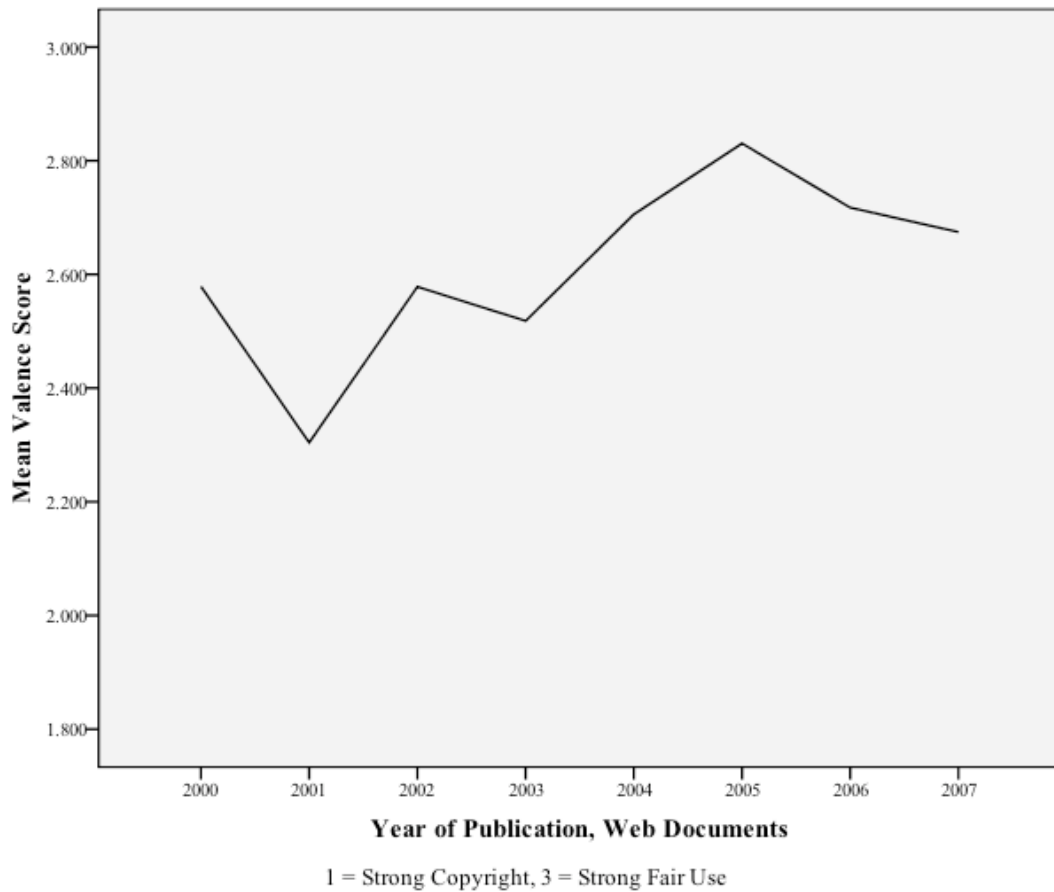
A web document's year and its rhetorical valence—where a score of 1 means support for strong copyright and 3 equals support for strong fair use—are positively correlated,[85] meaning that the online debate trended even more strongly toward the strong

---

[85] Pearson Correlation = .131, p = .006. Two caveats are in order. First, the valence data are highly non-normal, though a non-parametric test confirms this result. Relating categorical valence data based on the three rhetorical categories—strong copyright, mixed/neutral, or strong fair use—to document year, a Kruskal-Wallis analysis of variance is also clearly significant: Chi-square (df = 10, N = 438) = 21.67, p = .017.

As a second caveat, earlier debates have some effect on later debates, so data points are not truly independent; thus, both parametric and ordinary nonparametric tests (e.g.,

fair use coalition over time. While not central to testing this project's hypotheses, this finding is consistent with the findings in the last two chapters; over time, the strong use coalition gained strength in all three media. Compared to congressional hearings and print media, the high volume of copyright policy advocacy on the web is a recent phenomenon, and even this short time frame is long enough to suggest a similar trend toward a stronger position for those calling for stronger fair use. Even though the strong fair use coalition began from a position of clear strength, they extended their domination to a substantial degree.

There is no substantive relationship between rhetorical valence and topic of interest. The mean valence was virtually identical for documents found in documents discussing DMCA reform (291 documents, mean valence = 2.70, SD = .685) and those discussing the broadcast flag (198 documents, mean valence = 2.62, SD = .699). In short, each coalition spent roughly the same proportion of its energy on each topic.

Conclusion

By investing heavily in the online debate, the strong fair use coalition has carved out a unique forum. Unlike Congress and the elite print media, the web is not subject to much if any gate keeping; rather, the web is better described as a "peer-production model of filtering and discussion" (Benkler, 2006, p. 258). Compared to congressional and print

---

Kruskal-Wallis) are not the preferred means for establishing change over time. Here, these tests serve first and foremost as tools to establish that the online results are not a methodological artifact, rather than a means of testing central hypotheses, increasing the acceptability of using slightly inappropriate tests. With such a clearly significant result that is consistent with the trends identified in other media, it is even reasonable to use these results for reporting the unexpected result, caveat in hand.

media access, online success is far less dependent on financial backing. As a political medium, the web is more dependant on citizen-activist enthusiasm, technical skill, and a greater interest in expanding the scope of conflict. On all these counts, the strong fair use coalition has serious advantages over the strong copyright coalition, and in the debate over DRM policy, they turned these advantages into overwhelming online domination.

The online DRM debate flows through the hubs at the center of the strong fair use coalition. With so many passionately involved NGOs, scholars, technologists, and librarians linking to one another, central groups such as the Electronic Frontier Foundation and the Free Software Foundation have a torrent of topically relevant incoming traffic. Even second-tier sites within the strong fair use coalition still have more incoming links than all but the very most-linked sites in the strong copyright coalition.

The strong fair use coalition also publishes much more policy-relevant information online. If one were to learn about the DRM policy debate exclusively through searching the web, one might get the impression that the strong copyright coalition has few detailed arguments to make. The congressional record proves otherwise, but the strong copyright coalition appears not even to put much of this already-produced information online; the failure to exert even this small effort suggests almost no interest in putting their version of the issue into the online mix. In contrast, the strong fair use coalition provides a heavy information subsidy on the web. In the time frame studied, it was effectively no contest; the strong fair use NGOs and scholarly groups alone provided a veritable library of DRM policy developments. This numerical advantage even grew over time.

All this adds up to a substantial online advantage for the strong fair use coalition. This may primarily reflect intra-coalition dialogue—in this case, the web may be more of a resource for communicating with friends than strangers. Even if this is the primary significance of the results reported here, the power of this new tool is not to be underestimated. For a coalition composed of many diffuse, geographically dispersed groups and individuals without substantial capital to invest in this policy debate, even the chance to brainstorm and refine their arguments cheaply and easily is a new and powerful development.

Communication with an issue public—both those already mobilized and those potentially mobilized—is also an important goal and possible outcome. This study's methods do not measure the audience for these online messages—let alone the effects on such an audience. Yet these results at least suggest the real possibility of nearly one-sided persuasion of a nontrivial number of people via the web. For a typical citizen curious about the issue, the web is easily the most accessible research tool, and it is the most likely means by which most people will stumble onto this issue. One person might wonder why her laptop does not rip DVDs with the same ease as it rips CDs, and upon looking online for software that can solve this problem, find a host of activists deriding the law that makes one of these seemingly identical activities legal and the other illegal. Another might search for information about the transition to digital television broadcasting, discovering a policy paper warning him about the proposal to limit his ability to record and reuse broadcast television content. The strong fair use coalition is set up to capitalize on these opportunities, answering these kinds of questions while at the

same time politicizing them in a way that may draw these people into the battle as citizens whose lives are impacted by technology policy.

Even if a coalition is winning the war for hearts and minds, this victory does not necessarily translate into legislative victories. In part, this is because their internet message must compete with the messages from every other coalition in a similar situation on the full spectrum of policy issues, and this bottleneck of attention (Jones & Baumgartner, 2005) means that only a few under-resourced coalitions at a time can successfully wrangle the public attention to a sufficient degree to alter the policy dynamic. Those in the strong fair use coalition likely believe that, if they could somehow get enough citizens to learn about the issue and take action, they would be able to amend Title I of the DMCA.

There are countless advocacy coalitions in similar positions, fighting similar battles on innumerable issues: labor policy, environmental policy, food and nutrition policy, and so on. On many issues, there are undoubtedly one or more advocacy coalitions who also likewise face a competitive disadvantage in material resources but enjoy a mobilization advantage; they likely also believe that a surge in public attention could help them change the dynamic in their policy subsystem. Yet relatively few of these issues will rise to prominence in a given year, and none can stay atop the agenda for long. Thus, internet advocacy is perhaps best viewed as a metric for groups' organizational capacity to surf the next unpredictable wave of public attention. As long as the House and Senate floor and the nation's elite media pay little attention to an issue—as long as the only people paying attention are the relevant congressional committees and a

296

small slice of the electorate—those who have shaped the policy as it now stands will likely continue to have the upper hand.

It still makes sense for an under-resourced coalition to reach out to communicate on the web, even if they may not see immediate legislative results. The ability to communicate at a distance with the members of one's coalition for little financial investment is itself a remarkable new opportunity. Also, such a group has an obvious incentive to have as many sympathetic, informed citizens as possible the next time something dramatic brings a rush of public attention and a chance to redefine an issue. This happened in the copyright policy subsystem when Napster's entrance into the public consciousness had the electorate asking questions about copyright in the digital era (Vaidhyanathan, 2004), but the still-nascent strong fair use coalition was not yet strong enough to capitalize on this opportunity. As just one example, Public Knowledge, now the coalition's strongest link between online advocacy and policymakers' attention, was just getting off the ground. However, this will not likely be the last major event that brings the public's attention to the matter. While web advocacy has not yet led to a reordering of the policy subsystem, this organizational and rhetorical groundwork may prove invaluable to these insurgents when the next wave of public attention comes along.

CHAPTER NINE: COMPARING MEDIA

Taken together, the previous three chapters suggest that the copyright debate

looks quite different depending on the medium through which it is viewed. This chapter

formalizes that suggestion into an empirically demonstrated argument. I begin with a

brief methodological discussion, describing how these media can be compared despite the

internet's short history and shorter memory. Next, I describe how each of the three media

look quite different in terms of their overall rhetorical valence; those in the strong fair use

coalition use the internet to construct a rhetorical space quite distinct from those in

Congress and the major papers. Finally, I discuss how different media represent the

debate over digital rights management (DRM) regulation through the eyes of different

populations of witnesses; those sectors most strongly allied with the strong fair use

coalition take to the internet to make their case in numbers far greater than their

representation in the other two media.

Ensuring Timely Comparison

After a few years of debate, Congress passed the AHRA (*Audio Home Recording

Act*, 1992). In contrast, the World Wide Web did not explode in popularity until 1994

(Kelty, 2008). By the next year, Congress began holding hearings on the DMCA (*Digital

Millennium Copyright Act*, 1998b), passing the law while the internet was just getting

started as a major medium that would permeate US culture. Thus, two of the three

periods under study passed while the web's potential as a political tool was still years

298

away. Comparing Congress and the newspapers across all time periods is straightforward; both media were well established—to say the least—by 1989, and both have good archives. Comparing offline and online media, however, is a trickier proposition. To address this concern, all comparisons are within the 2003 to 2006 window. Thus, the AHRA and DMCA debates are set aside, and internet documents without a clear date or with a date before or after the 2003 to 2006 window are ignored for this chapter.

This study also must account for the internet's short memory. I carefully archived all internet documents as soon as possible because a substantial portion of the relevant documents have doubtless gone offline in the short period between the study's conception and its completion. These documents were all retrieved in one shot, rather than in repeated searches of the same sites, bringing up the potential that the earlier years in the time frame considered are not adequately represented. Thus, in addition to the comparisons of rhetorical valence based on documents dated 2003 to 2006, I also present the results of comparisons of valence based exclusively on 2006 documents. Doing so ensures that the central questions of this research are answered based on a thorough consideration of alternative hypotheses—including the supposition that the online debate might only look as strongly pro-fair use due to a systematic bias in each coalition's willingness to leave documents online years after their initial publication.

The comparison of 2006 documents only works for comparing web documents to congressional documents, as the set of relevant newspaper articles does not include any stories from 2006; even if it did, the number of relevant articles is so small that looking at any single year would likely entail examining too small a set to be helpful. The

comparison of 2006 documents between Congress and the web will have to suffice,

though as noted below, this comparison looks enough like that for the entire 2003 to 2006

period as to suggest little cause for fear about a false positive resulting from the internet's

short memory.

Despite these methodological corrections, comparisons between online and

offline media are less than ideal. Even though the comparisons here represent a

suboptimal solution, the differences between online and offline representations of the

DRM debate are so substantial as to allow for reasonably certain conclusions despite a

magnified margin of error.

Comparing Rhetorical Valence Across Media

The three media differed substantially in their representation of the prudence of

DRM regulation. More specifically, as one moves from those statements nearest the halls

of power to those farthest away—from Congress to the newspapers to the internet—one

moves toward ever-stronger support for the strong fair use position. The sharpest

difference is between the internet and the other two media, but even the difference

between congressional hearings and newspaper articles is worth noting. Since those two

offline media are most easily compared across all time periods, I begin there.

Rhetorical Valence: Congress versus Newspapers

In terms of rhetorical categories, the difference between newspapers and

congressional testimony seems at first glance to be quite sharp. See Table 9.1:

Table 9.1: Rhetorical Categories, Congress versus Newspapers

		Congress or Newspapers		
		Congress	Newspapers	Total
Rhetorical Categories	Strong Copyright	241	12	253
	Neutral/Mixed	29	30	59
	Strong Fair Use	165	16	181
	Total	435	58	493

These media are significantly different in terms of the rhetorical categories represented ($\chi^2 = 100.3$, $df = 2$, $p < .001$), and the difference is substantial (Cramer's $V = .451$, $p < .001$). Yet a good deal of this difference is the effect of a sharp difference in the relative propensity for a document to be mixed or neutral; these represent over half of the newspaper articles but fewer than seven percent of congressional documents. Removing neutral documents and considering only those documents that clearly support one coalition or the other, the differences between the media are a good deal less significant ($\chi^2 = 2.934$, $df = 1$, $p = .087$), representing a substantially smaller though still nontrivial effect ($\varphi = $ Cramer's $V = .082$, $p = .087$).[86] Among documents taking a clear side, a given congressional document is 1.95 times as likely to support strong copyright as is a newspaper article.[87] This difference may not be gargantuan, but it provides clear evidence

---

[86] These p values do not lead one to support the null hypothesis in this case. Because the documents consist of a census rather than a random sample, there is no concern that the outcome is actually the result of random chance.

[87] A non-neutral congressional document is 1.46 times more likely to support strong copyright than strong fair use, while a one-sided newspaper article is only .75 times as likely to support strong copyright. Framing in terms of odds ratios makes the effect sound larger, but taking the natural log of this result, .667, and converting to a d-like measure of effect size (Chinn, 2000) gives us an estimate of .367, representing a modest if nontrivial difference.

that newspapers are a slightly friendlier venue for the strong fair use coalition than are congressional hearings.

Comparing mean rhetorical scores paints a similar picture: a real if modest difference between newspapers and Congress. The congressional mean over all periods was 1.83 (SD = .95), while the newspapers averaged a rhetorical score of 2.11 (SD = .72).[88] The congressional average leaned modestly toward stronger copyright, while the newspaper average leaned slightly toward stronger fair use. This represents an effect size (d = .31) that is in the range of small but not trivial (Cohen, 1992). Measured in multiple ways, newspapers really are further toward the strong fair use end of the scale than congressional documents. Taken together, these figures provide solid support for Hypothesis 3: The ratio of strong fair use arguments to strong copyright messages is larger in elite newspapers than in Congress.

Rhetorical Valence: Online versus Offline Media

While the difference between newspapers and congressional documents is modest but nontrivial, the differences between offline documents and web documents in terms of rhetorical valence are remarkable. The internet is a haven for strong fair use messages in proportions that dwarf those offline. This is true when comparing only documents from 2003 to 2006, even though both newspapers and Congress shifted substantially toward

---

[88] Levene's test for equality of variances (F = 81.2, p < .001) indicates that variances are clearly not equal; assuming unequal variances, this difference is highly significant (t = -2.70, df = 86.1, p = .008).

the strong fair use coalition in the latest time period. Consider Table 9.2, which lays out the number of documents by rhetorical category in each medium:

Table 9.2: Rhetorical Categories by Medium, 2003 to 2006

		Number of Documents by Medium, 2003-2006 (2006 only)			
		Congress	Newspaper	Web	Total
Rhetorical	Strong Copyright	69 (27)	1 (0)	37 (13)	107 (40)
Categories	Neutral/Mixed	19 (12)	17 (0)	29 (5)	65 (17)
	Strong Fair Use	75 (26)	6 (0)	250 (90)	331 (116)
	Total	163 (65)	24 (0)	316 (108)	503 (173)

This highlights the exceptional differences between these media in terms of the rhetorical categories represented ($\chi^2 = 141.1$, $df = 4$, $p < .001$), and the difference is substantial (Cramer's $V = .374$, $p < .001$). Further, unlike the comparison between newspapers and Congress, the difference stays fairly sharp even after removing neutral documents and considering only those documents that clearly support one coalition or the other ($\chi^2 = 64.1$, $df = 2$, $p < .001$). The effect size even increases slightly (Cramer's $V = .383$, $p < .001$), staying well within the range of a medium to large effect. See Figure 9.1, which shows the proportion of documents that fall into each rhetorical category within each medium:

Figure 9.1: Rhetorical Categories by Medium, 2003-2006
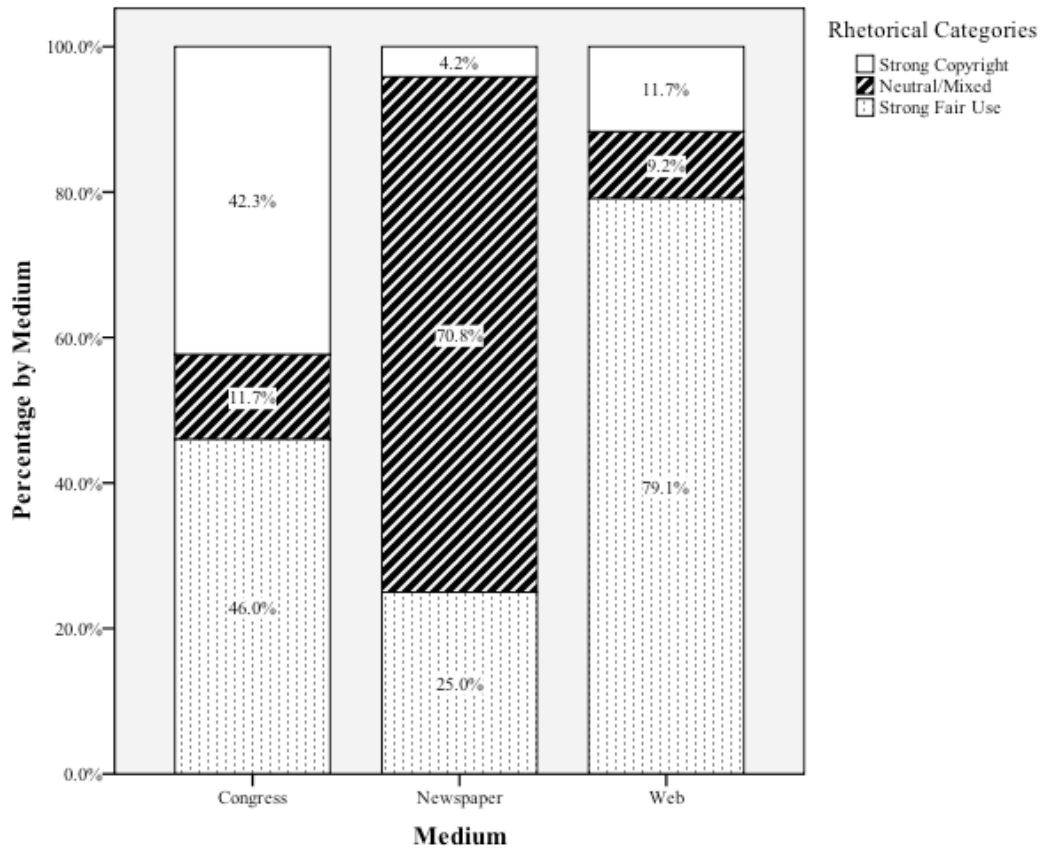
The comparison between online documents and congressional documents highlights a stark difference. Among documents taking a clear side, a given web document is 6.22 times as likely to support strong fair use as is a given congressional document, representing an exceptionally large effect size.[89] In 2006 documents exclusively, the same sharp contrast remains. Among 2006 documents taking a clear side,

---

[89] A non-neutral congressional document is 1.09 times more likely to support strong fair use than strong copyright, while a partisan web document is 6.76 times more likely to support strong fair use. The natural log of the odds ratio, 1.83, converted to a *d*-like measure of effect size (Chinn, 2000) gives an estimate of 1.01, representing an exceptionally large effect (Cohen, 1988; Morgan, Leech, Gloeckner, & Barrett, 2007).

a given web document is 7.19 times as likely to support strong fair use as is a given

congressional document.[90] This provides rock-solid support for Hypothesis 8: The ratio

of strong fair use arguments to strong copyright messages is larger online than that in

Congress.

The very low number of side-taking newspaper articles in the latest time period—

just seven such articles—makes similar comparisons with online documents less

meaningful. The ratios are quite similar, with an odds ratio of 1.13 times more partisan

newspaper articles supporting strong fair use than partisan web documents supporting

strong fair use. While this provides a reason to doubt Hypothesis 9, that the ratio of

strong fair use arguments to strong copyright messages is larger online than that in elite

newspapers, it is a thin reason because the bulk of newspaper articles are neutral or

mixed. With just 7 articles falling cleanly into either camp in the 2003 to 2006 time

period, it is hard to draw any clear conclusion. It is better to look for a way to consider

neutral newspaper articles as well, since these make up over 70 percent of the still-

regrettably-small group of articles.[91]

---

[90] For 2006 documents only, a non-neutral congressional document is .96 times as likely
to support strong fair use as strong copyright, while a partisan web document is 6.92
times more likely to support strong fair use. Applying Chinn's (2000) formula gives an
effect size estimate of 1.09, exceeding even the exceptionally large effect of 1.01 for the
entire time period.
[91] Consider the odds in each medium that a document will support stronger fair use
versus either neutral or strong copyright documents. Here, the odds ratio suggests that a
given web document is 11.36 times more likely to support strong fair use than a given
newspaper article, for an enormous d-like effect size of 1.34. This overstates the
difference between the media, but the tenfold difference in odds ratios highlights the
problematic nature of calculating an odds ratio in a way that ignores neutral or mixed

Comparing media based on mean rhetorical valence sheds additional light. These means are displayed in Figure 9.2:

Figure 9.2: Rhetorical Valence by Medium, 2003-2006

1= Strong Copyright, 3 = Strong Fair Use
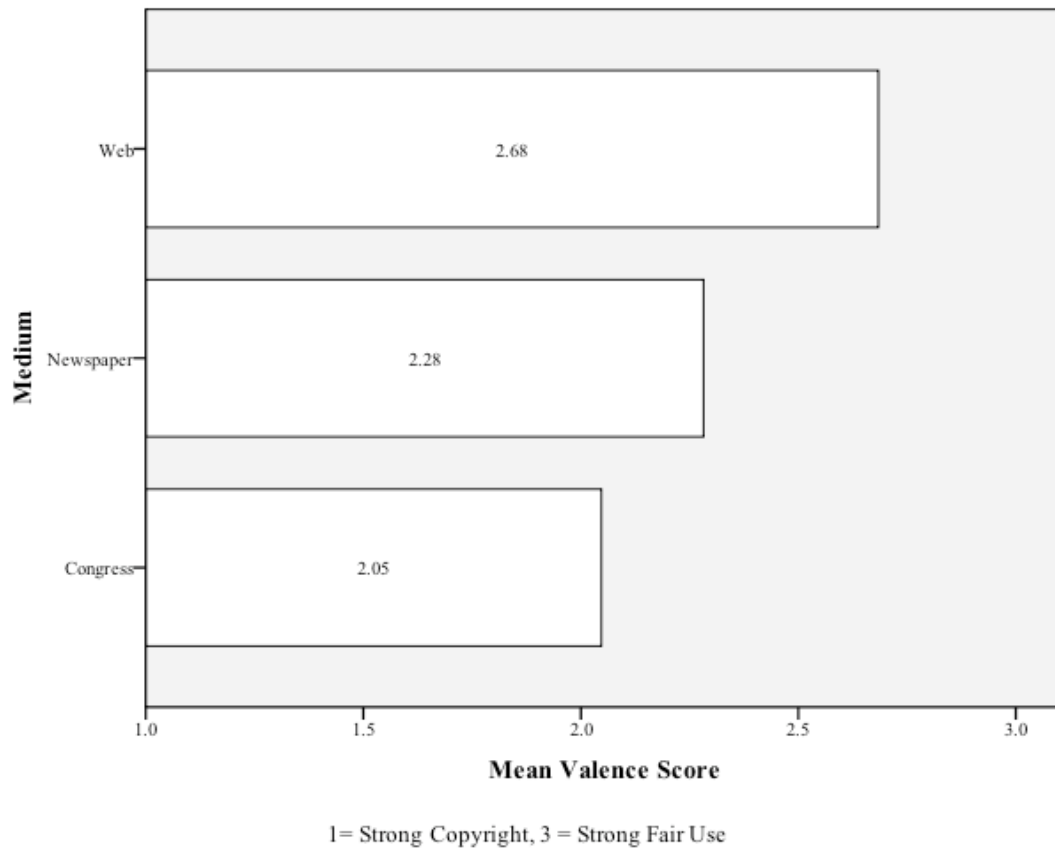
The differences between mean scores by medium are quite stark. Congressional documents are very near an average of perfect neutrality: 2.05 (*SD* = .95). In contrast, the newspapers average 2.28 (*SD* = .52), and web documents average 2.68 (*SD* = .67). The

newspaper articles. These media really are different, but a calculation accounting for neutral documents is prudent.

difference between web documents and newspaper articles is highly significant. Variances are not significantly different ($F = 1.27$, $p = .261$), and assuming that they are equal, $t = -2.87$ ($df = 171$, $p = .004$). This represents a medium to large effect size ($d = .61$). Since this comparison incorporates the neutral or mixed newspaper articles, it is much more appropriate than the odds ratio described above. Taken together, these findings offer solid support for Hypothesis 9: The ratio of strong fair use arguments to strong copyright messages is larger online than that in elite newspapers.

The difference between web documents and congressional documents is even more profound. Levene's test shows the variances to be significantly unequal ($F = 113.797$, $p < .001$), and without assuming equal variances, $t = -8.52$ ($df = 477$, $p < .001$). This represents an atypically large effect size ($d = .82$), meaning a profound difference in how these two media represent the debate. The difference between the media is larger still when considering only 2006 documents. The mean for web documents was 2.72, and it was 2.02 for congressional documents. With equal variances not assumed, $t = -5.33$ ($df = 104.7$, $p < .001$). Again, the effect size ($d = .90$) is larger than is typical in social scientific research (Cohen, 1988; Morgan et al., 2007).

The differences between online and offline communication around DRM policy are not only real and in the predicted directions, but they are so substantial as to create wildly different views of the debate. If one comes to learn about the copyright debate primarily via congressional hearings, supplemented by the light coverage in a major daily newspaper, one is getting a two-sided view of the matter; both sides in the DRM policy debate were well heard in these media between 2003 and 2006. In contrast, if one's

307

primary source of detailed information about DRM regulation is via the web, the current policy landscape—in which the DMCA stands, while proposals to expand the reach of DRM regulation are seriously considered—may be downright infuriating. The vast majority of online documents called for a halt in or reversal of the expansion of copyright as a vehicle for regulating digital technology. The structural differences between online and offline policy forums—including both incentive and ability to communicate—have created profoundly different versions of the same debate.

Comparing Sector Representation Across Media

Underlying the profound differences between media in their coverage of the DRM policy debate are sharp differences in the types of actors that are represented. Groups that are likely to support stronger copyright are relatively more likely to appear in congressional hearings and newspaper articles, whereas groups that historically support stronger fair use dominate the online debate. Before including the online debate—which requires that analysis be limited to the 2003 to 2006 period—consider the differences between the offline media in sector representation. Table 9.3 shows the number of documents in each medium that included quotations from each sector. Codes are allocated nonexclusively, accounting for the fact that many documents represent more than one sector. For instance, a given newspaper article may include quotations from the media sector and the technology sector, meaning that the article gets three codes—media, technology, and news. Thus, the total number of codes is higher than the number of documents. In particular, nearly all newspaper documents are coded as representing the news sector; the only exceptions are a few opinion pieces (op-eds or letters to the editor)

submitted by representatives of other sectors. The table also includes the share of total

documents in a medium that at least partially represented a given sector; these shares sum

to over 100 percent, as many documents represent multiple sectors.

Table 9.3: Sector Representation, Coded Nonexclusively, in Congress and Newspapers

Sector, Nonexclusive	Medium, Percentage of Total Documents by Medium				
	Congress (435 Doc's)	% of Doc's	Newspaper (58 Doc's)	% of Doc's	Total
Media	123	28.3	28	48.3	151
Lawyers	5	1.1	3	5.2	8
Appointed	33	7.6	8	13.8	41
Elected	85	19.5	9	15.5	94
News	8	1.8	55	94.8	63
Technology	121	27.8	26	44.8	147
Scholars	24	5.5	6	10.3	30
NGOs	31	7.1	13	22.4	44
Libraries	22	5.1	3	5.2	25
Education	10	2.3	0	0.0	10
Other	5	1.1	5	8.6	10
Total Codes	467		156		623

Note that the number of sector codes for newspapers (156) is 2.7 times as large as

the number of documents (58), while congressional documents (435 in total) rarely

represent more than one sector (467 codes, for a ratio of 1.1). Omit the news sector (101

codes, ratio of 1.7), and the average sector is roughly 1.6 times as likely to appear in any

given newspaper document as in any given congressional document. Among sectors with

double-digit frequency in either sector, this ratio between newspaper share and

congressional share is a good estimate: the media sector (1.7), appointed government

officials (1.8), the technology sector (1.6), and scholars (1.9) are all between 1.6 and 1.9

times as likely to appear in a given newspaper article as in a given congressional

document. Unsurprisingly, members of Congress were represented in more congressional documents (20%) than newspaper pieces (16%). The only major group that did markedly better in newspapers was the NGO sector, more than tripling their share from seven percent to 22 percent. This correctly suggests that NGOs such as Public Knowledge and the Electronic Frontier Foundation are constantly seeking publicity for their cause and are easily available to reporters. As the anchoring sector of the strong fair use coalition, the NGOs are constantly seeking to legitimize their position in the court of public opinion. Yet these groups have not, until recently, had reliable access to congressional hearings; thus, the first and second periods weigh down their average presence in Congress. Further, including three times as many quotations from NGOs helps to explain newspapers' slight lean toward the strong fair use coalition, especially as compared to the slight lean toward the strong copyright coalition conveyed in congressional documents.

Next, consider the differences between these offline media and the web in terms of sector representation. As with rhetorical valence, this analysis is limited to the period from 2003 to 2006. Table 9.4 shows the breakdown by sector:

Table 9.4: Sector Representation, Coded Nonexclusively, by Medium, 2003 to 2006

	Congress (163 Documents)	% of Documents	Newspaper (24 Documents)	% of Documents	Web (316 Documents)	% of Documents	Total
Sector							
Media	38	23.3	14	58.3	23	7.3	75
Lawyers	0	0.0	1	4.2	1	0.3	2
Appointed	6	3.7	6	25.0	21	6.6	33
Elected	50	30.7	5	20.8	8	2.5	63
News	0	0.0	23	95.8	10	3.2	33
Technology	34	20.9	11	45.8	58	18.4	103
Scholars	9	5.5	4	16.7	127	40.2	140
NGOs	20	12.3	10	41.7	101	32.0	131
Libraries	13	8.0	1	4.2	28	8.9	42
Education	3	1.8	0	0.0	0	0.0	3
Other	2	1.2	4	16.7	3	0.9	9
Total Codes	175		79		380		634

The differences between the web and offline media are substantial. For instance, media sector voices accounted for 23 percent of congressional documents and appeared in 58 percent of newspaper articles, but they appeared in just seven percent of web documents. Elected officials authored 31 percent of congressional documents—members of Congress take advantage of their pulpit—and appeared in 21 percent of newspaper articles, but just fewer than three percent of web documents included the voices of elected officials. Scholars helped shape just under six percent of congressional documents, but they appeared in 17 percent of articles and authored a whopping 40 percent of web documents. NGOs did better in Congress in the latest period (12%) than across all time periods (7%) but really shone in landing quotations in news stories (42%)

and putting documents online (32%). The technology sector was the only other group that appeared with double-digit frequency across congressional documents (21%), newspapers (46%), and the web (18%).

In most cases, groups' shares were comparable when using exclusive coding. Exclusivity was applied as follows: First, any news story was lumped into the news category. This had an unfortunate effect on newspaper documents—lumping all but one into the news sector—but it best reflects the sector that has the most say in the final product. Next, any document coded as representing the NGO sector was put there. This puts groups such as the Digital Future Coalition in the most reasonable category, as they function more as NGOs than as groups in any other sector. Also, several multi-sector web documents were roundtable discussions hosted by an NGO. Third, any remaining document coded as representing a scholarly group was included in that sector. This captured web documents from groups such as the Berkman Center for Internet and Society at Harvard,[92] and the Center for Internet and Society at Stanford Law School. Next came the other nonprofit sectors: libraries and education. Then, sectors were included in the order below, except for the technology sector, which was only applied when no other sectors were present, and "Other," which refers only to persons who do not clearly represent any sector. This reflects the sector's diversity and diverse views on DRM policy; when a technology sector representative co-authors a document with any other sector, the other sector is generally a much better predictor of rhetorical valence.

---

[92] At the time of document collection, the Berkman Center was part of Harvard Law School; while Berkman retains much of its focus on law, it is now associated with Harvard University.

Each sector's representation in each medium is illustrated in Table 9.5:

Table 9.5: Count and Share of Sector Codes, Coded Exclusively, by Medium

Sector	Congress	%	News-paper	%	Web	%	Total
Media	38	23.3	0	0.0	15	4.7	57
Appointed	6	3.7	0	0.0	13	4.1	22
Elected	50	30.7	0	0.0	3	0.9	53
News	0	0.0	23	95.8	10	3.2	33
Technology	31	19.0	0	0.0	37	11.7	68
Scholars	6	3.7	0	0.0	115	36.4	117
NGOs	20	12.3	0	0.0	95	30.1	111
Libraries	10	6.1	0	0.0	26	8.2	36
Other	2	1.2	1	4.2	2	0.6	5
Total[93]	163	100	24	100	316	99.9	503

Coding all news articles as being in that sector removed newspaper articles from useful consideration here, but the differences between congressional representation and online presence are comparable to those shown by nonexclusive coding. The raw number and share of congressional documents remained the same for media (23%) and members of Congress (31%), while these sectors had an even smaller share of web documents—five and one percent, respectively. The technology sector still did well in Congress (19%), and while their online share was a good bit smaller with exclusive coding (12%) than with nonexclusive coding (18%),[94] this still represents a substantial online presence.

---

[93] Totals may not sum to 100% due to rounding error
[94] 14 of the 21 web documents coded as the technology sector and at least one other sector were hosted by the Association for Computing Machinery, which functions much more as a scholarly organization than as a lobbying arm of the technology industry.

As with nonexclusive coding, the exclusive coding illustrates that NGOs and scholarly groups did far better online than in Congress. Retaining their 12 percent share of congressional documents, NGOs were the fourth most-common presence in hearings—after members of Congress and the media and technology sectors—but even this relative success pales compared to their 30 percent share of web documents. In an even starker contrast, scholars had just four percent of the exclusive share in Congress but a whopping 36 percent exclusive share online. Taken together, the nonprofit sectors—NGOs, scholars, and libraries—had an amazingly large 78 percent of web documents, even though these groups had a combined share of just 22 percent of congressional documents. Congress and the web differ significantly in terms the sectors represented (χ^2=189.4, df = 8, p < .001). In terms of effect size (Cramer's V = .629, p < .001), this represents a large to very large effect (Cohen, 1988; Morgan et al., 2007). These differences are substantial enough to account for a great deal of the sharp difference in rhetorical valence between these media.

Conclusion

The DRM policy debate looks very different depending on the medium through which one gets one's information. From 2003 to 2006, policymakers and readers of major dailies got a fairly two-sided view of the matter. Relative to earlier time periods, both forums had moved somewhat toward the strong fair use end of the scale, but the voices of stronger copyright—in particular, the media industry, the strong copyright wing of the technology sector, and strong copyright supporters in Congress—retained a powerful presence. In contrast, the web has presented a nearly one-sided view of the debate. NGOs

and scholarly organizations in particular have gone online in droves to rail against DRM regulation, and librarians and the strong fair use wing of the technology sector have joined this cause in substantial numbers.

In terms of online communication, this one-sidedness represents primarily a sharp difference in motivation. The strong copyright coalition has historically won or fought to a draw in every major debate—copyright as a tool for regulating technology always seems to expand rather than shrink, at least if one discounts judicial setbacks. These groups have reasonably solid access to the halls of power, so they have diminished incentive to divert energy away from making their case directly with policymakers in an effort to arouse public sympathy. Thus, media companies are remarkably uninterested in engaging in the online debate over copyright law. Elected officials rarely communicate their message within the online debate.[95] Even the technology sector—at least, those technology sector groups that are not also affiliated with a nonprofit sector such as scholars or NGOs—has a larger relative share in congressional documents than among online documents.

In sharp contrast, the strong fair use coalition is engaged in a herculean effort to generate as much sympathetic publicity as possible. These groups are trying to expand the scope of conflict (Baumgartner & Jones, 1993; Schattschneider, 1960) by turning it

---

[95] As described in the previous chapter, the tally of online documents does not include online reproductions of congressional hearings and documents from hearings on the House and Senate websites. Further, the total here is a few documents lighter than the 9 total online documents representing elected officials noted in the last chapter. The other 6 either had no discernible date or had a date outside the 2003-2006 time frame—generally, 2007.

into an issue for voters. They know that leaving the policy order untouched—working

within the established pecking order of congressional debates over copyright law—means

a continued expansion of the reach of copyright law. As an under-resourced coalition,

they use the web to coordinate with minimal expense and maximal leverage of

sympathetic online voices. As the challenger coalition, they are also attempting to use the

web to reach ordinary citizens. Online communication is not nearly as direct or powerful

a route to policy outcomes as congressional testimony—per message, it is almost

certainly not even as effective as favorable coverage in the *Times* or *Post*—but the

internet is the strong fair use coalition's home turf, and they certainly show a willingness

and ability to rack up many political points there. Faced with an opposition that enjoys

the status of dominant coalition and thus has little incentive to expand the scope of

conflict, the members of the strong fair use coalition confront little opposition online and

are thus able to carve out the medium as their own.

CHAPTER TEN: CONCLUSION

This dissertation has explored the debate over DRM regulation via copyright law, both as a topic of interest in its own right and as a vehicle for discussing the policymaking process in general. I began by asking who has access to which media of political communication: Who appears most often in Congress? Whose voices appear most regularly in the press? Who plays the most authoritative role in the online debate over DRM policy? I have demonstrated that the strong copyright coalition, which supports the expansion of copyright as a tool for regulating technology, has historically dominated the congressional debate, though this advantage in frequency of appearance was effectively erased by the 2003 to 2006 time period. Unexpectedly, I found that the strong fair use coalition, which opposes the growth of copyright as a tool for regulating technology, is somewhat better represented in major newspaper coverage, but the volume of coverage has been so low as to keep the DRM policy debate off the mainstream political agenda. In stark comparison to both of these offline media, the web's collective treatment of the DRM policy debate is remarkably one-sided toward the strong fair use coalition. Advocates for strong fair use have taken advantage of their superior numbers and the web's low barriers to entry to create a space in which their opinions define the issue.

The growth in online advocacy and in the strong fair use coalition's success in offline presence coincides with a change in the momentum of copyright policymaking. The 1990's saw the rapid expansion of copyright protection for DRM systems, while the

first years of the 21st century saw a halt in such expansions and a significant push to roll back earlier changes. Without making formal claims about causality, this dissertation strongly suggests that the strong fair use coalition has had at least some impact in reshaping the politics of copyright over the last 15 years.

I begin this chapter by looking at the present and future of copyright politics, evaluating the developments that have come to pass since 2006 and suggesting where each coalition is headed. Next, I evaluate the theoretical foundations of this research. Third, I suggest additional applications of this study's methods for studying online communication generally. Finally, I note some of the study's limitations and suggest avenues of future research.

The Politics of Copyright and DRM: Updating and Looking Forward

While this dissertation tells the story of the rise of the strong fair use coalition, the more recent past suggests a modest trend in Congress back toward the strong copyright side. This is due to the outcomes of larger political forces exogenous to the copyright subsystem. Copyright is not a particularly partisan issue; both coalitions have congressional allies in each party, though the strong copyright coalition has more allies in each party. Despite this, and somewhat counter-intuitively given Republicans' comparatively stronger belief in property rights, the shift toward Democratic leadership in Congress represents an exogenous setback for the strong fair use coalition. The change in the specific members of Congress who head the relevant committees harms their cause.

Since the Democrats won a majority in Congress in 2006, strong fair use coalition ally Republican Joe Barton is no longer running the House Committee on Energy and

Commerce. Thus, the strong fair use coalition's most powerful ally from 2004 to 2006 is no longer in a position to determine the hearing schedule and consider friendly legislation. Henry Waxman (D-CA) now chairs the committee. Waxman is not particularly well known for strongly held views on copyright—for instance, he authored no documents included in this study, despite being in Congress since 1975—but there is at least one good reason to believe he is more likely to side with the strong copyright coalition: his district includes a substantial portion of west Los Angeles, so a large number of his constituents are in the entertainment industry.

The other three relevant committee chairs are unlikely to step into Barton's shoes and become powerful allies of the strong fair use coalition. The new chair of the House Committee on the Judiciary, John Conyers, is a reliable member of the strong copyright coalition whose two documents in this study both supported stronger DRM regulation. Conyers was the 2008 recipient of the Business Software Alliance (BSA) Cyber Champion Award, granted because he "played an instrumental role in the passage of legislation to protect intellectual property and fight cyber crime" (Business Software Alliance, 2008). While not part of the DRM debate, Conyers has sponsored two bills that would expand copyright (*Fair Copyright in Research Works Act*, 2009; *Performance Rights Act*, 2009) within the first few months of the 111th Congress, strongly suggesting that his allegiance lies with the strong copyright coalition.

The chair of the Senate Committee on the Judiciary, Patrick Leahy (D-VT), is also a reliable supporter of stronger copyright law. For instance, just one month into the 111th Congress, Leahy sponsored a bill to mandate that radio stations pay additional

copyright royalties (*A bill to provide fair compensation to artists for use of their sound recordings*, 2009), the companion bill to the Conyers proposal introduced in the House (*Performance Rights Act*, 2009). Even Leahy's 2008 letter opposing the secretive, fast-track negotiation process and the proposed content of the Anti-Counterfeiting Trade Agreement (ACTA)—including escalated internet service provider liability for copyright infringement and possibly mandatory content filtering—does not suggest a sudden change of heart. Leahy's letter, co-authored with Arlen Specter (R-PA), reflects more concern about preserving congressional flexibility in light of future technological developments than a shift to support for the strong fair use cause (N. Anderson, 2008). The final chair of a relevant committee, the Senate Committee on Commerce, Science, and Transportation, is Jay Rockefeller (D-VA). He has no well-established track record on copyright law, but it seems unlikely that Rockefeller will step into Barton's shoes and provide a friendly venue for the strong fair use coalition.

President Obama also has no obvious leaning toward either coalition, though he has not given the strong fair use coalition much hope for a radical turnaround from the strong copyright sympathies of the Clinton and Bush administrations. On the contrary, at least one recent decision suggests they intend to continue in the same direction. Knowledge Ecology International (KEI) filed a Freedom of Information Act (FOIA) request for key documents related to the international negotiations on a proposed Anti-Counterfeiting Trade Agreement, a title that may not highlight the proposed treaty's broad reach into many areas of copyright law. KEI director James Love alleges that these documents have been "widely circulated to corporate lobbyists in Europe, Japan, and the

US," but his FOIA request was rejected on the grounds that the documents have been "classified in the interest of national security" (Love, 2009). Especially when shrouded in such stubborn secrecy, this looks like another round of the kind of policy laundering that helped yield Title I of the Digital Millennium Copyright Act, or DMCA (Herman & Gandy Jr., 2006).

While not nearly as significant as the Audio Home Recording Act (AHRA) or DMCA, the strong fair use coalition has suffered an additional legislative setback on DRM policy since the close of this study's time frame. Buried deep inside an enormous education bill that became law (*Higher Education Opportunity Act*, 2008) is a brief provision pushing college network administrators toward purchasing licensed means of downloading copyrighted works and filtering the content sent across their networks. The new law requires institutions of higher learning to certify that they have

> developed plans to effectively combat the unauthorized distribution of
> copyrighted material, including through the use of a variety of technology-based
> deterrents; and will, to the extent practicable, offer alternatives to illegal
> downloading or peer-to-peer distribution of intellectual property, as determined
> by the institution in consultation with the chief technology officer or other
> designated officer of the institution. (20 USC § 1094(a)(29))[96]

The demand that colleges move toward implementing "technology-based deterrents," generally interpreted to mean network-based filters to prevent the transmission of

---

[96] In H.R.4137, the bill as sponsored by Rep. George Miller (D-CA), this was in § 493(a)(1)(ii).

copyrighted works, is particularly objectionable for the strong fair use coalition. These critics object that DRM systems always inadvertently stop noninfringing uses of copyrighted works such as fair use. The copyright-specific addition to the bill was strongly opposed by several members of the strong fair use coalition, but it was too small a part of a bill that was too important not to pass. Once the strong copyright coalition succeeded in getting the amendment added to the bill, it was practically a done deal.

In addition to the minor loss as embodied in the Higher Education Opportunity Act, the strong fair use coalition is also unlikely to see the advancement of its most treasured potential reform, the loosening of the DMCA. While such proposals were already well established as unlikely to pass, the odds of passage may be nearly zero in the near future. The idea has been proposed repeatedly and failed consistently enough that it is unlikely to be taken seriously in the near future. In the 110th Congress, Representatives Rick Boucher, John Doolittle, and Zoe Lofgren offered a much milder version of a DMCA reform bill (*Freedom and Innovation Revitalizing U.S. Entrepreneurship Act of 2007*), and even that much more modest proposal failed to gain momentum. For instance, while 21 representatives added their names as sponsors, four of them later withdrew their sponsorship. As discussed in Chapter Two, policy ideas have a limited shelf life before other issues grab policymakers' attention, so the push to reform the DMCA will likely languish for a while until it can be reintroduced seriously.

Despite the harder challenges now facing those calling for stronger fair use, however, there is reason for them to be optimistic about the future—and for the strong copyright coalition to continue to be concerned about their ability to sustain recent gains

and advance additional proposals. First, this dissertation illustrates that the strong fair use coalition now has a well-established seat at the negotiating table. Barton's committee was not the only vehicle for slowing down proposals such as the broadcast flag and for advancing calls to reform the DMCA. Frequent appearances in hearings and news coverage reflect the institutionalized role of strong fair use advocates such as NGOs and scholars. Thus, the AHRA-era time of a virtual vacuum of opposition to DRM regulation, or even the DMCA-era in which the multi-sector organization of a strong fair use coalition was just getting started, seems like ancient history. Further, their total domination of the internet space is a sign of a reasonably large issue public that can be mobilized to contact Congress or engage in other policy actions.

A second cause for suspecting a potential sea change in copyright policy in the medium term is the potential outcome of economic change. Just as exogenous political events can affect a policy subsystem, so can exogenous economic events. The very large media companies at the heart of the strong copyright coalition face a number of challenges in this environment, and widespread infringement is just one of them. The widespread adoption of internet-connected computers has opened up a new world of possibilities for small-scale media producers to produce, place, and distribute their wares, possibilities that were not open to them ten or fifteen years ago (Benkler, 2006).

Insurgent companies with business models better suited to the new media environment continue to take business away from larger companies. Independent music labels in particular have done very well in this environment, seeing market share increase as fans do a lot of their marketing for them (Macworld, 2004). Faced with loss of revenue

323

when consumers choose downloading, loss of market share to indies and amateurs, bad publicity and ill will resulting from thousands of lawsuits, and a major recession, the major labels are closing divisions, laying off thousands of employees, and struggling to survive (Morelli, 2009). Movie studios and software companies are not in such imminent peril, but similar forces are affecting them. BitTorrent makes it feasible to share high definition video and 20-gigabyte design suites. Independent moviemakers, free software, and "Web 2.0" applications have made real gains in market share. If the profits of major record labels, movie studios, and proprietary software firms fall to a fraction of their current levels in the future, that will put a substantial strain on the resources available to the strong copyright coalition. If this combination of economic forces drives much of the concentrated capital out of these industries, the strong fair use coalition may be able to force a more favorable outcome into the law books.

Finally, while the entire debate has simmered well below the top of the public agenda, the whole policy subsystem is subject to a wide range of possible outcomes if the right event or series of events brings widespread public attention. For instance, the cases of Ed Felten (Lessig, 2004, pp. 155-157) and Dmitry Sklyarov (Richtel, 2002) were unmitigated public relations catastrophes for copyright holders, but this was at a time when the strong fair use coalition was still in its adolescence. Public Knowledge was not even founded until 2001. If a similar story happened today, well-established advocates whose phone numbers are in reporters' hands would spring into action, and the well-rehearsed story of copyright holders as villains could catch fire and bring substantial change. The subsystem like any policy subsystem is subject to substantial and sudden

324

change during an unexpected, sharp upswing in attention from the general public; if that attention comes, a sharp divergence in policy is a real possibility.

Theoretical Significance and Further Applications of Method

In Chapter 2, I draw upon and synthesize a great number of theories of policy process and political communication, hoping to say something about communicating policy actors—that is, about how policy advocates strategically use communication across multiple media. Such a theory is especially important in light of new media technologies such as the internet that substantially reduce barriers to entry into the policymaking process. In an era when communicating with policymakers and the public was substantially more expensive, it was not as important conceptually to separate communication from resources and access to policymakers; communicating required resources, and resources were generally closely related to access, so only those coalitions with money needed to be taken seriously. Now, the internet has given a substantial relative boost to groups with less financing, especially when those groups have or can recruit a substantial number of motivated sympathetic citizens. This means that the communication of policy ideas takes on a new significance that is less substantially tied to a coalition's financial and political capital. Seen in light of the recent developments in media technology over roughly the last fifteen years, the theories of political strategy and political communication invoked in this project do a good job predicting the behavior of the policy coalitions in the DRM policy debate.

The governing or dominant coalition, the strong copyright coalition, behaves largely as one would expect based on their incentives and resources. They have the

money and access to communicate directly with policymakers, so they do so. They have

the automatic credibility to feature prominently in much of the news coverage, and they

take advantage. Yet they also have the incentive not to encourage too much press

coverage of the DRM policy debate, and the low volume of stories reflects this incentive.

If they do have the incentive, such as if DRM policy becomes a point of major national

discussion, they likely will use their credibility and resources to land substantially more

DRM policy-related press coverage. Finally, they have little incentive to expand the

scope of conflict to include the general public and turn DRM policy into an electoral

issue. Thus, the strong copyright coalition makes little use of the internet—the medium

with which they could speak most directly to a potentially engaged issue public and least

directly to policymakers.

In contrast, the strong fair use coalition's behavior reflects the expected behavior

of an under-resourced challenger coalition. They constantly seek to expand the scope of

conflict as wide as possible, seeking to include the mainstream press and the general

public, by press outreach and by their heavy subsidization of the web debate over DRM

policy. This reflects their position as a challenger coalition with inferior access to

policymakers and a desire to effect substantial reform; these policy actors know that, if

the debate stays in the House and Senate judiciary committees and the US Copyright

Office, no substantial change will come about. They also take advantage of their

numerical advantage in mobilized individuals to leverage the web in a way that helps

ameliorate the effect of their substantial resource disadvantage.

Based on the networking theories that predict a power law distribution of hyperlinks (Barabási, 2003; M. S. Hindman, 2009), the sites included in this study are surely just the best-linked members of a large population of websites that are partially or wholly devoted to the debate over copyright law. This reflects my experience as both consumer and producer of such information. I have personally visited hundreds of blogs related to copyright law. While my experience reaffirms the study's conclusions that the sites at the center of the Issue Crawler graphs really are the central loci of the online debate, the wealth and diversity of sites that did not make this list is astounding. In my own small way, I am part of this collective online extension of the strong fair use coalition. The blog I founded (B. D. Herman, 2005), which has since migrated and become a group blog co-authored with several other University of Pennsylvania colleagues (B. D. Herman, Karpf, Tocci, & Tsui, 2006), is just one of what must be thousands of examples. We all know that we are deep into the long tail (C. Anderson, 2004), but as the author who writes most regularly about copyright, knowing that I might help even a few people to learn more about the copyright debate is part of my motivation. Judging by the identities of our comment authors, most of our audience appears to visit because they know us personally or because we link to them. Obviously, this represents a microscopic share of the total online audience.

The average copyright activist blogger's goal is not single-handedly to bring the strong fair use coalition to victory, but if a blog gets fifty people to think differently about copyright and five people to call their congressperson to voice the coalition's views, and if a few thousand other advocates do the same, this represents a substantial collective

impact. While few authors can prove direct and powerful effects of their personal writings, the strategy of distributed, collective action seems to have made quite an impact. Judging by my students, a substantial minority of young people seem to know a surprising amount about DRM and copyright law that is not available in mainstream news sources, and this includes people who themselves have not heard of the Electronic Frontier Foundation, let alone visited their webpage. The hub sites produce a reliable information subsidy, and thousands of sympathetic part-time activists amplify this information—sometimes several generations away and without explicitly seeking to mobilize an issue public.

Most of the other sites have a small audience, but strong fair use coalition supporters also carry these views into highly visible hub sites that are not specifically targeted to the copyright policy debate—sites like Slashdot and Digg—as well as social networking sites such as Facebook, MySpace, and Twitter. Few of these people can claim their own substantial audience, but collectively, they are possibly changing minds and laying the groundwork for the chance to effect more dramatic change. This is not the quickest way to get change, but it is remarkably cost-effective, and if and when copyright bursts forth as an electoral issue, the strong copyright coalition may face a mobilized, substantial issue public, springing into coordinated action in the name of fair use.

Anecdotal suggestions of effect aside, the strong fair use coalition's continued investment in online activism suggests that they believe it is an effective use of time and energy. Further, the high number of individuals and not-for-profit institutions at the core of the online debate reaffirms the role the internet plays in helping under-resourced

coalitions to have a voice. By way of contrast, consider the Consumer Electronics Association, or CEA. The group is an incredibly important member of the Washington, DC-based wing of the strong fair use coalition. They appear in Congress with reliable frequency, and they are the most reliably strong fair use trade group in DC today. Just like the EFF, the CEA has every incentive to expand the scope of conflict, so their online presence (28 relevant documents) is double that of the MPAA (14) and many times greater than that of the RIAA (1).

Yet the volume of CEA-hosted relevant information and their site's link share (28 relevant documents, 0.8% share) hardly puts them at the center of the debate. Compare this to the EFF (81 documents, 9.4% share), which is central to the online debate on an NGO-sized budget. The CEA apparently calculates that since they have direct access to policymakers, assured by the size of their industrial backers, these resources are better spent there. They have the resources to lobby policymakers directly; spending too much of their intellectual capital online takes away from this core strength. An NGO, however—even one in Washington, DC—has more incentive to stretch scarce resources by amplifying their message online. Unlike the CEA, Washington, DC-based nonprofit groups like Public Knowledge (74 relevant documents, 2.4% link share) and the Center for Democracy and Technology (44 documents, 5.6% share) cannot count on the economic weight of an industry to open doors for them. Thus, they go online, producing more content of more direct relevance to the debate than the CEA.

While these empirical results buttress the predictions about how communicating policy actors will use different media, there is far more work to do. The copyright debate

329

presents a particularly neat dichotomy between a well-resourced governing coalition, backed by a large industry, and an under-funded coalition that is primarily made up of nonprofit groups. As noted in the next section, it is worth using these methods to study similarly constituted policy spaces to see how well this particular model holds. For instance, are environmental groups worried about global warming particularly effective at dominating the online debate relative to the well-resourced industrial groups, such as the oil and gas industry, that oppose restrictions on carbon emissions? If this and other policy systems look similar, such an outcome suggests that the internet matters similarly to many or perhaps most under-resourced challenger coalitions.

Yet even such confirmatory results would be just the beginning of developing an acceptably complete theory of the communicating policy actor, which will need to be expanded and refined to account for the large array of possibilities for how a subsystem might look. How might things look different if any of these details change? The governing coalition might behave differently if they are composed primarily of ideological rather than economic groups. There are a few issues in which industrial groups are on the outside looking in and the governing coalition is composed largely of public interest groups. The debate over nuclear power comes to mind, though there are likely industrial interests on both sides. Do these industrial groups trying to change the status quo use the internet in a way that better reflects their positions relative to the policy in place now, or in a way that reflects their institutional cultures and bases of support? Further, there may be issues on which major changes to the status quo might best be

effected quietly because those pushing for reform are concentrated industrial interests who face wide public opposition to their proposals.

There are also issues for which the industrial versus public interest dichotomy simply does not fit. On one hand, some debates feature opposing coalitions that split primarily on partisan or ideological lines. The cliché example is abortion. It seems unlikely the pro-choice coalition would fail to fully use the internet to maximize public sympathy with their cause just because the status quo is closer to their goals, and while both sides may be composed primarily of NGOs, several are certainly better funded than some industry trade groups. Likewise, some debates, such as the debate over airline deregulation in the 1960s and 1970s (Brown & Stewart Jr., 1993), feature two or more coalitions composed primarily of well-heeled industry groups. In such subsystems, do any coalitions use the internet to a substantial degree?

This dissertation focuses on a debate with two coalitions that have very different organizational cultures, funding, and responsibility for current law. On each point, there was every reason to believe the strong fair use coalition would make heavy use of the internet, and the strong copyright coalition would act based on their superior access to policymakers. That is exactly what transpired and what will likely continue to happen. The theory is less clear about what to expect once these forces fall out of alignment. Pending further study, one must also consider the possibility that this study's results may be policy specific. The dedicated use of the internet by the strong fair use coalition is particularly likely given that coalition's technology-savvy members and strong belief in new technology. Future studies using similar methods to study strategic communication

331

within structurally diverse policy subsystems could help place these results in a broader context.

At a broader level, this research suggests new directions for political communication and policy studies. While this is not the first study to examine strategic communication within policy subsystems, such studies are still in short supply. Most scholars of political communication study electoral politics, and few policy studies scholars systematically examine communication. Yet as this study suggests, a great deal of the excitement of the internet is in the role it can play in reshaping the dynamics of the policy process. Other political resources such as financial capital and access to policymakers are still crucially important, but new technology has reduced their relative importance by giving policy advocates a new way to communicate with other coalition members and to mobilize an issue public. To properly study the exciting changes afoot, we would do well to expand the areas of overlap between political science and communication studies to include the systematic study of the strategies and outcomes of communication in policy subsystems.

Further Applications of Research Methods

While the methods in this dissertation are directly targeted at studying communication within policy subsystems, this is not the only potential application for the methods developed herein. In particular, this dissertation offers a reasonably systematic means for identifying and content analyzing online documents with an eye toward comparing them to offline documents. Databases such as news archives are well understood and have served as the basis for countless studies of media coverage, but

online researchers have still struggled with methods for forcing the unruly web to cooperate with research methods designed for more predictable source materials.

This dissertation's application of the Issue Crawler tools developed by Rogers (2004) was successful enough to warrant similar use by scholars similarly interested in manageably sized topical clusters of websites. Repeated crawls in this study worked well enough that, as long as one knows the terrain well enough to know the sites from which to start and to have a good idea about the quality of the results, this kind of long-term repeated querying is a good tool for identifying the bulk of relevant, important sites and estimating their share of incoming links. It also had the effect of smoothing over some of the potential noise in the data, such as when USA.gov inexplicably dominated the link share for two crawls.

This dissertation also made a contribution to the combination of web crawling and content analysis, initially developed by Farrall and Delli Carpini (Farrall, 2005; Farrall & Delli Carpini, 2004), specifically with an eye toward creating results that could be directly compared to offline media. One of the most substantial problems with internet research is identifying a set of documents that can credibly represent a larger population. Yet, given the power law distribution of hyperlinks, if one can identify the central sites in a related network, one has identified the sites that carry the content that has earned the links of the hundreds or thousands of minor sites that link into these hubs. This method does just that, and it then uses these sites as sources of content that can be analyzed much like textual content from older media such as newspapers.

333

The Rogers/Farrall and Delli Carpini method solves the problem of identifying the most central sites in a population of sites dedicated to an issue, but another problem is the difficulty of identifying and quantifying the relevant documents on each of these sites. Thankfully, Google is a free online tool that, with a little coaxing, can solve this problem. Starting from an established method for validating database search terms (Stryker et al., 2006), this dissertation offers a means for developing and evaluating search terms to extract a set of relevant documents from each of a group of websites. Because Google is opaque as to both their search technology and the times at which this technology has changed, and because the web changes substantially on an ongoing basis, this is unfortunately not a good method for strictly reproducing similar results at different times. It is, however, a good method for extracting comparable results across a modest number of websites if all the searches are done in short order. Combined with a simple procedure for estimating the search terms' recall and precision, this method provides a good tool for gaining a basic understanding of an issue across websites at one time. Further, it is powerful and accessible enough that repeated searches over a period of time may prove quite useful for studies that can tolerate a high margin of error.

Especially when taken together, these methods are applicable to a broad range of potential research problems. One can study the content on a group of websites as it relates to almost any range of possible issues—so long as those websites cluster together and the issues are amenable to searches on Google. Further, one can compare the online and offline coverage of an issue or a series of related issues, whether or not the topics are explicitly political.

Limitations and Suggestions for Future Research

This study answers many questions, but it leaves further questions still unanswered. This study's most regrettable limitation is the lack of interviews with policy actors. Several questions that remain unanswered would be far less mysterious after some in-depth interviews with those who took part in the history I describe here. Here are just a few of the potential areas of investigation: In the eyes of each coalition, how important is the internet relative to other communication media? How much communication with policymakers happens in private meetings, and what are the differences between private meetings and committee hearings? Do policy actors in the consumer electronics industry consider themselves to be in the strong fair use coalition, or do they represent a third group somewhat between the other two? These and dozens more questions—several of which might become apparent during interviews—call for a follow-up study using semi-structured interviews with policy actors.

Further, this study does little to establish the effectiveness of each coalition's communication strategy. This shortcoming opens a world of possibilities for future research. One avenue is the study of public opinion. One could conduct survey research to study the understanding of and opinions about copyright among the general population. Since online communication is incredibly one-sided toward the strong fair use coalition, one could estimate the effectiveness of these messages by looking for a relationship between sympathetic views and consumption of internet sources likely to voice these sympathetic messages. Taking one's cue from Zaller (1996) and taking advantage of the fact that DRM policy is not well-covered in offline media, one could

335

reach a reasonable measure of consumption of sympathetic messages in part by asking respondents basic factual questions about the DRM debate and the copyright subsystem.

Also, this study deliberately omits any consideration of the effectiveness of a given message; a timid but categorical document counts the same as a persuasive, bold document filled with powerful arguments. One could test the effectiveness of various copyright-related messages using laboratory experiments and focus groups, increasing the precision through which one can estimate the significance of these messages as deployed in context.

Finally, thanks to a lack of resources, this study falls short of fulfilling its own demand to focus on policymakers as an audience. In principle, congresspersons and congressional staffers should be amenable to study via the same methods used to study other members of media audiences, but in practice, their time and attention is so much more valuable that cooperation is presumably quite difficult to obtain. This does not mean that a researcher should not try, but it means that one must be, become, or develop contacts with the kind of person with intimate access to policymakers. Surveys require large numbers of participants to be meaningful, which may be hard to obtain from such a concentrated group of powerful people.

It is more realistic to try for the level of access required for semi-structured interviews and participant observation, which is somewhat realistic even for graduate students and junior faculty. Congresspersons themselves can be quite hard to reach, but congressional staffers are a good deal more accessible and have quite a good understanding of how policy issues get decided. Interviews with policy advocates

combined with interviews of policymakers—reasonably taken to include congressional staffers, who often make or suggest the specific policy choices made—could provide a much more complete sense of how communication across multiple media shapes policy decisions. This is true of the copyright debate specifically and of policymaking generally.

Conclusion

Prior to this study, there was little scholarship connecting the literature on policy systems to that on political communication, and absolutely none tying both traditions to the evolution of copyright and DRM policy. Through this combination, this dissertation offers additional insights into the nature and role of communication across multiple media in the DRM debate. Results suggest that the strong copyright coalition has historically enjoyed superior access to policymakers, but this advantage has diminished over time as the strong fair use coalition has organized and gained momentum. Results also suggest that the press coverage has leaned slightly toward the strong fair use coalition in tone, but coverage has been light enough to help the strong copyright coalition. Finally, the internet has presented the strong fair use coalition with a unique opportunity to spread their message virally with little financial cost, and they have maximized their use of this opportunity. They provide high volumes of relevant, regularly updated information subsidies for the rest of the web, and they have come to dominate the center of the online debate.

These results provide substantial support for the theory of communicating policy actors as developed in this dissertation; both coalitions make their respective strategic communication decisions as predicted. This theory merits further expansion and

337

refinement via study of other policy subsystems, both those that are structurally similar and those that are different. This study's methods also merit further exploration. The combination of methods for studying web communication in particular could prove useful in studying a range of possible issues, not just the online communication of policy actors. Further, by establishing a method for comparing online and offline communication, this dissertation opens up new avenues for further comparing and contrasting the two in other contexts.

The brief history of the debate over copyright as a tool for the regulation of DRM features many developments, but the most important development is the birth and growth of the strong fair use coalition as we know it today. As NGOs and scholars have joined and sharpened the efforts of fair use-minded technology groups, those advocates calling for less regulation of DRM have moved from the periphery to the bargaining table. Their communication strategies say a great deal about the copyright debate, and the copyright debate says a great deal about strategic communication within policy subsystems. With further expansion of these theories and research methods, additional interesting results are sure to follow.

APPENDIX A: A PRIMER ON COPYRIGHT

In this appendix, I provide a brief primer on some of the basics of copyright law so that even readers who know little about copyright will understand the rest of the dissertation. For a highly accessible introduction to the basics of copyright, see Fishman (2003). In this appendix, I first consider the scope and duration of copyright. Second, I delineate some of the notable exceptions to copyright law. Third, I provide a very brief introduction to copyright as a vehicle for the regulation of media technologies.

Scope and Duration

"Copyright is a legal device that provides the creator of a work of art or literature, or a work that conveys information or ideas, the right to control how the work is used" (Fishman, 2003, p. 2/2).[97] The last major overhaul of US statutory copyright law was in 1976 (*Copyright Act of 1976*, 1976). Since the 1976 Act went into force in 1978, copyright automatically applies to nearly any creative work that is "fixed in any tangible medium of expression" (17 U.S.C. § 102). This includes written text, musical scores, choreography, photographs, visual art, movies, paintings, recorded audio, and architectural works. It does not cover the underlying concepts, though these may be subject to patent protection. It also does not cover very short phrases, such as the old Ford Motor Company slogan, "Quality is job 1," though these may be subject to trademark

[97] Fishman's book is paginated by chapter, so page number "2/2" represents chapter 2, page 2.

protection. See Elias and Goldoftas (1996) for more on the differences between copyright, patent, and trademark. Copyright also does not apply to non-creative compilations of facts, such as the information in a telephone book (*Feist Publications, Inc., v. Rural Telephone Service Co.*, 1991) or the raw data in a database. Finally, no creative works produced or owned by the US Government are subject to copyright protection. Except for these caveats, nearly everything captured in a tangible medium is automatically copyrighted, from a toddler's scribbles to a CEO's email.

Subject to certain exemptions, copyright provides authors with a limited monopoly on six important rights:

> (1) to reproduce the copyrighted work in copies or phonorecords;
> (2) to prepare derivative works based upon the copyrighted work;
> (3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;
> (4) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly;
> (5) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and
> (6) in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission. (17 U.S.C. § 106)[98]

The author can then sell the exclusive or nonexclusive rights to do any or all of these things. Additionally, a firm can hire an author to produce a work; under most circumstances, the firm then automatically becomes the owner of this "work for hire" (see Fishman, 2003, pp. 8/4-8/18).

---

[98] US statutory copyright law is coextensive with Title 17 of the US Code. Unless otherwise noted, all statutory references are to the Copyright Act, including amendments.

In practice, large media companies hold most commercially valuable copyrights, and they become the "authors" who exercise the monopoly powers granted by copyright law. One who does any of these things with a copyrighted work without the express permission of the copyright holder is infringing copyright law, unless their actions qualify for an exemption.

Unlike some legal rights—such as the right to property—copyright is not indefinite. For a work created by an individual after 1977, copyright lasts for the life of the author plus 70 years. In the case of works-for-hire, copyright lasts for 95 years from the date of publication or 120 years from the date of creation, whichever comes first. Many works produced between 1923 and 1976 are also subject to long periods of protection, generally 95 years (Gasaway, 2003), though the details are too minute to recount here. Any work on which the copyright has expired or in which copyright never inhered, such as US Government documents, is in the public domain. Anyone can reprint and sell public domain works, such as exact copies of Shakespeare's works or federal court case decisions, with impunity.

The scope of copyright and the length of time for which a copyright lasts have grown dramatically since the first copyright statute (*Copyright Act of 1790*, 1790). Copyright originally applied only to books, maps, and charts; protection for other types of creative works, from music to visual art, would come later. The original term of copyright was 14 years with an option for a 14-year renewal; no work was under copyright for longer than 28 years. Until 1976, one had to register to obtain a copyright—it did not automatically apply to any creative work fixed in a tangible medium, as is the

case today. In comparison with today's copyright, the first copyright statute was extremely limited.

Exemptions and Defenses

As noted above, copyright does not apply to absolutely everything that is fixed in a tangible medium. Even if it does apply, there are a number of statutory limitations and possible defenses, permitting users of copyrighted works to engage in otherwise infringing activities such as copying or public performance. The list of statutory exemptions and mandatory licenses is quite extensive, so I recount but a few. Once a customer buys a physical copy of a copyrighted work, she may loan it to a friend, give it to a library, or resell it (§ 109). Libraries and archives are given some latitude in making copies of a work that they already own or previously owned but lost due to causes such as damage or theft (§ 108). Music is subject to a robust, complicated scheme for compulsory licensing (§ 115), covering activities such as radio airplay and recording of cover versions of songs. Educators have an exemption for the display of copyrighted materials in the classroom or, in the case of distance learning, the transmission of such works online (§ 110).

In addition to the mandatory licensing and statutory exemptions, there are a number of affirmative defenses to the charge of copyright infringement; even after a copyright holder has demonstrated an unauthorized use of a copyrighted work, a defendant has a number of potential legal arguments that can stop a finding of infringement. By far, the most important of these is fair use. The section that enumerates fair use reads, in part:

The fair use of a copyrighted work, … for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include —

(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;

(2) the nature of the copyrighted work;

(3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and

(4) the effect of the use upon the potential market for or value of the copyrighted work. (17 U.S.C. § 107)

Under fair use, a student may photocopy a chapter of a book for personal study, but he may not photocopy an entire book. A jazz band may borrow two measures of a copyrighted song without getting permission or a compulsory license, but they may not use the defense of fair use in creating a new arrangement of an entire song. A scholar may quote extensively from another's article as part of a scathing critique, but she may not reproduce another's article, claiming it as her own work, without committing both plagiarism and copyright infringement. In practice, these examples of fair use and infringement are clear enough that common sense and copyright law match quite well.

Despite some obvious examples of fair use and, in contrast, infringement, the doctrine of fair use leaves open a substantial gray area into which many activities will fall. One much-discussed example is the digital sampling of musical recordings. Many have argued that short samples should generally be seen as fair use (McLeod, 2001; Negativland, 1995; Vaidhyanathan, 2001). In the only clear ruling on the issue of musical sampling, a three-judge panel on the 6th Circuit provides quite a powerful argument for those who would argue that sampling is instead infringing. The court held that the section

106 exclusive right to prepare derivative works settles the question of unlicensed sampling:

> In other words, a sound recording owner has the exclusive right to "sample" his own recording. We find much to recommend this interpretation. To begin with, there is ease of enforcement. Get a license or do not sample. We do not see this as stifling creativity in any significant way. It must be remembered that if an artist wants to incorporate a "riff" from another work in his or her recording, he is free to duplicate the sound of that "riff" in the studio. (*Bridgeport Music, Inc. v. Dimension Films, et al.*, 2005, pp. 657-658)[99]

Rather than settling the debate, the decision has only sparked further fighting over the topic of musical sampling (Somoano, 2006). While the issue of musical sampling is distinct from the debate over digital rights management, it highlights the extremely debatable nature of fair use. Countless other examples, from unlicensed collegiate course packs to the reproduction of news clips for use in documentaries, bring out strong, contradictory opinions from both sides of the copyright divide. The debate over which activities should be excused under the doctrine of fair use is a central part of the debate over the proper scope of copyright law.

<div align="center">Copyright as Regulation of Media Technologies</div>

Over two decades ago, the courts declined to turn copyright law into a vehicle for regulating media technologies. Since that time, however, both Congress and the courts have begun to reshape copyright such that now it does play an important role in such regulation. In the famous *Sony* decision (*Sony Corp. of America v. Universal City*

---

[99] Citations omitted. While the *Bridgeport* court did not hold *per se* that sampling is not fair use, the decision's impact is captured in the above quotation—particularly, "Get a license or do not sample."

Studios, Inc., 1984), the Supreme Court ruled that Sony should not be held liable for their

customers' infringing uses of the company's videocassette recorders (VCRs). The court

held that the VCR is capable of and used for many noninfringing uses, such as recording

a show to watch in a different place or at a later time, and that the company is therefore

not responsible for the behavior of some customers who may use their product for

infringing purposes.

Since *Sony*, however, at least three major policy events have signified the high

degree to which copyright has become a tool for the regulation of media technologies.

First, in 1992, Congress passed the Audio Home Recording Act, or AHRA (*Audio Home

Recording Act*, 1992). This bill sought to address music industry anxiety about the

development of digital audio recording technology for the consumer market—in

particular, the invention of digital audiotape (DAT). The music industry as a whole—

including both music publishers (songwriters and their agents) and the record labels—

were profoundly scared by the prospect of consumers having the ability to make perfect

digital copies. For more on the policy process surrounding and legislative specifics of the

AHRA, please refer to Chapter 5.

Second, in 1998, Congress passed the Digital Millennium Copyright Act (*Digital

Millennium Copyright Act*, 1998b), including as its first title the WIPO Copyright and

Performances and Phonograms Treaties Implementation Act. This title provides new

legal protections for technologies designed to provide additional protections for

copyrighted materials. These are broadly classified as digital rights management (DRM),

and in most cases, the DMCA prohibits the act of circumventing DRM, as well as

developing and distributing tools that are capable of circumvention. For more, please see Chapter 5.

The third major event marking copyright's venture into the regulation of technologies is the Supreme Court's decision in *MGM v. Grokster* (*MGM Studios, Inc. v. Grokster, Ltd.*, 2005). Grokster and other peer-to-peer (P2P) services make it easy for anybody with an internet connection to exchange files, whether or not those files are subject to copyright. While observers disagree about the percentages, a substantial portion of the content traded on these networks represents the unauthorized transmission of copyrighted works, particularly music and movies. Not every downloaded file represents a lost sale, but users who may otherwise have purchased a work obtain a substantial number of files for free instead. Users often insist that they use P2P trading as a means of finding new products to buy—that this trading is primarily a form of free advertising for musical acts, movies, TV shows, and the like. In the belief that the value of the free publicity is outweighed by the volume of lost sales, copyright holders—in particular, the music and movie industries—have unleashed a legal storm on those who develop and use P2P programs (Electronic Frontier Foundation, 2008, 2009).

As part of this legal strategy, a coalition of copyright holders, led by MGM, sued the developers of P2P programs Grokster, Morpheus, and KaZaA. The court ruled that the P2P companies could be held liable for the infringing conduct of their customers. While the decision was unanimous, the court issued three separate opinions that left the meaning of *Sony* somewhat unclear. At least one renowned copyright expert believes the court created a new theory of liability and "that it did so deliberately to kill off *Sony*"

346

(Patry, 2006). This reading of the ruling is debatable, but if nothing else, the ambiguity created introduces a new threat of liability that gives copyright holders another vehicle for discouraging or suppressing the development of technologies not to their liking. Even if a developer of a new technology may ultimately prevail in court, the very threat of a protracted and expensive legal battle is often adequate to persuade technological innovators to work with copyright holders; *Grokster* only accelerated this compliance. In addition to the statutory protections for DRM, *Grokster* accelerates the role that copyright law plays in the regulation of new technologies.

Conclusion

Copyright is a limited monopoly, granted to creators of original works fixed in a tangible medium. It prohibits many kinds of uses unless authorized by the copyright holder; these include reproduction, public display and performance, and the making of derivative works. Yet these monopoly privileges are subject to a wide array of limitations and exemptions, ranging from the specific to the very broad and highly contested. For instance, copyright is of limited duration, though the number of years between a work's creation and its entry into the public domain has grown dramatically since the first copyright statute. The broadest and most debated exemption is that of fair use, which invites courts to weigh the public interest value of a given unauthorized use of a copyrighted work. The debate over the scope of fair use is one of the most important contests in the tussle over copyright.

As new media technologies have made it easier to make unauthorized uses of copyrighted works—whether those unauthorized uses are fair uses or infringements—

347

copyright holders have developed a high and understandable level of anxiety over the breakdown of their former business models. They have sought to use a combination of legal and technological limitations to slow or reverse the trend toward the totally free distribution of their works. While some of their legal and political strategies center on the pursuit of unauthorized uses of copyrighted works, recent years have seen a substantial trend toward the regulation of media technologies that may be associated with such unauthorized uses.

APPENDIX B: CODING INSTRUCTIONS

This appendix contains the complete instructions for coding newspaper articles, internet documents, and documents from congressional hearings. While each of these kinds of documents has some unique characteristics, the stages of coding are almost identical for each. They are:

1. In the case of congressional documents, establish that a legislative hearing is relevant to one or more of the legislative debates under study.

2. Establish that a document itself is relevant.

3. Identify certain basic characteristics about the document, including its source, date, and type of policy actors to whom it may most directly be credited.

4. Mark the document as taking one of three positions on the debate at hand: the *strong copyright* position, the *strong fair use* position, or some kind of mixed or neutral position.

5. If the document supports either the *strong copyright* or the *strong fair use* position, coding is complete for that document.

6. If and only if the document is mixed or neutral, identify the relevant paragraphs.

7. Mark each relevant paragraph as taking one of three positions on the debate at hand: the *strong copyright* position, the *strong fair use* position, or mixed or neutral position.

The instructions in this codebook are generally presented in this order. First, though, it contains a basic outline of the four legislative debates under consideration.

Coders are highly encouraged to read carefully the primer on legislative debates and, where necessary, seek additional background information on copyright law generally and the regulation of digital rights management technologies specifically. With that background and a broad understanding of the way in which the stages of coding fit together, each of the other sections of the codebook should stand on their own.

Primer on Legislative Debates

This project studies the debate about the extent to which federal copyright law should govern digital rights management (DRM), also known as technological protection measures (TPMs). DRM systems—most commonly encryption, though also including other tools such as watermarking—are designed to reduce the end user's ability to make infringing uses of copyrighted content.

DRM is generally vulnerable to circumvention by highly skilled end users. Once the DRM is circumvented, the copyrighted work is available for copying and distribution, whether fair uses (e.g., moving the songs from a CD to an MP3 player) or blatant piracy (e.g., copying DVDs for illegitimate resale on the street). Further, once one user has circumvented a DRM system, she may then share those instructions with other end users, permitting them to release still other copyrighted works from their DRM systems. The goal of regulating DRM is to discourage acts of circumvention and to reduce the availability of circumvention-enabling technologies, with the stated goal of reducing the number of infringing copies made available via circumvention.

This dissertation focuses on the legislative process that led to the passage of two bills, the 1992 Audio Home Recording Act (AHRA) and Title I of the 1998 Digital Millennium Copyright Act (DMCA). It also tracks the debates over two bills that were debated heavily during the 108th and 109th Congresses (2003-2006) but did not pass during that time. This primer provides a short description of each debate. Dates indicate the four-year period during which each proposal was most seriously considered.

Audio Home Recording Act (AHRA), 1989-1992

The AHRA was a legislative reaction to the recording industry's fear of the Digital Audio Tape (DAT) deck, developed by Sony in the 1980s. Unlike analog cassette tapes, DAT technology allows for a nearly infinite number of perfect clone copies. The recording industry therefore initially sought to prevent the importation of DAT machines. They held two sizable threats over the electronics manufacturers. First, they could refuse to release music in DAT format, reducing the consumer market for the new machines. Second, they threatened to sue manufacturers for "contributory infringement," hoping to impose legal liability on Sony and other DAT manufacturers for the potentially infringing uses of their products.

The AHRA was a legislative compromise between the recording industry and the electronics manufacturing industry. The legislation requires that all consumer-grade digital audio recording devices implement the Serial Copy Management System (SCMS), which permits end users to make unlimited copies of an original recording but not copies of copies. It also imposes royalties on recording devices and blank media, collected by the US Copyright Office and redistributed to copyright holders and artists. DAT decks were never widely adopted, but this bill is historically significant as the first regulation of digital rights management; in this case, the law requires it be installed into certain products. The AHRA was an early source of pointed disagreement over the direction of copyright as a vehicle for the regulation of technology.

This dispute sets up each coalition's stance in the DRM debate, and it includes a partial list of the participants in today's dispute. Copyright holders—in this case, the music industry—sought to impose a requirement that manufacturers implement a specific DRM system in their product design. Fearful of tools permitting consumers to make digital copies of their works, they sought to impose restrictions on these tools reducing their copying power. On this count, the AHRA is a success. Electronics manufacturers— here, Sony and other DAT manufacturers—wanted to be able to design and sell their DAT decks without being sued. These consumer electronics companies disliked the mandate to include SCMS, which increased manufacturing costs while reducing the value the machines offered to consumers. Nonetheless, they grudgingly accepted this mandate as a lesser evil than unending litigation. In short, this is the first legislative battle between members of the strong copyright coalition and the still-nascent strong fair use coalition.

Title I of the DMCA, 1995-1998

The most important copyright law of the last 30 years is the DMCA, enacted in 1998, and Title I is generally cited as its most important component. Unlike the highly targeted AHRA, the DMCA[100] regulates almost all DRM. It bans the act of circumventing DRM to gain unauthorized access to copyrighted works. It also bans trafficking in tools that effectuate the circumvention of a technological protection measure that controls access to works or limits certain uses, such as copying. Further, it bans the removal of copyright management information—digital watermarks inserted into works to mark the identity of the copyright holder and communicate other information. The bill imposes very stiff civil liability and, for those who violate the law for commercial gain, criminal penalties of up to one million dollars and 10 years in prison.

---

[100] Unless otherwise specified, references to "the DMCA" in this dissertation refer to Title I, which contains the anticircumvention provisions under discussion.

Except for a few narrowly targeted exemptions, the statute provides for few affirmative defenses to charges of violating this law. Even if one's behavior would normally be permitted as a noninfringing use—for instance, a fair use—such activity is still illegal even though one is not committing copyright infringement. For instance, a video hobbyist making a video remix of his DVD collection in the privacy of his home is violating the law because he needs to circumvent the encryption on the DVDs to do it. The copying itself is probably a fair use,[101] but even if a judge says so, she is handicapped to exonerate the video hobbyist because fair use is no defense to charges of having violated the DMCA.

The strong copyright coalition came out in full support of Title I of the DMCA; supportive witnesses came from industries including the music, movie, publishing, and software, as well as the Copyright Office and the Patent and Trademark Office. The members of the strong fair use coalition opposed Title I or pushed for it to be substantially weakened. These included representatives of educational institutions, the computer hardware manufacturers, librarians, and the nonprofit sector.

DMCA Reform, 2003-2006

The strong fair use coalition desperately wants to change Title I of the DMCA, and Representative Rick Boucher, Democrat of Virginia, has led the congressional charge. In the 108th and 109th Congresses, he introduced bills (H.R. 107 in 2003-2004 and H.R. 1201 in 2005-2006) that would reduce the impact of the anticircumvention provisions of the DMCA. These bills would nullify the ban on circumventing copy controls as applied to otherwise legal activities. The video hobbyist described above would be able to circumvent the CSS on his DVDs without fear of recrimination. These bills would also substantially reduce the reach of the anti-trafficking provisions, allowing companies to develop and sell tools to help the video hobbyist. The law would still prohibit the hacking DVDs en route to selling bootlegged copies on the subway; because a would-be bootlegger would have circumvented the encryption, he would be subject to the DMCA's civil and criminal penalties *in addition to* the civil and criminal penalties that apply to the acts of infringement. Because infringement is already illegal, opponents decry Boucher's bill as a gutting of section 1201.

Nonetheless, Boucher's bills earned serious attention and made some headway. For instance, the bill was discussed in a number of congressional hearings. Further, in the 109th Congress, the bill's 13 bipartisan cosponsors included House Committee on Energy and Commerce Chair Joe Barton. During a hearing on the broadcast flag in 2006, Barton strongly suggested that he would not easily permit a broadcast flag bill to pass without coupling it with something akin to H.R. 1201.

Broadcast Flag, 2003-2006

Accompanying the transition to digital radio and television broadcasts, the music, television, and movie industries express trepidation about the potential for viewers to tape perfect digital copies of broadcasts, edit out the commercials, and post them online. For them, the broadcast flag is a potential solution to this problem. The broadcast flag is a very small addition to a digital signal by which somebody encoding a digital signal can tell compliant devices which programs may be recorded, which cannot, and what can be done with permitted recordings. Device makers have every incentive to make noncompliant devices; consumers will pay more for tools that permit any and all recording, and incorporating flag-compliant technology is an additional manufacturing expense. Hence, the music, motion picture, and television

---

[101] Fair use is far too murky an area of law to permit one to state that something like this is *definitely* a fair use. In this non-lawyer's estimation, nonetheless, this sort of harmless reproduction is a good example of fair use.

industries seek to impose a federal mandate that all digital radio and TV tuners comply with instructions contained in a broadcast flag.

The effort to impose video and audio flags has gotten even more traction than Boucher's bills. In November 2003, the FCC passed a mandate requiring flag compliance of all digital TV tuners by July 1, 2005. The DC Circuit Court vacated the rules as outside the Commission's jurisdiction. New Jersey Republican Representative Mike Ferguson introduced legislation requiring flag compliance for digital radio tuners. Several congressional hearings from 2003 to 2006 considered audio and video flag mandates. Because recording is valuable both for consumers in general and for socially valuable noninfringing uses—for instance, the media studies professor who wants to record digital television signals as part of her research—the strong fair use coalition is strongly opposed to any broadcast flag mandates, setting up another rich debate featuring a very similar lineup as in the H.R. 1201 debate.

Include/Exclude Criteria for Congressional Hearings

At this stage, coders are searching for congressional hearings concerning any of four specific policy debates, as described in the primer at the beginning of these coding instructions. This section describes the conditions under which hearings would be included or excluded as representing a likely contribution to one or more of the policy discussions at hand. Once hearings are identified as likely to be relevant, individual documents from each hearing can then be coded.

Documents to be coded are LexisNexis summaries of hearings. Please mark documents as zero (0) for "negative" or one (1) for "positive" for relevance to one or more of the four policy debates. Code exclusively for debates within the hearings' allotted timeframes; for instance, when coding a summary of a hearing from 1996, only code it for whether or not it is relevant to the debate over the passage of the DMCA. (Practically speaking, these debates are almost entirely contained within the specified windows.)

The question coders are asking is: "Is this hearing likely to contain at least some degree of testimony relevant to one of the four debates?" For the 2003-2006 time frame, a hearing may consider both the broadcast flag debate *and* the DMCA reform debate. Coders need not worry about *which* debate is implicated; as long as the hearing is likely to include one or both of the debates, it simply gets coded as positive.

In coding a hearing as positive or negative, please be sure to read the entire description of the hearing. In addition to the title and summary, note that LexisNexis provides one or more sections entitled "Statements and Discussion". These follow the list(s) of witnesses. In these sections, one can often find valuable details that help settle whether a hearing is relevant.

Because LexisNexis produces searches in reverse chronological order, the documents are numbered such that smaller document numbers represent newer documents. Thus, to make the use of this codebook easier, these instructions provide specific guidance in the same reverse chronological order.

DMCA Reform Bill

Include if it meets ANY of the following:

1. Mentions a proposed reform of the anti-circumvention provisions of the DMCA in any detail

   In 2003-04, the bill sponsored by Rep. Rick Boucher (D-VA) to this effect was HR 107, the Digital Media Consumers' Rights Act of 2003. In 2005-06, it was HR 1201, which may be under-described as merely a bill that required labeling of copy-protected media such as compact discs. Any hearing mentioning these bills by name or number will be included, but any mention of DMCA reform will be included, even without the corresponding bill number.

2. Discusses the effect the DMCA has had on fair use and other noninfringing uses

   Keep an eye out for any hearings that are dedicated to the discussion of fair use or to meeting consumers' expectations; in such hearings, the DMCA may be one of many topics discussed. If so, be sure to code it as positive.

353

Do not include a hearing MERELY because it:

3. Discusses other policy issues potentially involving the manipulation of DRM-protected content

    For instance, a hearing may consider the issue of technologies that remove objectionable content from DVD videos for family-friendly viewing. Unless the hearing description meets one of the above "include" criteria, do not include it merely because it considers these technologies.

4. Discusses the prospective development and deployment of DRM systems—other than the broadcast flag, as detailed below

    For example, do not include a hearing merely because it considers whether or not college campuses and other broadband providers should implement (or should be encouraged or required to implement) DRM systems to reduce infringing uses of their networks.

5. Discusses other policy issues involving the unauthorized digital transmission of media files. In particular, do not include hearings merely because they discuss peer-to-peer (P2P) file sharing or other forms of alleged infringement.

    Examples include:

    A. Concerns about the use of P2P on college campuses

    B. Proposals to prevent the unauthorized digitization of analog signals, often called the "analog hole", and circulation of such digital files

    C. The US Supreme Court decision in *MGM v. Grokster*

6. Discusses the operations of the US Copyright Office

7. Discusses portions of the DMCA other than Title I, the anticircumvention provisions

    In particular:

    A.   Title II of the DMCA sets up a notice-and-takedown procedure for copyright holders to contact internet service providers and demand the removal of copyrighted content from the internet; this limits the legal liability of online service providers.

    B.   Title IV of the DMCA includes six miscellaneous provisions, including one that sets up a new system for webcasters to pay royalties on music broadcast over the internet.

    C.   Title V provides copyright protection for the design of vessel hulls.

    References to these and other portions of the DMCA that have nothing to do with the regulation of DRM are not, by themselves, grounds for including a hearing.

Include if it meets ANY of the following:

1. Discusses a proposed regulation requiring terrestrial digital broadcasters—radio or television—to implement a broadcast flag scheme.

   In 2006, Rep. Mike Ferguson (R-NJ) sponsored HR 4861, the Audio Broadcast Flag Licensing Act of 2006, which would have authorized the FCC to impose a broadcast flag on radio transmissions. Other unnumbered draft bills include the Broadcast Flag Authorization Act of 2005 and the HD Content Protection Act of 2005. Any mention of these bills by name or number is adequate for a hearing to be included, but this is not required for a hearing to be coded as relevant.

2. Discusses the FCC ruling that attempted to impose a broadcast flag mandate on digital television receivers

3. *Considers whether the FCC should impose an audio or video flag in general*

Do not include a hearing MERELY because it:

4. Discusses the transition to digital broadcasting for radio or television, *even if* unrelated copyright issues (e.g., licensing and royalties rates) are mentioned

5. Proposes to reform the licensing and royalty rates paid to copyright holders, regardless of the means of transmission (digital/analog, internet/broadcast/satellite)

6. Discusses other policy issues involving the unauthorized digital transmission of media files

   In particular, do not include hearings merely because they discuss peer-to-peer (P2P) file sharing or other forms of alleged infringement. Examples include:

   a. Concerns about the use of P2P on college campuses

   b. Proposals to prevent the unauthorized digitization of analog signals, often called the "analog hole", and circulation of such digital files

   c. The US Supreme Court decision in *MGM v. Grokster*

Title I of the DMCA

Include if it meets ANY of the following:

1. Mentions the implementation of the WIPO Copyright Treaty or the WIPO Performances and Phonograms Treaty or any of the bills toward this effect

2. Mentions the debate about whether to prevent the circumvention of DRM or the development, distribution, or sale of circumvention devices

In 1995, Rep. Carlos Moorhead (R-CA) introduced H.R. 2441, the NII Copyright Protection Act of 1995. That same year, Sen. Orrin Hatch (R-UT) introduced S. 1284, a similar bill with the same title. In 1997, Rep. Howard Coble (R-NC) introduced H.R. 2281, the WIPO Copyright Treaties Implementation Act. In 1998, Sen. Hatch introduced S. 2037, the Digital Millennium Copyright Act of 1998.

Any hearing mentioning any of these bills by name or number will be included, but any mention of a similar prohibition on the circulation of DRM will also be included.

Do not include a hearing MERELY because it:

3. Discusses portions of the DMCA other than Title I, the anticircumvention provisions

   In particular, do not include a hearing merely because it discusses these other titles:

   A.   Title II of the DMCA sets up a notice-and-takedown procedure for copyright holders to contact internet service providers and demand the removal of copyrighted content from the internet; this limits the legal liability of online service providers.

   B.   Title IV of the DMCA includes six miscellaneous provisions, including one that sets up a new system for webcasters to pay royalties on music broadcast over the internet.

   C.   Title V provides copyright protection for the design of vessel hulls.

4. Discusses other policy issues involving digital media

   Examples include:

   A.   Unauthorized online transmission of media files

   B.   Digital performing rights in sound recordings

   C.   Techniques for film preservation

5. Discusses the operations of the US Copyright Office

Audio Home Recording Act

Include if it meets the following:

1. Mentions any bills that require the implementation of DRM in digital audio recording devices such as DAT decks

   In 1990, Sen. Dennis DeConcini (D-AZ) introduced S. 2358, the Digital Audio Tape Recorder Act of 1990. In 1991, he introduced S. 1623, the Audio Home Recording Act. That same year, Rep. Jack Brooks (D-TX) introduced H.R. 3204, similar legislation of the same title. Rep. Cardiss Collins (IL) also introduced an "Audio Home Recording Act" in 1992; this one was bill number H.R. 4567; consideration of Collins' bill became folded into the debate over Brooks' bill, H.R. 3204.

Any hearing mentioning any of these bills by name or number will be included, but doing so is not required for inclusion. Any mention of a similar requirement for the implementation of DRM in digital audio devices would merit a hearing's inclusion.

Do not include a hearing MERELY because it:

2. Discusses other copyright policy issues

   Examples include:

   A.   Copyright for software or the designs of semiconductor chips

   B.   Royalty rates for uses of copyrighted works

   C.   International trade or treaty development

3. Discusses the operations of the US Copyright Office

Include/Exclude Criteria for Congressional Documents

At this stage, coders are searching for congressional documents concerning any of the four policy debates. This section describes the conditions under which congressional documents would be included or excluded as representing specific policy discussions. Before coding, please see the attached briefing on the four policy debates.

At this stage, we are testing the intercoder reliability of these include/exclude criteria. Please mark documents as zero (0) for "negative" or one (1) for "positive" for relevance to a given policy debate.

Unlike the other types of documents used in this study, congressional documents are not sent separately. Each document is identified in the spreadsheet by the hearing number (e.g. CIS 96-H521-17) and starting page number. (Note that the page number refers to the page in the hearing, not in the PDF; a document from a 1996 hearing might be identified as starting on page 55 but appear on page 26 of the PDF. Check the author or speaker's name to be sure you are considering the correct document.) Code only until the end of a given document; for instance, do not include question-and-answer sessions.

Older hearings (1989 to 1998) are scanned in from microfiche; these have two hearing pages per print/PDF page. These are essentially digitized photocopies, so file sizes are often very large and readability is uneven. (If a document is effectively unreadable, please feel free to skip this document and continue coding; make a note of this in the spreadsheet and discuss when returning it to the study's lead author.) To keep file sizes manageable, many are broken into several separate files. Newer hearings (2003-2006) are one hearing page per print/PDF page, and have much smaller file sizes; each complete hearing is in one file. Most of the documents in the newer hearings are of perfect digital quality, and even scanned documents are almost universally readable.

IMPORTANT: The two debates from 2003-2006 are somewhat related, so the two codes are not mutually exclusive. Documents that are obviously positive for one code still must be examined in detail to see if they are also positive for the other code. Even if a document is predominantly about one debate and barely provides enough coverage to qualify as a positive for the other (see length guidelines below), code it as positive for both debates.

A word on the extent of topical coverage required for inclusion

Any document with a minimum of at least 4 full sentences worth of relevant content will be included. (This means that any document of less than 4 sentences will be excluded.) This amount of content may occur together or be scattered throughout the document. This amount may also include contextual material that, read alone, would not seem relevant—so long as that contextual material is intimately linked and, when read in context, clearly relevant.

The minimum amount of relevant content must occur in the main text; do not code for content in footnotes or endnotes.

This is merely the first stage, during which the only concern is to identify relevant documents and exclude irrelevant documents.

Include if it meets the following:

1. Discusses the specifics of bills that require the implementation of DRM in digital audio recording devices such as DAT decks

   In 1990, Sen. Dennis DeConcini (D-AZ) introduced S. 2358, the Digital Audio Tape Recorder Act of 1990. In 1991, he introduced S. 1623, the Audio Home Recording Act. That same year, Rep. Jack Brooks (D-TX) introduced H.R. 3204, similar legislation of the same title. Rep. Cardiss Collins (IL) also introduced an "Audio Home Recording Act" in 1992; this one was bill number H.R. 4567; consideration of Collins' bill became folded into the debate over Brooks' bill, H.R. 3204.

   Any document discussing any of these bills by name or number will generally be included (subject to the length requirements detailed above), but doing so is not required for inclusion. Any mention of a similar requirement for the implementation of DRM in digital audio devices would merit including the document as relevant. Note that relevant documents may present a very heated discussion of a bill or offer a numbingly dry and exactingly neutral description of the legislation's meaning; both kinds of documents are included, as are those that fall somewhere in between.

   Some of these bills, including the final law as passed, also include royalty provisions; the end result is effectively a tax on digital audio recording devices, with the money divided among various stakeholders in the music industry. Many documents discuss these bills primarily or even exclusively in terms of the royalty provisions. Some may even discuss royalties without mentioning the bill by name. As with the DRM mandate, include documents that discuss the debate over whether to impose royalties on digital recorders and digital media.

   This rule is triggered *if* a congressperson or other speaker or author includes a few vague platitudes about the bill without discussing the specifics—that is, so long as such vague platitudes are at least specifically tied to the bill at hand. For instance, one congressperson's opening statement before a hearing includes the following:

   > I would also like to commend the major industries affected by this legislation for their hard work in bringing this compromise agreement to Congress. It has been a long time in coming, and you are to be commended for your efforts.
   >
   > This legislation would clearly help the equipment manufacturers and the record and electronic industries, but it is also important that we help the copyright owners, without whom there would be no need for this legislation.
   >
   > It is also important that this legislation be in the best interest of the public. From the birth of this country, copyright and patent law have been primarily designed not to serve the interest of the creators but to serve the overall public interest. It is our purpose here this morning to make sure that H.R. 3204 strikes the proper balance between the public interest on one hand and the proprietary rights of the creators on the other.

   Even though the speaker does not address the bill's specifics (royalties, copyright management mandates, etc.), it includes over four sentences in general support of the bill, so it is included as relevant.

Keep in mind the length requirements in applying this guideline. For instance, another congressperson's statement includes the following:

> ... I am pleased to join in these proceedings and to see the productivity which comes when divergent interests bring collective resolve to solve problems.
>
> We stand on the threshold of exploding technological advancements, and this bill embodies clear examples of both subtle and glaring questions of equity and fairness in contrast to the mere fiscal bottom line. Global competitiveness demands that we learn from this experience, so that the American marketplace, this industry and its artists do not fall victim to the politics of free enterprise. I hope this measure will find broad support and quick dispatch in this subcommittee and at the full committee level.

This is an example of a document that ***barely*** qualifies as relevant, with exactly 4 barely-relevant sentences. Anything document with less than this is coded as not relevant.

2. Discusses the likely economic, social, or cultural impact of the ***copying*** enabled by the introduction and potential mass adoption of digital audio recording devices such as DAT decks.

This might include a recording industry executive predicting that DAT will be the death of her industry. This could also include an electronics manufacturer discussing the legal threats that delayed the introduction of DAT to the US market.

To meet this standard, a document must discuss either:

A.   The effects on copyright holders, which here will generally mean the music industry, including record companies (e.g. RCA, Capitol), music publishers (e.g., ASCAP, BMI), songwriters, performers, etc. OR

B.   The effects of copyright concerns on the marketing or uses of digital recording devices.

Most of the examples of documents meeting Rule 2(A) will be copyright holders and their allies bemoaning the effects of digital copying on their business model. These should be easy to spot. But also include documents that make the opposite case—that Digital Audio Tape specifically or digital audio recording devices generally present no real threat to copyright holders. For instance, one witness in the technology sector makes the following point:

> [T]his bill started with concerns over copyright infringement. That is where I will start. I bought this boom box which has two cassette transports last week for $49.99. Here is a prerecorded tape of Beethoven, who also is not going to get too much out of this bill. Into here goes a blank tape that costs about a dollar. ... This is a principal tool of copyright infringement.
>
> It is my opinion that copyright infringement material is widely available in America today, and there is no one who is refraining from infringement because of concerns about sound quality. If you want a better boom box, you can buy one for $200. ...
>
> I think it is absured to suggest that many people are going to rush out and spend a couple thousand dollars on a CD player, a DAT recorder, just so they can make what to them are only slightly better copies onto blank tapes that cost $18.

While most documents that discuss the copying implications of DAT stress the potential threat from perfect serial copies, this speech dismisses such a threat. Either side of the argument is grounds for including a document.

In addition, be sure to keep an eye out for documents arguing that the legal threats from copyright holders have (or have not) slowed or stopped the development and marketing of digital audio recording devices. (This is described in Rule 2(B).) One witness bemoans:

> In particular at Tandy, we've been working to develop the market for digital compact cassettes, a technology that we have been developing in conjunction with Phillips, Europe's largest consumer electronics company. ...
>
> Yet we have been hesitant to manufacture and market this or other digital recording technologies in recent years. It's just been too risky because of the threat of litigation. The introduction of digital audiotape or DAT recorders is a perfect example. Two weeks after the delayed introduction of DAT recorders in the U.S. market, a group of music publishers and songwriters sued the manufacturers for contributory copyright infringement.

The witness supports the bill explicitly, but even had he not done so, his testimony would have been included as relevant due to the above passage alone.

As an example of how an article can contain points that meet both Rule 2(A) and Rule 2(B), consider the following excerpt from an article submitted for the record:

> In the belief that consumers had fallen so much in love with the idea of digital audio because of their exposure to CD, Japanese manufacturers reasoned that Digital Audio Tape (DAT) would be to the CD what the compact cassette was to the LP. Unfortunately, it didn't work out that way for a number of reasons. First, the record industry, spearheaded by the RIAA (Recording Industry Association of America), threatened lawsuits against any Japanese manufacturer who exported the DAT machines to the U.S. The RIAA was concerned about DAT's potential to make virtually perfect copies of CDs. (They seemingly missed the fact that, for most people, cassettes do the same thing. And despite that, *pre-recorded* cassettes have outsold both LP's and CD's combined since 1982! They've outsold blank tapes as well.) The threats of lawsuits were enough to stop DAT dead in its tracks, despite considerable accolades for the format in the audio and general press.

The parenthetical comment is another statement to the same effect as the witness's statement about dual-cassette boomboxes: consumers have long had the tools for home taping that meets their audio needs, and the recording industry still makes plenty of money. This is relevant under 2(A). The rest of the excerpt discusses how legal threats have harmed DAT sales; this meets Rule 2(B). Thus, the document is included.

3. Discusses the role of the Serial Copy Management System (SCMS) in restricting copies

    SCMS was the specific technology that was included in the final legislation. All consumer-level digital audio recording devices, such as DAT, must now include this system, which permits the user to make unlimited copies of a digital source (e.g., a CD or a DAT tape) but not copies of copies.

Several documents discuss SCMS, but not all call it by name. Under this rule, discussion of SCMS or any similar technology that limits digital reproduction of audio recordings, whether named explicitly or not, is relevant.

Some of the documents included under this rule may not contain the sort of strong rhetoric on copying and copyright as those included under the above rules. For instance, one witness actually brings a DAT deck into the hearing and demonstrates how SCMS prevents him from making a second-generation copy. He states:

> Now because these machines are equipped with SCMS circuitry, the Sony DAT recorder will not be able to make a digital copy of the first copy. Please watch.

> It is hard to see, but you will notice if you come closer that the word "prohibit" is flashing on and off. While the meters are moving, the level indicators are moving, no recording is taking place. The tape is not moving at all. In short, the Sony DAT recorder is refusing to make a second generation digital copy. ...

> In short, this system provides consumers with the ability to make a first generation digital copy of prerecorded copyrighted music for their own convenience but not to engage in the type of serial copying that has concerned the recording industry over the years. It does so without any effect on the quality of the prerecorded music.

This speech is included as relevant because it describes the role of SCMS in restricting copies.

Do not include a document MERELY because it:

4. Discusses the economic impact of digital audio recording devices such as DAT *without* reference to the devices' impact on copyright industries.

   For instance, a document that merely discusses digital recorders as an economic good that may have certain effects on the economy—for instance, reviving consumer spending on retail electronics—is not included.

   As an example of this point, consider one submitted article that bemoans the introduction of new digital formats (such as Mini Disc and Digital Compact Cassette) on the grounds that the plethora of choices will only lead to consumer confusion.

5. Discusses the technological characteristics of digital recording devices

   Unless they also meet one of the rules for inclusion detailed above, do not include documents discussing the technical advantages of digital audio recording (e.g., higher audio fidelity) or debating the merits of various formats on technical grounds.

   For instance, one article submitted as an exhibit is filled with technology-specific criticism such as the following:

   > "The designers of new formats are doing the engineers and the consumer a disservice by not designing high sonic quality into their standards," [audio engineer Jim] Berry noted. "The new DCC and Mini-Disc aren't bad formats but they do not raise the quality of duplicated products either."

This article is purely technical, asking: In terms of audio quality, is Format A better than Formats B, C, and D? Thus, it is excluded from this study.

Another describes the technical qualities of the Mini Disc in mind-numbing detail. For instance:

> The MD system employs two kinds of media: magneto-optical media for recordable blank discs and CD-type optical media for prerecorded sotware. The magneto-optical drive (MOD) technology in MD is similar to others already in use, but brings some clever ideas to the party. ...

And on it goes. Even though MD is a format regulated under the Act, this has nothing to do with the copyright debate and is thus excluded.

6. Presents graphical information that is pertinent to one of the relevant topics above

In applying this codebook, **a picture is worth zero words**. For instance, one written submission is a visual depiction of the copying restrictions imposed by SCMS. It shows that one can successfully make a first generation copy but not copies of copies. Then, it shows that the circumvention of SCMS permits unlimited generations of copies to be made. If this information were presented in written rather than graphic form, this document would probably be included. In this case, however, the document is excluded for failing to have at least 4 relevant sentences.

7. Discusses royalty rates for *other* uses of copyrighted works

This is the trickiest such guideline to apply. One point of discussion that comes up every so often in these hearings is the question of a **performance royalty** for recorded music. At the time of these hearings, recorded music could be played publicly—on the radio, in bars and restaurants, even in stadiums—with no royalty due to the record company or recording artist. (In contrast, these performances have long required the payment of royalties to songwriters and music publishers. This system is something akin to being able to play movies publicly and only paying copyright royalties to the screenplay's authors.) This can be cast as something that should be imposed on analog uses of music or on digital uses such as internet transmissions; these are all irrelevant points of concern.

As an example of what an irrelevant discussion of a performance royalty might look like, consider the following excerpt from a written submission:

> Recently, the Copyright Office issued a report requested by Senator DeConcini on the copyright implications of digital audio transmission services. In this report, the Register reiterated that Congress establish a performance right in sound recordings. The basis for the Copyright Office's recommendation is three-fold. First, new digital audio transmission technologies are likely to fundamentally change the manner in which sound recordings are marketed to and enjoyed by listeners, to the detriment of the sound recording copyright owner.
>
> ...
>
> In conclusion, we fully support S. 1623 as the appropriate solution to the difficult issue of audio home recording. At the same time, however, digital audio transmission technologies pose additional challenges to the rights of sound recording copyright owners beyond the home copying issues addressed by S. 1623. Accordingly, we encourage Congress to separately establish a performance right for sound recordings.

The last paragraph contains the only relevant discussion in the entire document, and it really belies its unrelated overall purpose: pushing an entirely separate proposal requiring the imposition of performance royalties for recorded music.

To be relevant under Rule 1 above, any discussion of royalties must be tied to physical copies of digital media such as DAT, Digital Compact Cassette, or CDs. Documents containing such discussions of relevant royalties will generally make clear that these royalties are tied to the problem of digital copies using physical media.

8. Discusses other issues in copyright policymaking

Examples include:

A.   Copyright for computer software or the designs of semiconductor chips

B.   International trade or treaty development

On this point, some relevant documents do reference an "Athens agreement", by which they mean a 1988 meeting at which the recording industry and the electronics industry agreed that they could both live with digital audio recording devices that incorporate SCMS. In this set, each of these documents is included as relevant. Other international treaties, negotiations, or other meetings are not necessarily relevant.

C.   The means by which Congress makes copyright law

One article included for the record is a very lengthy critique of congressional copyright lawmaking. (The author argues that Congress tends to rely on industry negotiations and that this creates bad law.) For instance, the author writes:

> Throughout its history, copyright law has had difficulty accomodating technological change. Although the substance of copyright legislation in this century has evolved from meetings among industry representatives whose avowed purpose was to draft legislation that provided for the future, the resulting statutes have done so poorly. The language of copyright statutes has been phrased in fact-specific language that has grown obsolete as new modes and mediums [*sic*] of copyrightable expression have developed. Whatever copyright statute has been on the books has routinely, and justifiably, criticized as outmoded. In this Article, I suggest that the nature of the legislative process we have relied on for copyright revision is largely to blame for those laws' deficiencies.

This argument is not relevant to this specific legislative debate, so it is excluded from coding.

9. Discusses the operations of the US Copyright Office

10. Represents or discusses procedural matters

Several documents represent procedural matters that have no bearing on the debate per se. Examples include:

A.   Committee chairs' introduction of witnesses

364

B. Letters asking witnesses to give written responses to enclosed questions

As an example of this type of document, consider the letter from one senator, requesting witnesses' written responses to his supplemental questions:

"Thank you for taking the time out of your busy schedule to testify at the hearing on the Audio Home Recording Act of 1991. ...

As mentioned during the hearing, Senator Leahy has enclosed some supplemental questions for inclusion in the hearing record. Please return the questions with your answers to the attention of Mara Mallin by November 18, 1991. ...

This letter is coded as irrelevant, as are any like it.

Title I of the DMCA

Include if it meets ANY of the following:

1. Discusses the implementation of the anticircumvention provisions of the WIPO Copyright Treaty or the WIPO Performances and Phonograms Treaty or any of the bills toward this effect

In 1995, Sen. Orrin Hatch (R-UT) introduced S. 1284, the NII Copyright Protection Act of 1995. That same year, Rep. Carlos Moorhead (R-CA) introduced H.R. 2441, a similar bill with the same title. In 1997, Rep. Howard Coble (R-NC) introduced H.R. 2281, the WIPO Copyright Treaties Implementation Act. In 1998, Sen. Hatch introduced S. 2037, the Digital Millennium Copyright Act of 1998.

Keep an eye out for documents mentioning these by name (subject to the length requirements detailed above), though doing so is not required for inclusion.

Any document discussing these bills in a holistic fashion will be included (again, subject to length requirements). For example, a speech from a Representative filled with categorical but nonspecific support for a bill would be included.

IMPORTANT AND TRICKY: Several of these bills contain provisions that are totally unrelated to the implementation of the WIPO treaties and the DRM regulations that eventually became Title I of the DMCA. In particular, the 1995 bills (H.R. 2441 and S. 1284) and Sen. Hatch's 1997 bill (S. 2037) also contained provisions dealing with completely separate topics—particularly, online service provider liability.

The 1995 bills would have placed internet service providers in a difficult legal position by making it a violation of copyright to transmit an unauthorized copy of a work from one location to another. This would have raised their potential legal liability for end users' violations of copyright. The provision in the 1997 Senate bill states that internet service providers should generally not be liable for their users' infringement of copyright so long as they meet certain conditions. Rep. Coble's 1997 H.R. 2281 is the only one of the four that considered DRM regulation separately from internet service provider liability (this was added later, but not until after the hearings you are coding).

365

Again, the question of internet service providers' liability for their customers' copyright infringement is NOT RELEVANT TO THIS STUDY. Thus, three out of the four bills of interest from the 1995-1998 timeframe contain both relevant and irrelevant portions, and coders must separate the wheat from the chaff in the congressional hearings that consider these bills.

Do not include a document simply for discussing internet service provider liability. For instance, a telecommunications industry representative may talk at length about how H.R. 2441 will drive her company out of business because of end users' copyright infringement. This is not relevant to the debate about DRM regulation, so unless she also touched on the parts of the bill regulating DRM, her speech would be excluded.

To help clarify how to handle documents discussing each of these bills, see the following table:

Table B.1: Relevance of Documents Discussing Multiple Topics Included in DMCA

| Year | Bill # | Topics included | | Include document if it discusses | |
		DRM Circumvention	ISP Liability	Only Whole Bill	Specifics
1995	S. 1284	Yes	Yes	Yes	Maybe
1995	H.R. 2441	Yes	Yes	Yes	Maybe
1997	S. 2037	Yes	Yes	Yes	Maybe
1997	H.R. 2281	Yes	No	Yes	Yes

If a document expresses nonspecific/categorical support for or opposition to any of the four bills (subject to the "4 sentence" length requirement), include it as relevant. Even if an author merely muddles about the importance of getting the bill right, include it as relevant. If a document discusses the specifics of H.R. 2281, it will certainly be relevant.

If a document discusses the specifics of any of the other three bills, ascertain whether the topic is DRM regulation or internet service provider liability. If it discusses both topics, it is included, as long as it meets the length requirement of four relevant sentences. In this case, make sure that the number of sentences discussing the entire bill plus the number of sentences discussing DRM-specific portions add up to at least 4 sentences total.

If a document includes specifics that are relevant only to the topic of ISP liability or other topics, exclude it, even if it has four or more sentences discussing the bill as a whole.

Consider this excerpt from a speech that is included because it discusses both parts of H.R. 2441:

> I want to thank you for holding this second round of hearings on the NII Copyright Protection Act and for your leadership on this issue. We all realize that this legislation will have a significant impact on the development of the brave new world of the global information highway and must carefully balance the rights of copyright owners and Internet users. Therefore, we are all anxious to do it right. ...
>
> Fair use is also a concern for many of you, and I would like to hear your comments on how the concepts of fair use will apply in cyberspace. Anticopy technology and encryption are of special interest to me. ...
>
> Strengthened copyright protections will be useless unless copyright owners can protect their rights through encryption. I am fearful, however, that section 1201 may be too far-

reaching and have some unintended consequences. Some fine tuning of the language may be in order.

Each sentence above counts as relevant. The document discusses the specifics of the anti-circumvention provision (2nd and 3rd paragraphs), so the vague pro-bill statement at the beginning also counts as relevant content. In this case, there are four sentences of anti-circumvention-specific content, but this example illustrates the broader point.

Consider this additional example. The document is much more concerned with issues such as internet service provider liability and copyright royalties. Yet, buried deep in the document, two relevant paragraphs read:

> Copyright Management Information: Copyright management information associated with a work—such as the name of the author or copyright owner and the terms and conditions for use of the work—will serve to promote licensing and reduce liability concerns. The integrity of this information will be important in the NII and H.R. 2441 is a positive step forward in promoting the development and use of reliable rights management information. In this regard, BMI supports the relevant provisions of section 4 of H.R. 2441.
>
> Technological protections: Technology protections for copyrighted works will flourish in the digital environment just as they are flourishing in the areas of system security, currency protection, credit and banking, and privacy. While BMI is not currently affected by the provisions of the bill that bar the circumvention of copyright protection systems, we believe that enactment of such provisions would be sound public policy.

Each of these two paragraphs deals with a topic relevant to the debate: whether it should be illegal to remove or alter copyright information such as digital watermarks, and whether it should be illegal to circumvent access and copy controls (e.g., those achieved via encryption). With a quick mention of each, this document is included as relevant.

Finally, consider this excerpt:

> The American Society of Composers, Authors, and Publishers (ASCAP) strongly supports H.R. 2441, the "NII Copyright Protection Act of 1995." ASCAP commends Chairman Moorhead and his co-sponsors, Reps. Schroeder and Coble, for introducing this important legislation. ...
>
> Copyright Protection Systems and Copyright Management Information
>
> We believe that the NII holds the promise of new forms of security for copyright works, and new means of conveying information about copyrighted works which will benefit both creator and copyright owners, and users. As H.R. 2441 recognizes however, it is essential that the integrity of such systems for copyright protection and copyright management information be protected, as provided for by section 4 of the legislation. While we recognize that there may need to be fine-tuning of this provision, we fully support the principle behind it.

This is a perfect example of a document that is mostly about other topics but is included under these guidelines. It begins with several sentences of categorical support, and then produces several pages of support for the provisions dealing with copyright holders' right to control electronic transmissions of their works. The 3-sentence quip above about copyright protection systems and copyright management information (in other words, DRM) is buried several pages into the

document. Yet it contains enough relevant sentences, combining generic support for the bill with those dealing specifically with the DRM provisions, to be included.

Look for other documents that qualify as relevant in a similar manner.

2. Discusses any proposal to ban or impede the circumvention of DRM

    Even if a document does not mention one of the above bills by name, title, or author, include it as long as it discusses the same basic idea. This could be a ban on circumvention of DRM, a ban or limit on tools to circumvent DRM, etc.

3. Mentions the debate about whether to prevent the circumvention of DRM or the development, distribution, or sale of circumvention devices.

4. Participates in or describes the debate over the pros and cons of circumventing DRM.

    The point of this rule is to include documents that are part of the debate over whether and when circumvention is sometimes valuable. Documents that meet this rule without also mentioning the law are very rare, but this rule may help a document with an otherwise brief discussion of the law *per se* to meet the overall length requirement.

Do not include a document MERELY because it:

5. Discusses portions of the DMCA other than Title I, the anticircumvention provisions

    In particular, do not include a document merely because it discusses these other titles:

    A. Title II of the DMCA sets up a notice-and-takedown procedure for copyright holders to contact internet service providers and demand the removal of copyrighted content from the internet; this limits the legal liability of online service providers.

    B. Title IV of the DMCA includes six miscellaneous provisions, including one that sets up a new system for webcasters to pay royalties on music broadcast over the internet.

    C. Title V provides copyright protection for the design of vessel hulls.

    As these ideas began to meld together during the legislative process, hearings began mixing the debates around the various proposals; one important point of this code is to separate these debates.

6. Discusses the WIPO treaties or other international concerns

    This is tricky; most documents discussing WIPO will actually be relevant for one of the reasons described above. For instance, if a document calls for the implementation of the portion of the WIPO treaties dealing with DRM circumvention, it is included. This does not mean that any discussion of the treaties is relevant, because the treaties also cover other topics. Much like the bills discussed in these hearings, one of the heated points of exchange is whether internet service providers (ISPs) should be liable for their customers' behavior.

    Examples of topics that would not lead to a document being marked as relevant include:

    A. Provisions stipulating that copyright shall apply online in the same way that it applies to physical media such as CDs and books

B. Concerns about liability for online service providers

C. Discussion of the treaty process or the resulting treaty *without* reference to implementing legislation in the US

7. Contains detailed, relevant questions

Several of the documents are written responses to congresspersons' questions. Do not count the questions in assessing the total amount of relevant content.

8. Discusses other policy issues involving digital media and copyright

9. Discusses the operations of the US Copyright Office

10. Discusses other policy issues involving encryption

During this time, there is also an ongoing debate over whether US firms will be permitted to export, sell, or even develop systems for strong encryption. One might see references to export controls, limits on key length (e.g. 40 versus 64 versus 128-bit encryption), or requirements that keys be disclosed to the government. This is not part of the DMCA debate.

For instance, one document (already used as an example above) contains the following excerpt:

> *Question 8B*. Current U.S. law restricts the report of strong encryption, including the Data Encryption Standard (DES) which is commonly used n the Internet and around the world. Given the global environment of the Internet and other computer entworks, will these export restrictions affect how U.S. copyright holders are able to use encryption to protect their on-line works? ...

> Answer 8B. Export restrictions on encryption have a negative impact on the ability of copyright owners to use the Internet and other inherently global networks to distribute copyrighted materials that are technologically protected against unauthorized access and use.

This content is not relevant, and since this document does not contain enough content that is relevant, it is excluded.

11. Discusses protection for patents or trademarks

A patent is given for a new invention, such as a newly developed drug or a new type of machine. A trademark is a brand name or other mark used to identify a brand—for instance, the name "Coca-Cola" is a trademark, as is the Coke-shaped soda bottle. The bodies of law governing patents and trademarks are separate and distinct from copyright law.

There is some overlap between these areas. For instance, software companies sometimes get patents on specific functions of their programs. Further, most companies in the debate have one or more trademarks—Microsoft, Sony, and Warner Brothers are just a few examples. All three (and a few other areas) fall into a broader category of "intellectual property" law. Do not include a document on the grounds that it discusses protections for patents or trademarks.

12. Represents or discusses procedural matters

Several documents represent procedural matters that have no bearing on the debate per se. Examples include:

A.   Committee chairs' introduction of witnesses

B.   Congresspersons urging speakers to meet certain guidelines for their testimony, including calls to tackle (rather than avoid) tough issues, speak up, and stay within time.

Broadcast (Audio and/or Video) Flag

Include if it Meets ANY of the Following:

1. Describes the proposed broadcast flag regulations or legislation in enough depth to meet the length guidelines above. Any document that discusses the proposed regulations or legislation in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals.

   Consider this example:

   > [A]ll but one of the broadcast flag technologies approved by the FCC prohibit all Internet redistribution, not just "mass, indiscriminate" redistribution. So if I want to email a copy of my appearance on the local news to my mother, the flag prohibits me from doing so. Essentially, the video flag permits me to retain my fair use rights circa 1992. Not a significant improvement over the audio flag, if you ask me.

   This speaker is clearly opposing the FCC mandate in sufficient detail (even based on this quotation alone) to merit inclusion.

2. Describes the potential social or economic impact of the broadcast flag and/or circumvention of the broadcast flag. Unlike DMCA Reform Bill Rule 6 above (under which one does not include documents merely because they discuss the pros and cons of deploying one or more DRM systems), this rule requires that one does include documents that participate in or describe the debate over whether the broadcast flag is a socially valuable DRM system.

   For instance, consider the following brief excerpt:

   > I applaud Representative Ferguson for introducing the Audio Broadcast Flag Licensing Act. This allows the songwriters to receive fair compensation for their work. We are all grateful for your insight. On behalf of everyone in the music community, I hope you will support this bill and create for all songwriters a secure digital future.

   This excerpt comes at the end of an otherwise irrelevant speech, and it contains just barely enough relevant sentences (4), but it meets this rule and thus this document is included.

   Also, be sure to look for mentions of the Broadcast Protection Discussion Group (BPDG), the inter-industry group (similar to the DVD Copy Control Association) that developed the flag standard. Documents discussing the impact of the flag standard as developed by the BPDG are to be included under this rule. Consider this example:

As I noted at the outset, a core consumer concern that drives the acquisition of new products is to receive compelling content for enjoyment at home. Therefore, CERC members endorse the goal of the "broadcast flag" initiative, which is, I believe, correctly stated in the staff draft: to curb the unauthorized redistribution to the public of content over the Internet, in competition with the original authorized distributor.

We also endorse the other core goal of the draft, which is to do this without depriving consumers of the functionality of any of the products already in their home, or on their home network. Accomplishing both of these core goals -- as the private sector Broadcast Protection Discussion Group ("BPDG") participants found in six months of discussion -- is no easy task. Some of these complications are evident in the staff draft as well.

This is just part of a much longer discussion of the social and economic impact of the flag. Later, this document briefly mentions the possibility of a federal mandate, but even if not, it would be included based on the above kind of content.

3. Describes the potential copyright implications of the adoption of digital radio or TV broadcasting in an obvious allusion to the debate over whether or not to impose a broadcast flag. The most common example of this is the question of whether fans should continue to be able to record programming from radio or TV to be replayed later.

For instance, any document with an extended statement to the effect that the adoption of digital and/or HD radio or television broadcasts forces us to reconsider the balance of copyright law—e.g., we cannot allow fans to record digital content that is of such pristine quality and so easily reproduced—should be included. Likewise, any statement that takes the opposite stance, insisting that copyright should stay the same—e.g., consumers should retain the right to record digital radio in the same way they have recorded analog radio programming—is also to be included.

Consider this example passage, from a document discussing the rollout of digital radio:

In the Audio Home Recording Act of 1992 ("AHRA"), Congress definitively addressed the issue of home recording of sound recordings and musical works. This Act was intended to be comprehensive, forward-looking legislation designed to end, once and for all, the "longstanding controversy" surrounding the home recording of prerecorded music. Indeed, then-President of RIAA, Jay Berman, described the bill that became the AHRA as "a generic solution that applies across the board to all forms of digital audio recording technology."

The Senate Report that accompanied the AHRA opened its discussion of the bill with the assertion that "[t]he purpose of S.1623 is to ensure the right of consumers to make analog or digital audio recordings of copyrighted music for their private noncommercial use." To this end, the provision of the AHRA providing the exemption for home copying, section 1008, was considered "one of the cornerstones of the bill" because it "removes the legal cloud over home copying of prerecorded music in the most proconsumer way possible: It gives consumers a complete exemption for noncommercial home copying of both digital and analog music, even though the royalty obligations under the bill apply only to digitally formatted music." The Ninth Circuit confirmed this conclusion in Recording Industry Association of America v. Diamond Multimedia Systems, Inc., 180 F.3d 1072 (9th Cir. 1999).

This is an unmistakable defense of the current right to record at home. Other parts of the document provide the context to clarify that this is meant as an argument for consumers' right to record

371

digital broadcasts at home, and thus as an argument against the broadcast flag. Thus, the document would be coded as relevant based on the above passage and its context.

Do Not Include a Document MERELY Because It:

4. Is or includes proposed legislation or current law without comment. Any document that is *simply* a bill, statute, etc., will always be excluded.

5. Describes details that are not relevant to the debate about the social, economic, and political values at stake with the flag, its rollout, or home recording of digital broadcast. Examples include:

   A.   The technical details of the technology

   B.   The legal subtleties of the law as it now stands

   C.   The FCC's (lack of) jurisdiction to impose such a mandate

   D.   Other details that do not otherwise shed light on the merits of the flag or its circumvention and/or the merits or costs of laws or policies that would require or limit its adoption.

   None of these topics provide any insight into the costs or benefits of the flag or any flag-related policy changes.

6. Discusses other issues related to copyright in digital media, including especially royalty rates or payments for various uses of digital content.

   With some frequency, many speakers use these hearings as an opportunity to bemoan or defend certain uses of digital content under one or more extant licensing schemes.

   For instance, XM satellite radio had just released a receiver that allows consumers to record and save up to 50 hours of music from the satellite service. They read the law to say that they were responsible for paying royalties on the devices as laid on in the Audio Home Recording Act (a single levy on the device); this is in addition to the royalties they already pay for the digital transmission of every song they broadcast. Several witnesses insist that this is too low a price— that they should also pay royalties in line with those paid for purchases, e.g. those paid by iTunes. They often call for reforms in the laws governing statutory royalties. Witnesses who disagree insist that these recordings do not count as purchasing downloads and that the law on royalties should remain the same. This entire debate is irrelevant to this study.

Include If It Meets ANY of the Following:

1. Discusses a proposed DMCA reform, including its practical, social, economic, or other implications.

   In 2003, Representative Zoe Lofgren (D-CA)introduced HR 1066, the BALANCE Act, which would have permitted circumvention of DRM for otherwise noninfringing purposes and would have permitted the design and marketing of tools for such circumvention.

   Also in 2003, Rep. Rick Boucher (D-VA) introduced a very similar bill, HR 107, the Digital Media Consumers' Rights Act of 2003. In 2005-06, it was HR 1201, which may be under-described as merely a bill that required labeling of copy-protected media such as compact discs. Any document mentioning these bills by name or number will generally be included, subject to the length requirement of 4 relevant sentences, but any mention of DMCA reform will be included, even without the corresponding bill number.

   For instance, one written submission that is primarily about the broadcast flag also contains the following excerpts:

   > We worked with this Committee and the motion picture industry on the Digital Millennium Copyright Act of 1998 (the "DMCA") as well. Yet, we have also been surprised at some of the later interpretations of this law, and at the reluctance of some to consider the clarifications proposed by Chairman Barton and Congressman Boucher. ...

   > Any "Flag" Provision Should Be Proven Necessary And Accompanied By H.R. 1201 ...

   > [W]e respectfully urge that this subcommittee give renewed attention and impetus to protecting consumers, libraries, and educators by taking affirmative action on H.R. 1201.

   These four sentences, scattered throughout the multi-page document, provide just enough relevant content about HR 1201. Thus, though it is much more about the audio broadcast flag, it is also coded as relevant to the DMCA reform debate.

2. Any document that discusses the likely impact of proposed reforms in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals.

3. Describes costs/benefits of DRM-specific portions of DMCA (17 USC §§ 1201-1204), especially the anti-circumvention provision of DMCA (§ 1201).

   Documents meeting this criterion will almost certainly have an implicit or explicit message akin to one of the following:

   A. Keep this statute as it is to maintain these important social benefits. In one example, a representative argues:

      > The DMCA is a second example of a law being updated for the digital age. As an increasing number of pirated files became available for downloading, copyright owners realized that the federal government would only be able to

pursue the worst of the pirates. The need for self-help measures became clear. However, self-help measures did not make sense if tools to circumvent these self-help measures were readily available. In the analog world, we put locks on our doors to safeguard our possessions, and penalize those who possess burglary tools. A digital equivalent of laws to penalize those with burglary tools was needed. It was also necessary to align United States law with our international treaty obligations. Congress then enacted the Digital Millennium Copyright Act, the DMCA, in 1998 to penalize digital versions of burglary tools. I think the law has been a success. Our digital economy has exploded in size, making digital content available to everyone with a computer. However, it seems that the DMCA has been blamed for everything under the sun. This is unfortunate since the DMCA is the foundation of our nation's digital economy.

The representative never mentions the DMCA reform bill specifically, but the thrust of his comments is quite clear: he supports the anticircumvention provision as it now stands and thinks any attempt to reduce its reach would be a terrible idea. Further, he specifically mentions several of the bill's advantages: reduced piracy, international treaty compliance, and economic growth. Thus, this document is included for coding.

OR

B. Reform this statute or we will continue to suffer its negative effects. Such a document might never explicitly refers to a DMCA reform bill, but it should be included if it is reasonably read as building the case for just such a measure.

OR

C. There is a bitter debate afoot over whether or not to reform the law. One news story recounts the outcome of an ACLU suit that sought to challenge the constitutionality of several of the anticircumvention provisions. It includes summaries or quotations of arguments from both sides and the court's ruling, including:

As part of this rule, include documents that make (or describe) arguments about whether the Copyright Office's triennial rulemaking is properly determining exemptions to the law against circumventing DRM protections. A document may be included if it argues that the rulemaking is (not) properly balancing the interests of copyright holders and those who want to circumvent DRM—subject to the length guidelines outlined above. As with (c) above, one should also include documents that describe the debate over the rulemaking.

4. Participates in or describes the debate over the pros and cons of circumventing DRM.

The point of this rule is to include documents that are part of the debate over whether and when circumvention is sometimes valuable. Documents that meet this rule without also mentioning the law are very rare, but this rule may help a document with an otherwise brief discussion of the law *per se* to meet the overall length requirement.

For instance, consider this speech, in which the speaker insists that circumvention sometimes enables otherwise legal uses:

I am concerned that some attempts to protect content may overstep reasonable boundaries and limit the consumer's legal options particularly in light of the emerging technologies that we are beginning to see in the marketplace. It boils down to this. I believe that when I buy a music album or a movie DVD, it should be mine once I leave the store. Who does

not believe that? Does that mean that I have unlimited rights to use that DVD or that album? No, of course not, but the law should not restrict my fair use right to use my own property. Current law provides that I am liable for anything I do that amounts to infringement but current law also prevents me from making legal use of the content that is techno- logically locked even if I have the key. That just does not make sense to me. In defending this conflict, some say that fair use leads to piracy. Some even say that fair use is piracy. I do not believe that. I do not think it is. By definition, fair use is the use that does not infringe upon the owner's rights.

This is clearly part of the debate over whether there are sometimes proper reasons for circumventing DRM systems.

In addition to the impact on fair uses generally, pay careful attention for the debate around interoperability. Much of the debate about this portion of the DMCA revolves around the (in)ability of technologists to create products that integrate properly with systems that use DRM technologies.

IMPORTANT: **Do not** consider broadcast flag-specific discussion in applying this rule. But **do** include documents that are primarily about the flag if they would otherwise qualify—for instance, documents that also contain 4 or more sentences discussing a proposed DMCA reform bill. This is even if the document is mostly about the broadcast flag debate.

Do Not Include A Document MERELY Because It:

5. Is or includes proposed legislation or current law without comment. A document that contains both an entire bill and a meaningful discussion of its likely effects (e.g., a typical committee report) would be included, as long as it meets one of the "include" rules and the length requirements detailed above. But any document that is *simply* a bill, statute, etc., will always be excluded.

6. Describes the provisions of the law as it now stands. If it does not otherwise meet one of the rules for inclusion, a document should not be included merely because it contains a description of the anticircumvention provisions, such as a tutorial designed to teach somebody what would and would not count as a violation of the law.

    For example, one document that is coded as positive for broadcast flag but negative for DMCA reform includes the following passage, in which the author expresses reservations about the FCC interpreting copyright law:

    > While Congress itself has placed limitations on the exclusive rights of copyright owners in the course of mandating certain technologies, I am unaware of any precedent for a federal agency doing so.

    > About the closest precedent involves the Copyright Office, not the FCC. In the course of its triennial rule-making under Section 1201(a)(1)(C) of the Copyright Act, the Copyright Office is empowered to analyze whether the anti-circumvention provisions of the DMCA are adversely affecting non-infringing uses of copyrighted works. If the Copyright Office finds such adverse effects, it is empowered to create limited exemptions from the anti-circumventions provisions to protect the adversely-affected non-infringing uses.

    > While the Copyright Office's DMCA rule-making provides some parallels to the broadcast flag context, it is in many ways inapposite. The Copyright Office is an agency

that has tremendous experience with interpreting copyright laws, and is explicitly required by the DMCA to engage in the triennial rule-making. The FCC has no such copyright law experience of which I am aware.

While this speaker probably does have an opinion regarding the DMCA reform bill, this particular document is not part of that debate. It merely describes the law as it now stands.

As this example also illustrates, do not include a document merely because it discusses the Copyright Office's triennial rulemaking. Documents discussing the legislative basis and logistics of the rulemaking, without detailing the debate over the law and its interpretation, are excluded.

7. Discusses the pros and cons of deploying or mandating one or more DRM systems, whether retrospectively or prospectively

Debates about whether to implement a specific DRM system are not necessarily relevant to the debate about whether to reform the DMCA. The most obvious example is the broadcast flag, the other topic in this time frame. Most broadcast flag-relevant documents are not relevant to the DMCA debate.

Other possible DRM systems may be discussed, and coders should also treat them as irrelevant. For instance, some documents may be from hearings discussing whether or not certain providers of networked communication services (broadband providers, YouTube, universities, etc.) can or should deploy copyright-filtering technologies on their networks. This subject alone would not qualify a document for a positive coding.

Also, note that the occasional document may make brief reference to the DMCA in the context of the broadcast flag debate. For instance, one speaker remarks:

> One of these values is embodied in DMCA section [1201(c)(3)]. That section reflects a policy and decision by the Congress that consumer electronics and computer products not be required to respond to particular technological measures. This was a correct decision by the Congress and one that should not be overturned. It should certainly not be overturned by a grant of jurisdiction to the FCC and then subsequently, in essence, a retransfer of jurisdiction from the FCC to a small group of industry players.

This speaker is using the current DMCA merely as a wedge for arguing against the broadcast flag. Thus, it is not relevant to the DMCA reform debate.

8. Discusses other copyright-related issues

A.   Analog Hole

In the included hearings, the most commonly discussed irrelevant subject may be the so-called "analog hole," which is another way of saying re-digitization of analog content. At some point, a computer or other digital media player must turn bits into output that is consumed by humans. End users can record this content in an unencrypted format. This analog output represents a "hole" in DRM systems.

Suppose I buy a song in a DRM-restricted format that tethers the song to my computer. If I want another copy for my second computer, I could circumvent that DRM by hacking the code that tells my computer to disallow copying. (This would violate the DMCA, and even discussing this behavior would be relevant to this debate.) Alternately, I could record the audio output from my computer and resave that content as an unencrypted

376

MP3. The latter takes advantage of the analog hole. Because it is legal under the DMCA, discussion of the analog hole is not by itself relevant to the DMCA reform debate.

B.   Fair use of digital media

Several documents from the hearings under study are excluded even though they discuss the importance—or debate the contours—of fair use in the digital media environment. For instance, consider the following excerpt:

> The history of copyright law is a history of law adjusting to new technology. Often these laws cannot keep up with the state of technological advances. As we know, the internet and digital technology have created new possibilities for methods of distribution, of popular entertainment such as music and film in addition to enhancing academic studies.
>
> Determining how fair use is applied in this digital environment in the concept of appropriate fair use is something we as policymakers must carefully consider as we contemplate new laws to protect the interest of creators while maintaining access for consumers. In the past, traditional methods of copyright enforcement often involve the holder against a middleman. Illegal replication and distribution were more centralized in the activities of a bootlegger or an innocent infringer. Today, digital technology has cut out the middleman which makes copyright enforcement more challenging. In addition, as the public's consumption of digital products grows, the law and technology increasingly focus on digital means to protect copyright interest because of the great risk of piracy inherent in digital media exchanged over the internet. Thomas Freedman in his book, The World is Flat, talks in great depth about this very issue and the pros and cons involved in what the technology today is allowing us to do.

This document is excluded even though it discusses the contours of fair use in digital media. Additionally, it is excluded despite broaching the subject of the pros and cons of deploying DRM systems; this is discussed in Rule 7, above.

Include/Exclude Criteria for Internet Documents

At this stage, coders are searching for documents concerning two specific policy debates. This section describes the conditions under which documents from the internet would be included or excluded as representing specific policy discussions.

One debate ("DMCA Reform") is the question of whether the anticircumvention provisions of the Digital Millennium Copyright Act should be loosened to allow more kinds of circumvention and/or the development, marketing, and distribution of devices that are capable of circumvention. The other debate ("Broadcast Flag") considers whether the federal government should impose a requirement that all devices capable of receiving digital radio or television recognize copy management technology most commonly called the "broadcast flag". Terms such as "audio flag" and "video flag" are also common.

Before coding, please see the attached briefing on the two policy debates. Coders are qualified based on their background knowledge of DRM policy debates, but the briefing provides a more specific sense of the debate that is of interest in this project.

At this stage, we are testing the intercoder reliability of these include/exclude criteria. Please mark documents as zero (0) for "negative" or one (1) for "positive" for relevance to a given policy debate.

IMPORTANT: These topics are very closely related, so the two codes are not mutually exclusive. Documents that are obviously positive for one code still must be examined in detail to see if they are also positive for the other code. Even if a document is predominantly about one debate and barely provides enough coverage to qualify as a positive for the other (see length guidelines below), code it as positive for both debates.

A word on the extent of topical coverage required for inclusion

Any document with a minimum of at least 4 full sentences worth of relevant content will be included. (This means that any document of less than 4 sentences will be excluded.) This amount of content may occur together or be scattered throughout the document.

Note, however, that the 4 full sentences must occur in the main text; do not code for content in footnotes or endnotes. Further, if the document contains multiple authors or voices—e.g., a congressional hearing— demand 4 sentences from the same author or speaker. (Do not keep documents merely because 4 different people have made glancing references to a subject.)

This is merely the first stage, during which our only concern is to identify relevant documents and exclude irrelevant documents.

Coders do not necessarily need to read long documents in their entirety. Rather, proceed through the following steps to manage documents that are too long to read quickly:

1. Exclude documents that are obviously irrelevant to the copyright debate, e.g. a congressional hearing on US broadcasting policy in Cuba.

2. For documents that are related to copyright, one can save a good deal of time (identifying true positives quickly) by searching for the following keywords:

    A.   Broadcast Flag debate:

            i. Flag

            ii. FCC

            iii. Digital

            iv. TV

            v. Television

            vi. Radio

    B.   DMCA reform debate:

            i. DMCA

            ii. Digital Millennium

            iii. 1201

            iv. Boucher

            v. Circumven

    Always be sure to read these keywords in context. Further, do not take the lack of keywords (or the lack of relevant content surrounding these keywords) as definitive evidence that the document should be excluded.

3. For documents with an executive summary, introduction, and/or table of contents, exclude if *none* of these elements provides anything possibly of relevance. Err on the side of further examination; the goal is to include all relevant documents.

    A.   If the table of contents includes even potentially relevant sections, read those sections.

    B.   Within this guideline, treat congressional hearings as long documents, and look at the chair's opening remarks[102] and witness list as one would an executive summary and table of contents. If either of those elements suggests potentially relevant witnesses, read further to see if the hearing should be flagged as a positive.

---

[102] Occasionally, the chair's first speech is on a subject other than the hearing (e.g., it is her/his last hearing as chair of that committee, and s/he wants to reflect on their tenure). In these cases, search for the first time the chair or another representative provides a detailed description of the intended contents of for the hearing. Then, treat that speech as the introduction.

Recall that in congressional hearings, insist on four developed, topical sentences from within at least one individual submission—generally, a Representative's speech, a witness's testimony, or a written submission. Do *not* code a hearing as positive merely because four or more speakers make one-sentence references to a topic.

4. For documents that are long enough to be broken into sections but contain none of these elements, read any sections that may potentially contain relevant content. Essentially, use section headers as one would use a table of contents; if any section looks potentially relevant, please read through it.

DMCA Reform Bill

Include If It Meets ANY of the Following:

1. Discusses a proposed DMCA reform, including its practical, social, economic, or other implications. Any document that discusses the likely impact of proposed reforms in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals. As detailed in Rule 6 below, this does not include any document that *is* a proposed bill; it must *discuss* a proposed bill.

    As an example of a document that obviously meets this rule is a press release by the congressional office of one of the sponsors of a reform bill. The release is entirely about the details of the bill, complete with the Representative's spin on the subject.
    Not every document will need to be this obviously about the bill in order to be included, but those that provide some sort of meaningful discussion of the bill—even if not by name or bill number—will still be included. Another example is a reprinted article from the *Library Journal* declaring Rep. Boucher as the journal's choice for "Politician of the Year 2006". It provides an overview of Boucher's library-friendly policy stances. As part of this document, the author includes the following quote from Boucher:

    > "The creator of content in the future will deliver it in digital format, protected with a password, and to get around the password the user will have to pay every time.... This is the great concern, and I think every librarian is worried about it. I've introduced a bill that would solve this problem by allowing users to bypass technical protection measures for purposes that do not infringe on copyright, purposes exercising the right of fair use." Boucher says there is "terribly strong" opposition to the bill because it "calls into question" the original intent of those who wanted the DMCA. It seems almost designed to kill fair use.

    This is roughly the last third of a section that details Boucher's stance on the anticircumvention provisions of the DMCA, but even if not, this document would be included based on the above-quoted section alone.

2. Describes costs/benefits of DRM-specific portions of DMCA (17 USC §§ 1201-1204), especially the anti-circumvention provision of DMCA (§ 1201).

    Documents meeting this criterion will almost certainly have an implicit or explicit message akin to one of the following:

    A.   Keep this statute as it is to maintain these important social benefits. In one example, a representative argues:

The DMCA is a second example of a law being updated for the digital age. As an increasing number of pirated files became available for downloading, copyright owners realized that the federal government would only be able to pursue the worst of the pirates. The need for self-help measures became clear. However, self-help measures did not make sense if tools to circumvent these self-help measures were readily available. In the analog world, we put locks on our doors to safeguard our possessions, and penalize those who possess burglary tools. A digital equivalent of laws to penalize those with burglary tools was needed.

It was also necessary to align United States law with our international treaty obligations. Congress then enacted the Digital Millennium Copyright Act, the DMCA, in 1998 to penalize digital versions of burglary tools. I think the law has been a success. Our digital economy has exploded in size, making digital content available to everyone with a computer. However, it seems that the DMCA has been blamed for everything under the sun. This is unfortunate since the DMCA is the foundation of our nation's digital economy.

The representative never mentions the DMCA reform bill specifically, but the thrust of his comments is quite clear: he supports the anticircumvention provision as it now stands and thinks any attempt to reduce its reach would be a terrible idea. Further, he specifically mentions several of the bill's advantages: reduced piracy, international treaty compliance, and economic growth. Thus, this document is included for coding.

OR

B.   Reform this statute or we will continue to suffer its negative effects. One document begins with:

This document collects a number of reported cases where the anti-circumvention provisions of the DMCA have been invoked not against pirates, but against consumers, scientists, and legitimate competitors.

The document never explicitly refers to a DMCA reform bill, but it is obviously building the case for just such a measure. Of course, with an opening like that, it does not take much reading to verify that it meets the length guidelines above; the very purpose of the document is to collect stories of the law's misapplication.

OR

C.   There is a bitter debate afoot over whether or not to reform the law. One news story recounts the outcome of an ACLU suit that sought to challenge the constitutionality of several of the anticircumvention provisions. It includes summaries or quotations of arguments from both sides and the court's ruling, including:

The ACLU's suit, filed against filtering-software company N2H2 last July, claims the law unconstitutionally interferes with researchers' ability to investigate and evaluate the effectiveness of Internet filtering software.

"There is no plausibly protected constitutional interest that...outweighs N2H2's right to protect its copyrighted property from an invasive and destructive trespass," U.S. District Judge Richard Sterns wrote.

> Sterns' ruling dealt a sore blow to opponents of the DMCA, who claim it violates free speech rights protected by the First Amendment, …

None of these arguments are developed in detail, but it nonetheless meets both this rule and the length guidelines above.

As part of this rule, include documents that make (or describe) arguments about whether the Copyright Office's triennial rulemaking is properly determining exemptions to the law against circumventing DRM protections. A document may be included if it argues that the rulemaking is (not) properly balancing the interests of copyright holders and those who want to circumvent DRM—subject to the length guidelines outlined above. As with (c) above, one should also include documents that describe the debate over the rulemaking.

One policy actor's website includes a treatise interpreting the entire DMCA. Nearly the entire section discussing the anticircumvention provisions (Title I of the Act; Section 2 of the document) is an objective description of the current statutory and case law, as well as administrative rulings via the US Copyright Office. If that were true of the entire section, the document would be excluded.

However, the document does include one lament:

> Unfortunately, the triennial rulemaking has been largely ineffective at protecting fair use and other noninfringing activities impaired by technical protection measures. First, the Librarian is not empowered to grant any exemptions to the trafficking prohibitions contained in 1201(a)(2) or 1201(b). Accordingly, only those who have the technological ability to circumvent are directly assisted by the exemptions. Second, the Copyright Office has imposed a heavy burden on those who seek exemptions.

Even though this is an isolated aside, this document is included as a positive for the DMCA Reform debate.

This part of this rule is particularly tricky when coding the opening statement of a congressional hearing. If the chair even hints at any of the above topics, search the hearing text in greater depth in search of relevant text.

3. Participates in or describes the debate over the pros and cons of circumventing DRM. Consider one document primarily dedicated to arguing that it is necessary to circumvent DRM in order to achieve many socially valuable fair uses. As it states:

> Digital rights management (DRM) technologies are aimed at increasing the kinds and/or scope of control that rights-holders can assert over their intellectual property assets. In the wake of the Digital Millennium Copyright Act's (DMCA) ban on the circumvention of DRM technologies used to control copyrightable works, DRM restrictions are now backed up with the force of law. In essence, copyright owners now have the ability to write their own intellectual property regime in computer code, secure in the knowledge that the DMCA will back the regime with the force of law.

This would be included *even if* the DMCA were not explicitly invoked. (E.g., it may have noted the technical know-how required to circumvent DRM and argued that this is regrettable because circumvention is sometimes socially desirable.)

The point of this rule is to include documents that are part of the debate over whether and when circumvention is sometimes valuable. These instructions include an example that still mentions the

DMCA because those that meet this rule without also mentioning the law—and thus meeting Rule #2—are very rare.

In addition to the impact on fair uses generally, pay careful attention for the debate around interoperability. Much of the debate about this portion of the DMCA revolves around the (in)ability of technologists to create products that integrate properly with systems that use DRM technologies.

IMPORTANT: Do *not* apply this rule if the only DRM discussed is the broadcast flag. Further, exclude broadcast-flag specific discussion when searching for minimum length requirements. But *do* include documents that are primarily about the flag if they also include adequate discussion of the costs or benefits of circumvention in general, even if the ultimate rhetorical thrust is toward the broadcast flag debate.

Do Not Include A Document MERELY Because It:

4. Is or includes proposed legislation or current law without comment. A document that contains both an entire bill and a meaningful discussion of its likely effects (e.g., a typical committee report) would be included, as long as it meets one of the "include" rules and the length requirements detailed above. But any document that is *simply* a bill, statute, etc., will always be excluded.

5. Describes the provisions of the law as it now stands. If it does not otherwise meet one of the rules for inclusion, a document should not be included merely because it contains a description of the anticircumvention provisions, such as a tutorial designed to teach somebody what would and would not count as a violation of the law.

    For example, one document that is coded as positive for broadcast flag but negative for DMCA reform includes the following passage, in which the author expresses reservations about the FCC interpreting copyright law:

    > While Congress itself has placed limitations on the exclusive rights of copyright owners in the course of mandating certain technologies, I am unaware of any precedent for a federal agency doing so.

    > About the closest precedent involves the Copyright Office, not the FCC. In the course of its triennial rule-making under Section 1201(a)(1)(C) of the Copyright Act, the Copyright Office is empowered to analyze whether the anti-circumvention provisions of the DMCA are adversely affecting non-infringing uses of copyrighted works. If the Copyright Office finds such adverse effects, it is empowered to create limited exemptions from the anti-circumventions provisions to protect the adversely-affected non-infringing uses.

    > While the Copyright Office's DMCA rule-making provides some parallels to the broadcast flag context, it is in many ways inapposite. The Copyright Office is an agency that has tremendous experience with interpreting copyright laws, and is explicitly required by the DMCA to engage in the triennial rule-making. The FCC has no such copyright law experience of which I am aware.

    While this speaker probably does have an opinion regarding the DMCA reform bill, this particular document is not part of that debate. It merely describes the law as it now stands.

As this example also illustrates, do not include a document merely because it discusses the Copyright Office's triennial rulemaking. Documents discussing the legislative basis and logistics of the rulemaking, without detailing the debate over the law and its interpretation, are excluded.

6. Discusses the pros and cons of deploying one or more DRM systems, whether retrospectively or prospectively. For instance, some documents may be from hearings discussing whether or not certain providers of networked communication services (broadband providers, YouTube, universities, etc.) can or should deploy copyright-filtering technologies on their networks. This subject alone would not qualify a document for a positive coding.

Broadcast (Audio and/or Video) Flag

Include if It Meets ANY of the Following:

1. Describes proposed broadcast flag regulations or legislation in enough depth to meet the length guidelines above. Any document that discusses the proposed regulations or legislation in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals.

    Consider this example:

    > [A]ll but one of the broadcast flag technologies approved by the FCC prohibit all Internet redistribution, not just "mass, indiscriminate" redistribution. So if I want to email a copy of my appearance on the local news to my mother, the flag prohibits me from doing so. Essentially, the video flag permits me to retain my fair use rights circa 1992. Not a significant improvement over the audio flag, if you ask me.

    This speaker is clearly opposing the FCC mandate in sufficient detail (even based on this quotation alone) to merit inclusion.

    Here is another example from a chair's opening speech for a congressional hearing (which is also coded positive for the DMCA reform debate) meeting this rule:

    > And if a modest bill such as one for a broadcast flag standard be deemed necessary by our members to allow consumers to enjoy new digital copyrighted broadcasts and prevent piracy, it seems to me that our Subcommittee will be the home for the drafting and the review of same.

    As described above, this mention would be too brief to merit inclusion on its own. In an opening statement for a hearing, however, this is reasonably read as an allusion to more substantial discussion below. Thus, this statement would lead coders to examine the hearing more carefully.

2. Describes the potential social or economic impact of the broadcast flag and/or circumvention of the broadcast flag. Unlike DMCA Rule 6 above (under which one does not include documents merely because they discuss the pros and cons of deploying one or more DRM systems), this rule requires that one does include documents that participate in or describe the debate over whether the broadcast flag is a socially valuable DRM system.

    For instance, be sure to look for mentions of the Broadcast Protection Discussion Group (BPDG), the inter-industry group (similar to the DVD Copy Control Association) that developed the flag

standard. Documents discussing the impact of the flag standard as developed by the BPDG are to be included under this rule. Consider this example:

> Digital Television Broadcast Flag Rulemaking
>
> As I noted at the outset, a core consumer concern that drives the acquisition of new products is to receive compelling content for enjoyment at home. Therefore, CERC members endorse the goal of the "broadcast flag" initiative, which is, I believe, correctly stated in the staff draft: to curb the unauthorized redistribution to the public of content over the Internet, in competition with the original authorized distributor.
>
> We also endorse the other core goal of the draft, which is to do this without depriving consumers of the functionality of any of the products already in their home, or on their home network. Accomplishing both of these core goals -- as the private sector Broadcast Protection Discussion Group ("BPDG") participants found in six months of discussion -- is no easy task. Some of these complications are evident in the staff draft as well.

This is just part of a much longer discussion of the social and economic impact of the flag. Later, this document briefly mentions the possibility of a federal mandate, but even if not, it would be included based on the above discussion, couched in terms of the BPDG.

3. Describes the potential copyright implications of the adoption of digital radio or TV broadcasting in an obvious allusion to the debate over whether or not to impose a broadcast flag. The most common example of this is the question of whether fans should continue to be able to record programming from radio or TV to be replayed later.

    For instance, any document with an extended statement to the effect that the adoption of digital and/or HD radio or television broadcasts forces us to reconsider the balance of copyright law—e.g., we cannot allow fans to record digital content that is of such pristine quality and so easily reproduced—should be included. Likewise, any statement that takes the opposite stance, insisting that copyright should stay the same—e.g., consumers should retain the right to record digital radio in the same way they have recorded analog radio programming—is also to be included.

    Consider this example passage, from a document discussing the rollout of digital radio:

    > In the Audio Home Recording Act of 1992 ("AHRA"), Congress definitively addressed the issue of home recording of sound recordings and musical works. This Act was intended to be comprehensive, forward-looking legislation designed to end, once and for all, the "longstanding controversy" surrounding the home recording of prerecorded music. Indeed, then-President of RIAA, Jay Berman, described the bill that became the AHRA as "a generic solution that applies across the board to all forms of digital audio recording technology."
    >
    > The Senate Report that accompanied the AHRA opened its discussion of the bill with the assertion that "[t]he purpose of S.1623 is to ensure the right of consumers to make analog or digital audio recordings of copyrighted music for their private noncommercial use." To this end, the provision of the AHRA providing the exemption for home copying, section 1008, was considered "one of the cornerstones of the bill" because it "removes the legal cloud over home copying of prerecorded music in the most proconsumer way possible: It gives consumers a complete exemption for noncommercial home copying of both digital and analog music, even though the royalty obligations under the bill apply only to digitally formatted music." The Ninth Circuit confirmed this conclusion in Recording

385

Industry Association of America v. Diamond Multimedia Systems, Inc., 180 F.3d 1072 (9th Cir. 1999).

This is an unmistakable defense of the current right to record at home; while this document is already included for meeting other rules above, it would be coded positive based on the above passage and its context—again, a discussion about the rollout of digital radio.

Do Not Include a Document MERELY Because It:

4. Is or includes proposed legislation or current law without comment. A document that contains both an entire bill and a meaningful discussion of its likely effects (e.g., a typical committee report) would be included, as long as it meets one of the "include" rules and the length requirements detailed above. But any document that is *simply* a bill, statute, etc., will always be excluded.

5. Describes details that are not relevant to the debate about the social, economic, and political values at stake with the flag, its rollout, or home recording of digital broadcast. Examples include:

    A.   The technical details of the technology

    B.   The legal subtleties of the law as it now stands

    C.   The FCC's (lack of) jurisdiction to impose such a mandate

    D.   Other details that do not otherwise shed light on the merits of the flag or its circumvention and/or the merits or costs of laws or policies that would require or limit its adoption.

As one example, consider a news story about the case before the DC Circuit questioning whether the FCC stepped outside its authority when it attempted to mandate that all new digital televisions implement the broadcast flag. Nearly the entire story discusses legal issues that do not concern the debate about whether or not the broadcast flag should be adopted. For instance, the article included the following quotations from judges questioning the Commission's jurisdiction:

> "You're out there in the whole world, regulating. Are washing machines next?" asked Judge Harry Edwards. Quipped Judge David Sentelle: "You can't regulate washing machines. You can't rule the world."

The article also includes discussion about whether or not those who sued to stop the rule from taking effect have "standing"—that is, whether they have the legal right to sue in this case. For instance:

> "You have to have a harm that distinguishes you from the public at large," Sentelle said during oral arguments. "If there is not a particularized harm, you do not have standing...There may be someone from the industry who can come forward." Edwards also said he was concerned about the groups' "standing," referring to the judicially recognized right to sue. Special rules exist for organizations suing federal agencies.

None of this content provides any insight into the costs or benefits of the flag or any flag-related policy changes. The same story does provide a scintilla of the debate as context. Drawn from two separate parts of the article, here it is in its entirety:

> [The groups that sued to stop the rule argued] that librarians' ability to make "fair use" of digital broadcasts will be unreasonably curtailed. …
>
> From the perspective of the entertainment industry, the broadcast flag is needed to encourage over-the-air distribution of valuable content. Without the FCC's action, the Motion Picture Association of America has argued, the threat of Internet piracy would imperil the future of digital TV.

These brief discussions of the social and economic impacts of the flag are just barely too short (3 total sentences) to meet our length guidelines outlined above. Thus, this document is not included.

Include/Exclude Criteria for Newspaper Articles

At this stage, coders are searching for newspaper articles concerning any of the four policy debates. This section describes the conditions under which newspaper articles would be included or excluded as representing specific policy discussions. Before coding, please see the attached briefing on the four policy debates.

At this stage, we are testing the intercoder reliability of these include/exclude criteria. Please mark articles as zero (0) for "negative" or one (1) for "positive" for relevance to a given policy debate.

A word on the extent of topical coverage required for inclusion

Any article with a minimum of at least 4 full sentences worth of relevant content will be included. (This means that any article of less than 4 sentences will be excluded.) This amount of content may occur together or be scattered throughout the article. This amount may also include contextual material that, read alone, would not seem relevant—so long as that contextual material is intimately linked and, when read in context, clearly relevant. This threshold for relevance may be met by any one positive rule below (1, 2, or 3), or it may be met by any combination—for instance, 2 sentences from rule 1, and 2 from rule 3.

This is merely the first stage of coding newspaper articles, during which the only concern is to identify relevant articles and exclude irrelevant articles.

Audio Home Recording Act

Include if it meets the following:

1. Discusses the specifics of bills that require the implementation of DRM in digital audio recording devices such as DAT decks

   In 1990, Sen. Dennis DeConcini (D-AZ) introduced S. 2358, the Digital Audio Tape Recorder Act of 1990. In 1991, he introduced S. 1623, the Audio Home Recording Act. That same year, Rep. Jack Brooks (D-TX) introduced H.R. 3204, similar legislation of the same title. Rep. Cardiss Collins (IL) also introduced an "Audio Home Recording Act" in 1992; this one was bill number H.R. 4567; consideration of Collins' bill became folded into the debate over Brooks' bill, H.R. 3204.

   Any article discussing any of these bills by name or number will generally be included, subject to the length requirements detailed above, but doing so is not required for inclusion. Any mention of a similar requirement for the implementation of DRM in digital audio devices would merit including the article as relevant.

   These bills also included royalty provisions; the end result is effectively a tax on digital audio recording devices, with the money divided among various stakeholders in the music industry. Many articles discuss these bills primarily or even exclusively in terms of the royalty provisions.

For instance, one column urges Congress not to pass the bill, describing the royalty as "parasitic." Subject to the length requirement of four or more relevant sentences, include articles like this.

2. Discusses the likely economic, social, or cultural impact of the **copying** enabled by the introduction and potential mass adoption of digital audio recording devices such as DAT decks.

    This might include a recording industry executive predicting that DAT will be the death of her industry. This could also include an electronics manufacturer discussing the legal threats that delayed the introduction of DAT to the US market.

    To meet this standard, an article must discuss either:

    A.   The effects on copyright holders, which here will generally mean the music industry, including record companies (e.g. RCA, Capitol), music publishers (e.g., ASCAP, BMI), songwriters, performers, etc. OR

    B.   The effects of copyright concerns on the marketing or uses of digital recording devices.

    As an example of 2B, consider the following excerpt:

    > The copy is a virtual clone of the original in every tonal nuance - and if the original is protected by copyright, as most recordings are, there are legal limits to what you may do along these lines. (Basically, you can make copies for yourself, but you cannot sell or otherwise distribute them.) ... [For Nakamichi to bring its DAT recorder to the US] was to brave the ire of the Recording Industry Association of America, which threatened legal action against anyone importing DAT equipment for home use unless such devices had provisions to prevent unlimited copying of CD's. The Nakamichi has no such provisions, yet its sale in this country has not been challenged legally.

    Only these four sentences are relevant under these instructions, but this is just enough content that the article is coded as relevant.

3. Discusses the Serial Copy Management System (SCMS)

    SCMS was the specific technology that was included in the final legislation. All consumer-level digital audio recording devices, such as DAT, must now include this system, which permits the user to make unlimited copies of a digital source (e.g., a CD or a DAT tape) but not copies of copies.

    Several articles discuss SCMS, but not all call it by name. For instance, one article includes the following 4 relevant sentences from different parts of the article:

    > With its demonstration of a new audio format, Philips set in motion what is sure to be a lengthy slugfest among equipment manufacturers and record companies hoping to influence the next standard for music listening in the home and car. ...

    > Far from a certain success, Philips's system faces opposition from the Recording Industry Association of America, the trade group that believes record companies are being deprived of millions of dollars in revenue by consumers who violate copyrights by recording songs onto tapes. After mounting a copyright challenge to DAT, the group was successful in winning a concession from DAT recorder makers in the form of a microchip that allows users to copy a CD, but does not permit a digital copy of that tape copy.

Philips plans to include the same circuitry in its system, but that may not satisfy songwriters or record companies.

These four sentences are all relevant; the first two sentences are relevant under Rule 2B, and the next two sentences are relevant under this, Rule 3. Thus, the article has the required number of relevant sentences and is included.

Under this rule, discussion of SCMS or any similar technology, whether named explicitly or not, is relevant.

Do not include an article MERELY because it:

4. Discusses the economic impact of digital audio recording devices such as DAT *without* reference to the devices' impact on copyright industries.

   For instance, one article includes the following:

   > [DAT machines] will be sold in the United States under an agreement, scheduled to be announced this week, between electronics manufacturers and record companies that had opposed the technology for years because they feared it would increase the pirating of copyrighted recordings. While the agreement is a big victory for consumers, who will be able to make tape recordings with the distortion-free quality of compact disks, it is also brimming with uncomfortable lessons about the state of American competition with Japan.

   These two sentences are the only ones discussing the dispute over the machines' power to make perfect copies. The rest of the article uses DAT as an opportunity to bemoan the loss of US manufacturing capacity in small electronics. If this article still contained at least two more relevant sentences—for instance, arguing that we're allowing foreign companies to destroy our domestic music industry—it would still be included. As it stands, though, the above two sentences are all that is relevant, so it is excluded as irrelevant.

   Similarly, if an article merely discusses digital recorders as an economic good that may have certain effects on the economy—for instance, reviving consumer spending on retail electronics—this is not sufficient to meet Rule 2.

5. Discusses the technological characteristics of digital recording devices

   For instance, one article's content is best summed by this excerpt:

   > For Americans, the year's most far-reaching audio event was the introduction of digital audio tape (DAT) after a long delay caused by by [*sic*] legal and political wrangling about possible copyright violations. The timing of the introduction was inopportune. In an increasingly cloudy economy, few people were eager to spend nearly $900 on a new type of tape recorder that wouldn't play their old cassettes. But from a technical point of view, the advantages of DAT are indisputable and make its eventual success seem certain. And, as more companies enter the market with DAT recorders, competition will inevitably cause prices to drop. DAT is greatly superior to the familiar audio cassette. DAT is to conventional cassettes what CD's are to LP records -- both DAT and CD's make use of digital technology, with all its advantages: silent background, no tape hiss, frequency response to the upper and lower limits of audibility, rock-steady pitch that ends flutter and wow, and a dynamic range that extends from a whisper to a thunderclap.

The first sentence is the only relevant content in this entire article (meeting Rule 2B), so it would be coded negative. The bulk of the article discusses the technical advantages of digital audio recording, which is not sufficient for it to be coded as positive.

6. Discusses other copyright policy issues

    Examples include:

    A.   Copyright for computer software or the designs of semiconductor chips

    B.   Royalty rates for *other* uses of copyrighted works, e.g. songs played in bars and restaurants

    C.   International trade or treaty development

7. Discusses the operations of the US Copyright Office

Title I of the DMCA

Include if it meets ANY of the following:

1. Mentions the implementation of the anticircumvention provisions of the WIPO Copyright Treaty or the WIPO Performances and Phonograms Treaty or any of the bills toward this effect

    In 1995, Rep. Carlos Moorhead (R-CA) introduced H.R. 2441, the NII Copyright Protection Act of 1995. That same year, Sen. Orrin Hatch (R-UT) introduced S. 1284, a similar bill with the same title. In 1997, Rep. Howard Coble (R-NC) introduced H.R. 2281, the WIPO Copyright Treaties Implementation Act. In 1998, Sen. Hatch introduced S. 2037, the Digital Millennium Copyright Act of 1998.

    Any article mentioning any of these bills by name or number will generally be included (subject to the length requirements detailed above), but doing so is not required for inclusion. Content discussing a similar prohibition on the circulation of DRM would also be included in deciding whether an article is relevant.

2. Mentions the debate about whether to prevent the circumvention of DRM or the development, distribution, or sale of circumvention devices.

3. Participates in or describes the debate over the pros and cons of circumventing DRM.

    The point of this rule is to include articles that are part of the debate over whether and when circumvention is sometimes valuable. Articles that meet this rule without also mentioning the law are very rare, but this rule may help a document with an otherwise brief discussion of the law *per se* to meet the overall length requirement.

Do not include an article MERELY because it:

4. Discusses portions of the DMCA other than Title I, the anticircumvention provisions

    In particular, do not include an article merely because it discusses these other titles:

    A.   Title II of the DMCA sets up a notice-and-takedown procedure for copyright holders to contact internet service providers and demand the removal of copyrighted content from the internet; this limits the legal liability of online service providers.

    B.   Title IV of the DMCA includes six miscellaneous provisions, including one that sets up a new system for webcasters to pay royalties on music broadcast over the internet.

    C.   Title V provides copyright protection for the design of vessel hulls.

    As these ideas began to meld together during the legislative process, hearings began mixing the debates around the various proposals; one important point of this code is to separate these debates.

5. Discusses other policy issues involving digital media

    Examples include:

    A.   Unauthorized online transmission of media files

    B.   Digital performing rights in sound recordings

    C.   Techniques for film preservation

6. Discusses the operations of the US Copyright Office

DMCA Reform Bill

Include If It Meets ANY of the Following:

1. Discusses a proposed DMCA reform, including its practical, social, economic, or other implications. Any article that discusses the likely impact of proposed reforms in meaningful detail will be included regardless of whether that article discusses the arguments for or against such proposals.

2. Describes costs/benefits of DRM-specific portions of DMCA (17 USC §§ 1201-1204), especially the anti-circumvention provision of DMCA (§ 1201)

    Alleged benefits will generally be for the benefit to copyright holders, including decreased piracy, increased revenues, and increased investment in creative products such as music, movies, software, and video games.

    Alleged costs will be somewhat more variable. Potential examples include:

    A.   Legal threats scare academic researchers and thus discourage certain kinds of research or teaching.

Particularly likely to be mentioned on this count are computer science research into areas such as encryption and security research. For instance, consider one story about an academic conference at which scholars discussed the problems copyright presents for them in their teaching and research. (This topic would not, by itself, merit the article's inclusion.) Here is an excerpt:

> Edward W. Felten, a professor of computer science at Princeton University, was at the center of a legal battle in 2001, when representatives of the recording industry threatened to sue him and the university over the publication of a paper analyzing a set of digital watermarking technologies designed to secure music files. The recording industry based its claim on the 1998 Digital Millennium Copyright Act, which makes it a crime to circumvent antipiracy measures built into digital media.

> "After a long legal fight involving withdrawal and later resubmission of our paper, and our filing of a lawsuit against the parties who tried to suppress our work," Professor Felten wrote in response to a call from conference moderators for tales of copyright woe, "we won the right to publish our paper. Attempts to create a research exemption to the D.M.C.A. have failed thus far."

Because of the DMCA, Felten and his colleagues faced grave legal threats for conducting their research.

Another example of academic research that might be harmed is communication and media studies research using copyrighted works (e.g., movies on DVD) that are generally found in encrypted media. A professor who wants to make use of several DVDs in her class, for instance, may need to make excerpts to make class time more efficient. To do so, however, she would have to hack the encryption on the DVDs. This is another example of a relevant discussion of the costs and benefits of circumventing DRM.

Do *not* apply this rule for just any discussion of the copyright concerns of media studies scholars. The same article discusses the painful process of securing permissions from copyright holders. That is not relevant to this study.

B.   Legal concerns discourage certain kinds of publishing, speech, etc.

For instance, one article discusses a publisher's decision not to publish a book in light of legal threats due to the DMCA:

> ANDREW HUANG, an engineer and programmer in San Diego, has written a book called "Hacking the Xbox: An Introduction to Reverse Engineering." It has also been an introduction to copyright law in the digital age.

> Wiley Technology Publishing, a unit of John Wiley & Sons, agreed last year to publish the book. But after Mr. Huang delivered the manuscript five months ago, the publisher backed out over concerns that the Digital Millennium Copyright Act of 1998 made it illegal to disseminate information about how to circumvent copyright protection.

Consider what this implies on a larger scale: there are certain topics that cannot be published, and certain books that might not be written, because of this law.

C.   Legal concerns discourage certain kinds of archiving and preservation of digital materials

One article provides an excellent example of this rule. Here is an excerpt:

The personal computer industry began less than three decades ago, but already some of the early software programs that defined the era are an endangered species, the potential victims of "bit rot," according to a prominent digital archivist.

The warning came from Brewster Kahle, chairman of the nonprofit Internet Archive, who spoke at a meeting that the United States Copyright Office convened last week in Los Angeles. The session was held to discuss the impact of the Digital Millennium Copyright Act, the 1998 law meant to protect digital material from unauthorized copying.

Mr. Kahle (pronounced kale) said the law's stringent anticopying provisions, and the decay of the floppy disks and other magnetic media used to store early PC software, could allow early programs like Apple Computer's AppleWriter, a word-processing program, and the VisiCalc spreadsheet software to be lost forever.

Kahle is arguing that the DMCA's anticircumvention provision (mistakenly identified in the article as the "anticopying provision") prevents him from archiving certain digital materials, including important early software products.

D. Legal threats have slowed or discouraged companies from creating interoperable technologies

Many companies have used digital rights management as a tool for creating a monopoly on interoperable technologies such as machines, machine parts, and software. Then, when competitors reverse engineer around these DRM systems, the makers of the DRM-laden technologies sue under the DMCA. "You circumvented our DRM, which is illegal," they allege. Most scholars believe the issue here has little to do with copyright infringement. Rather, it is a matter of some companies using the combination of technology (DRM) and the law (DMCA) to prevent competition.

For instance, the printer company Lexmark installed a small computer chip into its printer cartridges to enable them to communicate with their printers. These chips contain a bit of secret code that the printer looks for. If the printer cartridge does not contain the secret code, the printer will not work with an otherwise compatible cartridge. Competitors such as Static Control Components, a small company that remanufactures printer cartridges, can work around this by cracking the secret code and building it into the chip on their printer cartridges. One article describes it this way:

Passed in 1998, the [DMCA] is designed to protect copyrighted works in an age when the material easily can be illegally copied and distributed over the Internet. The music industry uses the DMCA to sue Internet song-swappers it maintains are violating copyright law. But another provision of the law—Section 1201—expressly prohibits individuals from circumventing technological measures erected by copyright holders to protect their works.

Ever since, businesses that make products as diverse as voting machines, electronic pets and garage-door openers have turned to the law to protect their digital turf. Lexmark International Inc., one of the world's largest printer companies, joined the parade last December when it cited the law to sue Static Control.

Lexmark alleged that the company illegally copied some of the code used by computer chips in Lexmark cartridges to enable the remanufactured cartridges to work. The chips

monitor the level of toner and tell users when it is running low. More important, they make the cartridges compatible with the printer -- if the two do not execute an electronic "secret handshake" activated by the chip, the copier will not work.

By figuring out how to emulate that handshake, Static Control circumvented Lexmark's ability to protect its copyrighted works, Lexmark's attorneys argued. In February, Lexmark won an injunction that stopped Static Control from making its chips.

The question that arises in this instance is whether one company should be able to dictate the terms by which other companies can make interoperable products through the use of DRM and the DMCA. Obviously, Lexmark thinks this is a valuable and justifiable use of the law, while Static Control Components thinks it is terrible. Any document detailing any of these cases or debates about these cases is relevant to this study.

In each of these examples, somebody is alleging (implicitly or explicitly) that the DMCA prevents or discourages some socially valuable activity. These are the activities most commonly associated with longstanding exceptions to copyright: teaching, research, library lending and preservation, the creation of interoperable information products, etc.

One can debate whether this cost is more or less important than the benefit to copyright holders whose business models build upon this law. Nonetheless, it is clearly an impact of the anticircumvention provisions of the DMCA. Thus, any article discussing any of these topics is included as relevant.

As part of this rule, include articles that discuss whether the Copyright Office's triennial rulemaking is properly determining exemptions to the law against circumventing DRM protections. An article may be included if it includes voices arguing that the rulemaking is (not) properly balancing the interests of copyright holders and those who want to circumvent DRM— subject to the length guidelines outlined above.

3. Discusses the pros and cons of circumventing DRM.

As discussed in newspapers, this includes virtually any discussion of DRM circumvention. (About the only way DRM circumvention would not be relevant is if it were confined merely to technical discussion, which of course is not to be found in the newspaper.)

Topics considered include:

A.   Why do people circumvent DRM?

B.   When people circumvent, what are the social/economic/legal/moral/technological ramifications?

C.   Does circumvention lead to copyright infringement? A lot or a little?

D.   Does circumvention permit technological innovation? What does that look like?

E.   Who is happy about/defensive of circumvention? Who is unhappy? Why?

One article, on those who hack Microsoft's Xbox video game console, touches on many of these issues. A representative excerpt is here:

All sorts of new software is indeed running on Xbox consoles these days, and they are in fact becoming home-entertainment hubs, but it is not Microsoft doing the amazing.

Rather, an online confederacy apparently numbering in the thousands -- including accomplished hackers of varied motives and everyday technophiles like the Manhattan financial executive (who shared his experience on the condition of anonymity) -- is taking the lead. Those involved often call their efforts "unleashing" or "unshackling" -- freeing the Xbox to express its inner PC. Technology industry executives, however, often call such activity a bald attempt to hijack the Xbox illegally.

It is a battle that involves many of the ethical and legal issues facing the technology and media industries at this digital moment. What rights do consumers have to tinker with products they own? How far should companies go to protect their intellectual property? What happens when the desires of consumers conflict with the business models of companies they patronize? Who gets to decide just what a particular product may be used for?

Again, this article touches on most of the reasons to include an article with this rule. End users want to hack the Xbox to make creative uses of its powerful computing capabilities. This is one instance of the oft-repeated clash between the technically savvy folks who believe in the "freedom to tinker" and copyright holders who are trying to use technology to stop certain types of tinkering.

In applying this rule, pay careful attention for the debate around interoperability. Much of the debate about this portion of the DMCA revolves around the (in)ability of technologists to create products that integrate properly with systems that use DRM technologies.

IMPORTANT: *Do not* consider broadcast flag-specific discussion in applying this rule.

Do Not Include MERELY Because It:

4. Describes *other* parts of the DMCA. These provisions deal with topics such as:

    A. Copyright procedures for online service providers and service providers' liability for users' infringement

    B. Licensing for digital music services

    C. Circumstances under which distance educators do not need copyright licenses to use copyrighted works

    In each of these instances, an article may explicitly mention the DMCA. They are all part of the act because it is a big act that covers many topics. This study is only concerned with the provisions dealing with digital rights management technologies such as encryption and watermarking.

5. Describes the provisions of the law as it now stands. If it does not otherwise meet one of the rules for inclusion, an article should not be included merely because it contains a description of the anticircumvention provisions, such as a tutorial designed to teach somebody what would and would not count as a violation of the law.

6. Discusses the pros and cons of deploying one or more DRM systems, whether retrospectively or prospectively. For instance, some articles may discuss the debate over whether certain providers of

networked communication services (broadband providers, YouTube, universities, etc.) can or should deploy copyright-filtering technologies on their networks. This subject alone would not qualify an article for a positive coding.

Broadcast (Audio and/or Video) Flag

Include if it Meets ANY of the Following:

1. Describes the proposed broadcast flag regulations or legislation in enough depth to meet the length guidelines above. Any document that discusses the proposed regulations or legislation in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals.

2. Describes the potential social or economic impact of the broadcast flag and/or circumvention of the broadcast flag. Unlike DMCA Reform Bill Rule 5 above (under which one does not include articles merely because they discuss the pros and cons of deploying one or more DRM systems), this rule requires that one does include articles that participate in or describe the debate over whether the broadcast flag is a socially valuable DRM system.

   For instance, be sure to look for mentions of the Broadcast Protection Discussion Group (BPDG), the inter-industry group (similar to the DVD Copy Control Association) that developed the flag standard. Documents discussing the impact of the flag standard as developed by the BPDG are to be included under this rule.

3. Describes the potential copyright implications of the adoption of digital radio or TV broadcasting in an obvious allusion to the debate over whether or not to impose a broadcast flag. The most common example of this is the question of whether fans should continue to be able to record programming from radio or TV to be replayed later.

   For instance, an article arguing that the adoption of digital and/or HD radio or television broadcasts forces us to reconsider the balance of copyright law—e.g., we cannot allow fans to record digital content that is of such pristine quality and so easily reproduced—should be included. Likewise, any statement that takes the opposite stance, insisting that copyright should stay the same—e.g., consumers should retain the right to record digital radio in the same way they have recorded analog radio programming—is also to be included.

Do Not Include an Article MERELY Because It:

4. Is or includes proposed legislation or current law without comment. Any document that is *simply* a bill, statute, etc., will always be excluded.

5. Describes details that are not relevant to the debate about the social, economic, and political values at stake with the flag, its rollout, or home recording of digital broadcast. Examples include:

   a. The technical details of the technology

   b. The legal subtleties of the law as it now stands

   c. The FCC's (lack of) jurisdiction to impose such a mandate

d. Other details that do not otherwise shed light on the merits of the flag or its circumvention and/or the merits or costs of laws or policies that would require or limit its adoption.

None of these topics provide any insight into the costs or benefits of the flag or any flag-related policy changes.

Basic Document Details

At this stage, documents have been identified as relevant to one or more of the debates at hand. With these relevant documents in hand, coders are studying documents, seeking both minimal background information and identifying the overall rhetorical valence of the document.

This stage describes the process of describing documents based on three basic characteristics: date, document type, and kinds of organizations or persons represented.

Date

What year was a document produced? This test is most important to code in relation to internet documents, as newspaper articles and congressional hearings both have definitive dates.

Importantly, code for the document's original production date unless the document is described as having been substantially revised or updated—in which case, use the date of the latest revision. For instance, many websites will automatically display an updated copyright notice reflecting the year of the web visit; ignore such dates. If a website has archived a document from 1999, posting it beginning in 2004, enter the date as 1999.

If the date of a document's production is unclear, please leave that cell blank. For all dates, be sure to enter complete, four-digit figures (e.g., 2004) rather than 2-digit figures (e.g., '04).

Type(s) of Organization(s) Represented

All of the documents to be coded are authored by policy actors from one or more of the following groups. These are the groups that regularly participate in the copyright debate—even if some do so with much higher frequency than others.

For our purposes, treat these codes as *mutually exclusive*. If a person speaks on behalf of more than one sector or if a document quotes sources from multiple sectors, you will still mark just one sector. See the instructions below for the priority order, but to give an example, a newspaper article that cites people from more than one industry would be marked as being in the "News" category and only in that category.

Types of organizations include:

1. Media

    Examples include:

    > Media companies of any size (e.g., Paramount Pictures, Vivendi Universal) and other entertainment sectors with valuable media holdings (e.g., National Football League)

    > Lobbying organizations (e.g., Recording Industry Association of America, Future of Music Coalition, Digital Media Association)

    > Creative "Talent" such as authors, artists, performers, etc., whether well-known (e.g., Johnny Cash, Debbie Gibson) or lesser-known and simply identified as authors (e.g., Andrew Huang, author of *Hacking the XBox*)

2. Lawyers and legal associations

   If an attorney is identified as representing one or more other groups described here, categorize them with that group and *not* here. An attorney representing a media firm, for instance, counts as "Media" and not as a lawyer.

   This category is for anybody identified simply as an intellectual property attorney *and not* as representing a specific client in a case or topic under discussion. Disregard former occupations. For instance, when cited in a 2004 newspaper article, former Patent and Trademark Office chief Bruce Lehman is identified as a practicing attorney. Because that is his then-current occupation, he would count under this category—and *not* as an appointed government official.

   This also includes any statements or comments by officers of any legal association (e.g., American Intellectual Property Law Association).

3. Appointed government officials

   This is for government officials who were not directly elected by voters and who do not work in the offices of those who were elected.

   Examples include the Register of Copyrights, the Commissioner of Patents and Trademarks, or officials from the Department of Commerce. Also include officials who work for Congress generally, such as the Congressional Research Service. Additionally, judges and other judiciary officials go here.

   Finally, include representatives of international trade bodies such as the World Intellectual Property Organization (WIPO).

4. Elected officials

   Mostly, this means members of Congress. Also include congressional staffers and spokespersons, as well as committee staff.

5. News

   Apply this code to print news and to online documents produced by news organizations (e.g., Wired), with the singular exception of opinion pieces by authors who fall into one of the other groups on this list.

   Thus, for instance, every news article is coded as News, but an Op-Ed or letter to the editor would generally not be.

   Do *not* apply this code to persons speaking on behalf of the news business, e.g. an owner of newspapers or an industry trade group. Code these as representing media interests.

6. Technology firms, organizations, and trade groups

   This includes traditional firms in computer hardware or software (e.g., Dell, Microsoft) as well as consumer electronics (e.g., Toshiba, Tandy).

   This category also includes firms that produce and sell various DRM technologies such as watermarks and encryption (e.g., Macrovision, Digimarc).

This also includes trade associations of various flavors (e.g., the Consumer Electronics Association, the Business Software Alliance, the Home Recording Rights Coalition, Computing Professionals for Social Responsibility).

It also includes nonprofit organizations and other groups organized around nontraditional business models. Examples include the Free Software Foundation (FSF), the Mozilla project, and the Internet Corporation for Assigned Names and Numbers (ICANN).

The Association for Computing Machinery counts as both a scholarly organization and a technology group.

Many employees of technology companies speak publicly on copyright issues but disavow any right to speak for their employers. In these instances, *do not* code these speakers or authors within this category; classify them as "Other," as described below.

7. Scholars

This includes any public statements by those affiliated with institutions of higher learning who *do not* speak on behalf of their institutions. Generally, this means professors, whether their field is law (e.g., Lawrence Lessig), the humanities or social sciences (e.g., Siva Vaidhyanathan), or technical fields such as computer science (e.g., Ed Felten). Also include undergraduate or graduate students who are not speaking on behalf of any organization that falls into another category.

Also include scholarly associations (e.g., Society for Cinema and Media Studies).

Also include institutes housed at universities (e.g., the Berkman Center for Internet & Society at Harvard, the Stanford Center for Internet & Society, etc.)

Also include student groups (e.g., Free Culture).

The Association for Computing Machinery counts as both a scholarly organization and a technology group.

8. Nonprofit groups

These groups may be explicitly dedicated to advocacy (e.g., Public Knowledge, the Consumer Project on Technology). They may see themselves more as think tanks (e.g., the Progress and Freedom Foundation, the CATO Institute). They may even have a different mission still (e.g., Creative Commons).

This is for any nonprofit group that is not tied to one or more of the other sectors identified here.

Express skepticism when groups identify themselves as representing the public *and* a specific sector. For instance, the Home Recording Rights Coalition is really a technology industry group.

Two groups in this set are coded as representing both nonprofit and other sectors: the Digital Future Coalition (representing nonprofit, technology, scholars, libraries, and education) and the Chilling Effects Clearinghouse (nonprofit and scholars).

9. Libraries

This includes both librarians and representatives of library associations (e.g. the American Library Association).

This also includes law librarians and their associations; place them here, not with the lawyers.

10. Education

    This includes any educational administrators who speak for their institutions (e.g., university deans, provosts, or spokespersons) or anybody speaking on behalf of groups of educational institutions (e.g., the Association of American Universities).

11. Other (e.g., unaffiliated individuals, individuals other than scholars who disavow the right to speak for their employers)

Include a person as representing one of these groups even based on job title alone. For instance, if a newspaper article quotes somebody as a librarian, code the article as representing libraries. The only exception is when a person disavows the right to speak on behalf of their employer. In that case, code her or him as "Other".

For documents such as congressional speeches, the type of organization represented will generally be very straightforward; the speaker introduces herself, identifies her employer, describes their overall business model, and begins to speak to the debate at hand. If the speaker says she represents the Motion Picture Association of America, mark the document as representing the media.

For those documents representing people from multiple categories, code the document as representing the highest-ranked group on the following list:

A. News

B. Nonprofit groups

C. Scholars

D. Libraries

E. Education

F. Media

G. Lawyers and legal associations

H. Appointed government officials

I. Elected officials

J. Technology

K. Other

Thus, a newspaper or magazine article with relevant quotes from a congressperson, a scholar, and a librarian should be coded as news. An NGO founded by scholars, librarians, and tech sector representatives should be coded as an NGO.

Rhetorical Valence

This portion of the codebook deals with "rhetorical valence", which is another way of saying that coders are to identify which side of the debate a document advances. There are two sides to this debate; one side advances stronger copyright law (and thus, weaker exemptions such as fair use), and the other side argues for stronger fair use (and thus, weaker copyright law).

A coding unit will be characterized as falling into one of three categories. Either it is clearly in the **strong fair use** camp, clearly in the **strong copyright** camp, or in some sense in between these two extremes. This third option, **neutral**, may mean that a unit advances no position, and it may also mean that it makes an earnest attempt to consider both the strong fair use position and the strong copyright position.

In order to be coded as neutral by offering arguments from both sides, a coding unit must advance *relevant* arguments from both positions—arguments that deal to some degree with one or more of the four policy debates at hand. (Please review the included primer on these policy debates.) For example, consider a document that offers relevant arguments on the strong copyright side and strong fair use arguments that are not relevant to any of the four DRM debates identified in this study; this document would be coded as strong copyright because, for the DRM debate as defined, it only offers arguments on that side. This is true of only a few documents, but coders should be prepared to identify them and code them as within the appropriate camp (strong copyright/strong fair use).

IMPORTANT: a unit is not to be coded neutral if it presents the arguments of one side en route to rebutting these arguments. For instance, a strong copyright advocate may discuss one or more arguments of the strong fair use side and then rebut these arguments. In this case, the document should be coded as strong copyright and *not* coded as neutral.

Units will often be keyed to the specific policy proposals studied in this project. Thus, support for the AHRA, Title I of the DMCA, and any form of mandate for either the audio or the video broadcast flag is generally to be coded as being an argument for strong copyright. Opposition to any of these is generally coded as strong fair use. Likewise, support for Boucher's bill to reform the DMCA is generally to be coded as strong fair use, while opposition to that reform is strong copyright.

While many units are explicitly in support or opposition of a proposal, a unit is not to be coded as neutral simply because it disavows a specific policy position or fails to take an explicit position on the legislation at hand. For instance, consider the congressional witness who claims neither to be for nor against a bill mandating the video broadcast flag but then proceeds to spend her/his entire time voicing concerns that the proposal may be imprudent. This speech or written submission would be coded as "strong fair use" despite the witness's explicit claim of neutrality.

Finally, a document may be coded as being in one camp even though their stance on a particular piece of legislation would normally suggest being in the other camp. For instance, several copyright holders oppose proposed DRM regulations as not doing enough in their interest. Songwriters were opposed to early drafts of what later became the Audio Home Recording Act because, in addition to the imposition of mandatory DRM solutions, they wanted royalties from digital recording devices and blank media; several also suggested the DRM regulations were not strong enough to be effective. These are still coded as being strong copyright documents *even though* the authors oppose a bill that would mandate a specific form of DRM. Likewise, some believe Rep. Boucher's proposals to moderate the anticircumvention provisions of the Digital Millennium Copyright Act are inadequate—that the bans on DRM circumvention should be further diminished or repealed entirely. Arguments to this effect are still coded as being strong fair use even though they weaken the case for a bill that would have the effect of expanding the right to make fair use of digital media.

To help coders spot relevant arguments, this document first provides several example arguments from each coalition.

Then, it outlines the specific instructions for coding each unit, which come in three stages:

1. Code each *document* as strong fair use, strong copyright, or mixed/neutral

2. If mixed/neutral, identify each relevant paragraph

3. Code each *paragraph* as strong fair use, strong copyright, or mixed/neutral

Most documents fall cleanly into one of the two camps; for these, coding ends with Stage 1.

Example Arguments

The members of each coalition tend to make similar arguments. The following examples are illustrative, *not* exhaustive. Coders are *not* being asked to identify these arguments as codes, but the ability to spot them will help coders to identify the valence of coding units.

While these are not intended to be exhaustive, later research may warrant attempting to create such an exhaustive list. Toward that end, if one's coding or knowledge of the debate suggests that other important arguments seem to be missing from these examples, please bring this to the attention of the lead researcher.

This sub-section presents paired arguments from each coalition. Then, it presents a smaller list of examples of statements that suggest a document is mixed or neutral.

Strong copyright coalition

Those in the strong copyright camp support proposals to increase the scope and effect of copyright law and oppose efforts to scale back copyright. In the DRM debate under study in this research project, they generally support the DRM regulations in the AHRA and the DMCA, support the requirement of a broadcast flag on radio or television receivers, and oppose efforts to reform the DMCA.

In their support for these positions, advocates of strong copyright often make arguments from among the following, each of which includes an example (real or hypothetical) of what such an argument might look like:

1. Fair use is narrow. Exemptions to copyright, including especially fair use but also including other exemptions such as those for libraries, are or should be narrow in general, or they are inapplicable in a specific set of circumstances.

   "The electronics industry would stretch fair use so far that it would destroy the very purpose of copyright law."

2. New technology is a threat to copyright. A specific new technology (excluding DRM) or new technologies in general are a threat to copyright holders and thus must be contained.

   "Without some sort of restriction, DAT will kill the music industry."

3. Piracy is rampant. The degree and impact of infringement is out of control, is killing the copyright industries, has reached a level that must be stopped, is a worldwide epidemic, and so on.

"Millions of young people are stealing our works online, and something must be done to stem the tide of rampant theft."

4. The technology sector is responsible for causing and helping to fix this problem. Technology companies are morally culpable—and should be legally liable—for tools that enable customers' infringing uses.

    "Technology companies must design their products to reduce their infringing capabilities."

5. DRM is good. DRM systems are an important or valuable tool for copyright holders to prevent infringement.

    "The Serial Copy Management System will greatly reduce infringement."

6. End users accept DRM, or the only activities they are blocked from doing are infringing or otherwise not valuable.

    "Look at iTunes and DVDs; these have proven wildly successful in the marketplace." Or:

    "If you buy a song from the iTunes Music Store, all of the fair uses you could possibly need are built into the software; hackers who circumvent the restrictions are trying to get something they didn't pay for."

7. DRM regulations catch wrongdoers. Those who are arrested or sued under proposed or current laws that ban the circumvention of DRM are generally or entirely involved in infringement or other lawbreaking.

    "These would-be hackers should be discouraged from circumventing access controls. You have no right to break into my house just to learn more about how locks work."

8. DRM regulations reduce piracy.

    "Thanks to these actions enforcing the anti-circumvention provisions of the DMCA, there are now fewer pirated copies of these movies online."

9. There is strong inter-industry consensus behind a given DRM regulation. A proposed DRM regulation is or should be acceptable to its usual opponents because it represents an acceptable compromise. DRM regulations are necessary to get the content industry or other players on board with a given technology.

    "The recording industry, songwriters and publishers, and electronics industry have all come together to negotiate this 'DAT pact' so that consumers can get these high-end machines and copyright holders can be compensated fairly." Or:

    "Intel and four other companies developed the Digital Transmission Content Protection (DTCP) standards specification to provide protected digital entertainment in the home. Imagine never being able to record cable television on a digital video recorder (DVR) for later viewing. Or not being able to sign up for video on demand. Without DTCP, Hollywood studios and other content owners would have been reluctant to ever allow video on demand or pay-per-view digital movies, much less permit a DVR to receive digital television content. Their fear: piracy."

10. A proposal that would strengthen DRM regulations is basically sound. Even if modest amendments may be in order, these should not obstruct the bill's passage.

405

"We think the approach in this bill is basically sound. We just want to tweak a few things that may clarify it going forward."

11. The 1996 World Intellectual Property Organization treaties require implementing legislation that punishes those who make and distribute devices that circumvent DRM.

"If we don't prevent people from developing circumvention devices, we're reneging on our treaty obligations."

12. The triennial rulemaking held by the Copyright Office to determine exemptions to the DMCA is adequate for protecting fair use. Exemptions granted show that the system works. Thanks to the exemptions process, the few affected classes of users will be able to make fair uses of encrypted works.

"If you need to circumvent DRM, there's a vehicle for that. These folks haven't even applied for an exemption, and now they want to gut the law." Or:

"We had a good hearing that fairly represented all interested parties, and the Copyright Office determined that all this rhetoric about lost fair use rights doesn't match the empirical record." Or:

"It is important to understand the purposes of this rulemaking, as stated in the law... This is not a broad evaluation of the successes or failures of the DMCA."

13. A given proposal to put the force of law behind DRM is bad because there should be even stronger protections for copyright holders.

"Some executives in the music industry would have preferred more stringent restrictions on copying or a royalty surcharge on recording equipment and blank tape that would have been used to compensate copyright holders. 'It's quite easy to circumvent any of the technical solutions they put forward,' said Edward P. Murphy, president of the National Music Publishers Association in New York. Moreover, he said, technological solutions must be tailored for each technology, while a royalty would cover all formats."

14. A party accused of violating the AHRA or DMCA should be found guilty or liable. The law should be interpreted such that a socially undesirable act is discouraged.

"While working for Elcomsoft, Dmitry Sklyarov hacked the protections for eBooks, and he did so intending to sell this hack to the general public. The effect would be to destroy the market in eBooks, encouraging the widespread online distribution of these books as unencrypted PDFs. This is clearly a violation of the DMCA, and he and Elcomsoft should be held responsible for this behavior."

15. The market cannot be trusted to sort out the future of DRM. Laws against circumventing DRM prevent market failure.
"The rhetoric of 'let the market work' is disingenuous. The market needs rules to prevent free-riders."

16. The code in software designed to circumvent DRM is not speech and does not deserve First Amendment protections.

"The movie studios, on the other hand, deny that DeCSS involves any form of speech referring to it as a 'digital crowbar' designed to deliberately circumvent copyright protection technologies. They continue that the prohibition on its distribution is just the same as measures prohibiting 'the

provision of gambling devices, trafficking in satellite theft devices, and trafficking in cable signal theft devices' and is not a content based restriction on speech."

Strong fair use coalition

Those in the strong fair use coalition oppose proposals to increase the scope and effect of copyright law and support efforts to scale back copyright. In the DRM debate under study in this research project, they generally oppose the DRM regulations in the AHRA and the DMCA, oppose the requirement of a broadcast flag on radio or television receivers, and support efforts to reform the DMCA.

In their support for these positions, advocates of strong fair use consistently make arguments from among the following, each of which includes a hypothetical example:

1. Fair use is important and should be interpreted expansively. Exemptions to copyright, including especially fair use, are or should be broad or interpreted broadly in general or in a specific set of circumstances. Fair use is of preeminent importance, protects the First Amendment, trumps copyright claims, and so on.

    "They would have you believe that any unauthorized use is an infringement of copyright, which would be nice for them but would destroy fair use." Or,

    "Teachers need the ability to make use of video clips as they see fit without fear of being sued."

2. Copyright is a threat to new technology. The regulation of new technologies via copyright is a threat to the development of that technology or of new technologies in general. Copyright should not contain or constrain new technologies. The proposed DRM regulation might unnecessarily constrain the development of future technology.

    "This bill might be appropriate for the technologies of today, but what about tomorrow's technology? This proposal introduces unnecessary constraints into the process of technological innovation. We need to keep the copyright lawyers out of the laboratories."

3. There is little real threat from piracy. The level of infringement is overstated, has little negative effect on copyright holders, or has some positive effects.

    "Internet piracy is not the reason for the downturn in music sales."

    Alternately, piracy may or may not threaten specific economic interests but will not harm society at large.

    "Musical creativity won't come to a grinding halt just because the RIAA's business model dies. Those who can adapt will still make music and will profit handsomely."

4. Individual users are responsible for their own behavior. Technology companies are not or should not be liable for tools that enable customers' infringing uses.

    "If you want to sue users when they break the law, fine, but don't come to us and demand that we design less useful products."

5. DRM has important social, political, or economic costs. DRM gives copyright holders a censorious veto over criticism, DRM locks consumers into specific technologies, which reduces the incentive to innovate new technologies, and so on.

407

"Sure, these locks might help reduce infringement, but they've locked out criticism and commentary." Or,

"These technologies aren't compatible with one another, so a user who buys an iPod has to buy their online music from iTunes; that creates lock-in, which reduces competition and innovation."

6. End users reject or resent DRM. Rights such as the ability to make backup copies, experiment with one's media collection, or engage in other DRM-blocked activities are valuable and important to end users.

    "A consumer buys a DVD burner, connects it to his DVD player, and even though he's invested all this money, he can't make a backup copy of his daughter's movie. He's mad, but he doesn't think, 'Oh, I guess this has DRM on it.' All he knows is that it doesn't work, so he thinks it's broken. And in a way, it is."

7. DRM regulations catch innocent folks. Otherwise law-abiding, well-meaning citizens are or might be arrested or sued under proposed or current laws that ban the circumvention of DRM systems.

    "Under the DMCA, encryption researchers have been threatened, sued, or arrested merely for doing their jobs."

8. DRM regulations do not reduce piracy.

    "Regardless of whether a song, movie, or software program is released with DRM, it's available online for free, regardless of the DMCA." Or,

    "In general, we believe that serious questions remain as to whether the broadcast flag proposal will be sufficiently effective."

9. There is substantial disagreement about a given DRM regulation, or important groups (e.g., consumers, educators) have been left out of a so-called compromise.

    "This so-called 'DAT pact' was negotiated without consumer input. Consumers don't want to pay higher prices for digital audio technology that is crippled by a copy management system." Or,

    "The [Broadcast Protection Discussion Group] Final Report represents the deliberations of a group that was express limited in its mission, which was to 'evaluat[e] **technical** solutions for preventing unauthorized redistribution' of digital TV content (emphasis added). ... Over time, however, as other technical and policy issues are dealt with, a broader consideration of consumer concerns will be needed, and this process must include consumer organizations as well as industry."

10. A proposal that would strengthen DRM regulations is so flawed that it should not be adopted without substantial alterations (e.g., much broader exemptions) or should not be adopted at all. Another proposal should be adopted in its place.

    "Tomorrow morning I intend to offer some comments and recommendations with respect to the other important issue that we have pending before us, and that is the anti-circumvention question, and what we refer to in the last Congress as [the] Section 1201 set of issues. It is my intention, Mr. Chairman, to offer separate legislation on this question that will provide what I think is a more promising approach than the bill that is pending before the Subcommittee at the present time."

11. The 1996 World Intellectual Property Organization (WIPO) treaties DO NOT require implementing legislation that punishes those who make and distribute devices that circumvent DRM.

    "The subcommittee should know that the WIPO convention which drafted the treaty rejected language which would have required punishment of device manufacturers. … It is not required by the treaty. And yet, the device-oriented approach having been rejected at Geneva by the treaty conference, has now resurfaced and is at the core of the legislation proposed by the Administration to implement the treaty."

12. The triennial rulemaking held by the Copyright Office to determine exemptions to the DMCA is inadequate for protecting fair use. Despite exemptions, the system is broken.

    "The Copyright Office rulemaking can't exempt the tools to circumvent, so even if I get the right to circumvent, unless I'm an expert programmer, I need somebody else to break the law just so I can exercise my rights."

13. A given proposal to diminish the force of law behind DRM does not create wide enough exemptions or go far enough to repeal a DRM regulation.

    "Boucher's DMCA reform bill is substantially watered down. We should just repeal the bans against circumvention and against trafficking in circumvention devices."

14. A party accused of violating the AHRA or DMCA should be found not guilty or not liable. The law should be interpreted such that a socially desirable act is permitted.

    "All Skylink did was design remote controls compatible with Chamberlain's garage door openers. They did not infringe on Chamberlain's copyrights along the way, so the DMCA should not apply. That law is intended to prevent massive copyright infringement, not protect device manufacturers from competition in the market for compatible devices."

15. It is fine for content providers to deploy DRM, but the government should not give those DRM systems the force of law.

    "The marketplace should determine the success or failure of DRM technologies but, increasingly, content distributors are turning to legislatures or the courts to erect new legal mandates to replace long-standing copyright regimes. DRM systems should be mechanisms for reinforcing existing legal constraints on behavior, not as mechanisms for creating new legal constraints."

16. The code in software designed to circumvent DRM is speech and does deserve First Amendment protections.

    "In its reply brief, the Electronic Frontier Foundation, on behalf of the defendant, argues that 'DeCSS itself has no non-speech elements' and similarly that its 'dissemination .. by a member of the media covering an issue of public concern is pure speech.' They liken the computer program to 'blueprints and instructions for a photocopier, recipes, books about fixing cars, and videos on baby care' and argue that just because somebody 'might use [it] to do something' does notmean that it is any less protected as speech."

Again, these example arguments are not exhaustive. Look for arguments from either list, but also look for similar arguments; once you spot enough to know that a coding unit falls into either camp or somewhere between, you are done with that unit.

Neutral or Mixed

These are examples of statements that suggest a document is neutral or mixed. These are generally not controlling; if a document contains one or more neutral or mixed points as well as one or more points suggesting the document's author supports one of the coalitions, that document will usually not be coded as neutral.

Use your judgment. If a document seems to be making an honest attempt to present both sides to an issue, please code it as neutral. If a document presents both sides but chooses one, please code it as supporting one coalition.

1. A proposed DRM regulation is soon going to be considered, is actively under consideration, or has been passed.

    "On November 4, the Federal Communications Commission published a Report And Order And Further Notice Of Proposed Rulemaking approving the 'broadcast flag' scheme for copy protection for broadcast digital television."

2. There is an ongoing debate over a given DRM policy proposal. Both sides have something to say on the issue of DRM regulation.

3. A given authority is mixed or hedged regarding the wisdom of a given proposal.

    "In its recent order, the FCC acknowledged that substantial controversy surrounded this proposed process [for approving broadcast flag technologies]; in particular, commentators were concerned that such a process could give one industry inappropriate authority over another." Or,

    "We approach our task with an open mind and have not drawn conclusions, but it is proper to [consider imposing a broadcast flag mandate] due to our commitment to the DTV transition. ... At this point, we have drawn no conclusions that a 'broadcast flag' system is necessary or appropriate, or that the Commission has jurisdiction to adopt such a system. Nevertheless, I believe it is entirely fitting and proper that the Commission undertake this examination."

4. Some proposals for exemptions to the DMCA should be accepted, while others should be rejected.

    "Several proposed classes of works were justified by enough empirical evidence to support granting these exemptions, but several more classes were not so justified."

5. One form of DRM mandate should be implemented, but another is not yet ready for implementtaion.

    "Dealing specifically with the broadcast video flag, I am glad to see that all stakeholders worked together to reach an agreement and, given the court's decision on the FCC's ability to implement such rules, it is now incumbent upon us to act in a responsible manner to ensure the protection of the video content. A matter of greater contention surrounds the issue of audio flag. Unlike the video flag, I believe it would be in the best interest of all consumers and the future of the industry to sit down with all affected parties and again try to work out a privately negotiated agreement rather than rely on a legislative solution."

First, coders are to identify the valence of the overall document. Documents will fall into one of three mutually exclusive categories. The key question is:

> Is this document clearly in support of either the strong copyright position or the strong fair use side?

If the answer is "yes" because the document is clearly in either camp, code it as such. Valence coding for that document is then finished.

If the answer is "no" because the document is not clearly in either camp, code the document as neutral. Some of these documents offer arguments from both sides of the debate, considering the interplay between them in a more-or-less balanced manner. Others simply describe the proposals that are the subject of debate and provide a sense of how these policies would change facts on the ground. Any document not cleanly in either camp is coded as being neutral.

IMPORTANT: Do not code a document as neutral if it includes arguments from one camp only to refute or rebut them. For instance, a document might argue:

> The motion picture industry claims that peer-to-peer trading is a threat to their ability to make money. I will concede that it's a threat to their *current* business model, but that's no more profound than saying that the car was a threat to the horse carriage industry. They need to adapt. People will pay for legitimate peer-to-peer downloads, and Congress has no obligation to prevent the music industry from going through these growing pains—especially at the expense of technological innovation. We need to keep the lawyers out of the laboratories.

Based on this paragraph, at least, this document would be coded as being within the strong fair use camp. Even though it includes arguments from both sides, it only includes a strong copyright argument ("peer-to-peer trading is a threat") in order to rebut it.

Coding for document valence often requires a reasonably thorough survey of the document. For a slim minority of documents, it will be so clear that coders can stop after reading the title and first paragraph of a document. In most cases, though, coders will not need to read every word, especially not in the case of longer documents. The overall purpose of most documents in the population will be clear after a summary read-through.

If a document is neutral, coders move to the second and third stages. At that point, the coding unit becomes the paragraph. Before moving to that part of the codebook, however, consider some examples of each category of document:

Example Documents: Strong Copyright

A document need not express categorical cheerleading for a proposal to expand copyright to fall into this category. For instance, one document examining the DMCA concludes:

> A prudent compromise approved by the House Commerce Committee last week would delay the anti-circumvention rule for two years while the Commerce Department and the Federal patent and copyright offices study the effect of the prohibition on users. The Commerce Secretary could waive the rule for any class of works where technological shields were impeding the lawful use of

411

copyrighted matter. The situation would be reviewed every two years. Both the content producers and the libraries and schools are willing to accept this more fluid approach. Congress should adopt this plan in the final version of the digital copyright legislation.

Because it says explicitly, "Congress should adopt this plan," it falls into the strong copyright category.

In response to proposals to reform the DMCA, a strong copyright document argues against reform and in favor of the status quo. For example, one document considers proposals to reform the section of the DMCA that concerns encryption research. It evaluates comments on both sides, but here is the very last sentence:

> As such, we conclude that is it premature to suggest alternative language or legislative recommendations with regard to Section 1201(g) of the DMCA at this time.

By coming down on the side of those defending the DMCA's current form, this document gets classified as strong copyright.

One tricky issue around the broadcast flag is the question of whether or how the FCC should limit restrictions on the flag's use. If the flag passes, should broadcasters be able to use the flag to prevent any recording of any material whatsoever, or should the FCC ensure that end users can make uses of certain kinds of materials, such as news and public affairs programming? Should the flag allow a certain amount of copying, and under what circumstances?

Those in the strong copyright camp generally oppose efforts to limit broadcasters' ability to use the flag on certain types of programs, and they generally oppose efforts to expand the potential noninfringing uses of flagged content. One person makes the following argument:

> I have no problem with the FCC on a policy basis mandating use of the broadcast flag technology. While I am generally opposed to broad Government mandates on technology, I have long considered it appropriate in limited circumstances for the Government to order the use of certain tesensus has emerged. …
>
> That being said, I do have some concerns about the broadcast flag rulemaking, in particular, what some parties are asking the FCC to do.
>
> Numerous comments have been filed asking the FCC to ensure that any broadcast flag technology allows consumers to make various uses of the digital TV programming it protects. These commentators purport to cite various copyright law doctrines, including first use, as the Chairman discussed, and first sale, as guaranteeing consumer utilization of copyrighted TV programming in the ways they hope to protect.
>
> It is these claims about copyright law and the role of the FCC in analyzing them that gives me pause about the broadcast flag rulemaking. I am unaware of any precedent for the FCC interpreting the Copyright Act as part of an FCC rulemaking or in any other capacity, nor am I aware, for that matter, of the FCC ever mandating that copyright owners surrender any of their exclusive rights to consumers.

This person clearly supports the broadcast flag and opposes efforts to reduce the flag's impact. This is a strong copyright document.

Documents may be neutral even if they are quite one-sided. For instance, a newspaper article that seems rather one-sided may still present the other side in earnest. Look for language that indicates a legitimate debate. A good example is:

> To [the DMCA's] proponents—the studios and their allies—copyright legislation would unleash digital commerce by toughening prohibitions against pirating movies and other data from the Internet. Critics don't object to that but claim that the proposed anti-piracy rules are so tight that they would restrict the free flow of information. They would (say critics) create a "pay-for-use" world in which publishers, studios and others could lock up more and more information in computer memory—which could be used or seen only by dropping coins into a cyber meter.

> There is a legitimate debate here.

Rarely will examples be this explicit, but something to this effect should be adequate to illustrate that a document really is neutral.

A somewhat less explicit example reproduces several arguments against the broadcast flag proposals, but then contains the following phrase:

> Not everyone felt it was time to develop an audio flag, given that the HD radio industry was still emerging. Each side in the debate accused the others of stalling or otherwise acting in bad faith. Others felt that the legislation would undercut existing fair use provisions regarding home recording. Content providers argued that it would not, as they had no issue with timeshifting or recording during broadcast.

A paragraph such as this will generally be adequate to move a document into the "neutral" category.

The DMCA reform debate also has the occasional less-than-perfectly-neutral document that still qualifies as neutral. In one hearing, Boucher's proposal to reform the DMCA is under discussion. One speaker says that the DMCA earned his initial support, but it is worth reconsidering at this time. The passage that clinches this as a neutral document is as follows:

> It has now been six years since the DMCA became law, and it is important for this Committee to review its progress. These hearings will allow us to explore whether the DMCA has achieved a proper balance after all, or whether further action is required. The Committee on Energy and Commerce should closely examine the current system to find the appropriate balance that protects scholarship, research, and innovation while protecting the legitimate interests of copyright owners.

Here is another example:

> Speaking at the Future of Music Coalition policy summit, Baird said critics of DRM and U.S. laws that make it illegal to bypass anti-piracy technologies are focusing on "a sliver of an industry." In the evolving "ethereal world of digital distribution, everybody is trying to fight to keep or get a piece" of the market, he said.

> But Michael Geist, an Internet researcher at the University of Ottawa, has waged a war against DRM because he said the tools "can be used to take away rights that consumers have." Although a Copyright Office review of the Digital Millennium Copyright Act found no major problems with its anti-circumvention language, Geist said the law is flawed. He noted a "more unbiased review" of anti-circumvention methods by an Australian parliamentary panel, which flagged dozens of concerns.

413

A few neutral documents are exceptionally long transcripts of conference panels and the like. For example, one such document begins with a very long pro-broadcast flag statement, and only after many pages does somebody on the panel state a reason not to support the broadcast flag. Deep within this document, one finds this statement:

> I think the broadcast flag really just took the first approach that was offered. You know, the [Federal Communications] Commission essentially took the first approach to this problem that was offered to it, and it's not the effective one. It's not the one that works. And it's one that imposes external costs that are really intolerable, both economic costs generally and economic costs of specific industry sectors and to consumers. And I think that's a mistake.

Thus, this document may appear to be strong copyright, but deep within, it becomes clear that it is a neutral document.

Example Documents: Strong Fair Use

Sometimes, a document will take be very explicit about a strong fair use stance at the very beginning. For instance:

> Two key Congressional committees have approved legislation on the transition to digital television (DTV) without having broadcast flag language added. That's the good news. The bad news is that there is still plenty of room for mischief in the legislative process.

This author thinks the broadcast flag is a bad decision—a strong fair use position—and opens with that opinion.

Other documents will be a bit more difficult to categorize but will still make clear the call for stronger fair use. One document evaluating the impact of the DMCA concludes:

> As mentioned previously, CCIA and others anticipated the need for additional exceptions of this nature and we are not surprised that the advance of technology and innovation has exceeded the foresight of the drafters of the DMCA. For this reason, we argued that acts of circumvention should be unlawful only when tied to actual acts of piracy or violations of copyright. We continue to urge the Copyright Office and Congress to revisit this critical issue and consider the harm to innovation and competition in the software industry that is already apparent from enactment of the DMCA. At a minimum, the Copyright Office and Congress should consider clear exceptions to cover circumvention of technological protections in instances similar to those outlined above.

The thrust of this document is that the DMCA should be loosened to allow for more kinds of circumvention; while not as obvious as the first example, this still puts this document solidly into the strong fair use coalition.

Identifying Relevant Paragraphs

Documents have already been identified as relevant using the same criteria discussed below. Now, code any paragraph as relevant if it has *at least* one relevant sentence.

Technical Matters

Ignore technical discussions that are not immediately connected to the effects of those technologies on stakeholders (e.g., users, copyright holders, manufacturers).

For instance, as described more fully in the briefing on the AHRA, the law mandates a DRM system called Serial Copy Management System, or SCMS. If one is faced with a highly technical document that outlines the circuitry specifications of SCMS, it is not necessary to go through several pages of such highly technical descriptions of the underlying specifications. Consider only paragraphs that discuss this design's effects in restricting copies.

Likewise, some paragraphs may describe the technical characteristics of another DRM technique or the technical specifics of the broadcast flag. Unless these are immediately tied to one of the criteria below for relevance, please do not code a paragraph as relevant based only on these terms.

Reading in context

In determining whether a paragraph is relevant, keep in mind the context of the neighboring paragraphs in making this determination. Look at the paragraph before and after. For example, consider the following two paragraphs:

A.   There are so many battles going on about the Internet these days -- control of privacy, sex and con artists, to name a few -- that you might not have focused on one concerning copyrights. But a lot of other folks have, and their ongoing war has reached a crucial stage.

B.   Last Thursday, by a 99-to-0 vote, the Senate approved the Digital Millennium Copyright Act, a bill that would implement two copyright treaties adopted in 1996 by the United Nations' World Intellectual Property Organization to cover property rights over written material, sound recordings and software in the online world. What's got some people upset are sections that they say would restrict access for private use to a whole range of material on the Net.

Paragraph B discusses the DMCA quite explicitly. In this context, Paragraph A is clearly relevant; the author contends that the "ongoing war" over copyright has reached a crucial stage, meaning the Senate's unanimous passage of the DMCA. As read in context, both paragraphs are relevant. Thanks to the second paragraph, some of the vague words in the first clearly come to say something directly about the DMCA.

This is not to say that any paragraph adjoining a relevant paragraph is always relevant, even if it discusses copyright. Consider the next paragraph from the same document:

C.   To explain: The notion of copyright in this country is as old as the Republic itself, with the Constitution giving Congress the power "To promote the Progress of Science and the useful Arts, by securing for limited times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries. . . ."

This paragraph is about copyright, it is mentioned to make a point in the debate about the DMCA, and one would correctly suppose that the author will come back around to discussing the DMCA more directly; all the same, this paragraph is not relevant, even when read in context.

Include a paragraph if it meets the following:

1. Discusses the specifics of bills that require the implementation of DRM in digital audio recording devices such as DAT decks

   In 1990, Sen. Dennis DeConcini (D-AZ) introduced S. 2358, the Digital Audio Tape Recorder Act of 1990. In 1991, he introduced S. 1623, the Audio Home Recording Act. That same year, Rep. Jack Brooks (D-TX) introduced H.R. 3204, similar legislation of the same title. Rep. Cardiss Collins (IL) also introduced an "Audio Home Recording Act" in 1992; this one was bill number H.R. 4567; consideration of Collins' bill became folded into the debate over Brooks' bill, H.R. 3204.

   Any paragraph discussing any of these bills by name or number will generally be included, but doing so is not required for inclusion. Any mention of a similar requirement for the implementation of DRM in digital audio devices would merit including the document as relevant. Note that relevant paragraph may present a very heated discussion of a bill or offer a numbingly dry and exactingly neutral description of the legislation's meaning or prospects; both kinds of paragraphs are included, as are those that fall somewhere in between. Here is an example of the latter:

   > This technical reference document is provided to facilitate the implementation of legislation relating to digital audio recording ("DAR") devices, known as the "Audio Home Recording Act of 1991" ("the Act").

   > This Techncial Reference Document establishes the standards and specifications that are necessary to implement the Serial Copy Management System ("SCMS") under the Act. It draws in part from specifications proposed…

   This document is mostly a highly technical discussion of the specifications for SCMS—which, as detailed below, is not adequate for a paragraph's inclusion as relevant—but these two paragraphs both mention the AHRA specifically, so they are both included.

   Some of these bills, including the final law as passed, also include royalty provisions; the end result is effectively a tax on digital audio recording devices, with the money divided among various stakeholders in the music industry. Many paragraphs discuss these bills primarily or even exclusively in terms of the royalty provisions. Some may even discuss royalties without mentioning the bill by name. As with the DRM mandate, include documents that discuss the debate over whether to impose royalties on digital recorders and digital media.

   This rule is triggered ***even if*** a congressperson or other speaker or author includes a few vague platitudes about the bill without discussing the specifics—that is, so long as such vague platitudes are at least specifically tied to the bill at hand.

2. Discusses the likely economic, social, or cultural impact of the ***copying*** enabled by the introduction and potential mass adoption of digital audio recording devices such as DAT decks

   This might include a recording industry executive predicting that DAT will be the death of her industry. This could also include an electronics manufacturer discussing the legal threats that delayed the introduction of DAT to the US market.

416

To meet this standard, a paragraph must discuss either:

A.   The effects on copyright holders, which here will generally mean the music industry, including record companies (e.g. RCA, Capitol), music publishers (e.g., ASCAP, BMI), songwriters, performers, etc. OR

B.   The effects of copyright concerns on the marketing or uses of digital recording devices.

Most of the examples of paragraphs meeting Rule 2(A) will be copyright holders and their allies bemoaning the effects of digital copying on their business model. These should be easy to spot. But also include documents that make the opposite case—that Digital Audio Tape specifically or digital audio recording devices generally present no real threat to copyright holders. Either side of the argument is grounds for including a document.

Here is an example of a paragraph that meets rule 2(A):

> Digital audio tape machines use the technique employed in compact disks to record and play music that is sharp and free of noise. But the machines, unlike compact disk players, allow consumers to tape their own music. Record companies, which fear that the ability to make high-quality tapes will reduce sales of compact disks, have fought to keep the machines off the market.

In addition, be sure to keep an eye out for documents arguing that the legal threats from copyright holders have (or have not) slowed or stopped the development and marketing of digital audio recording devices. (This is described in Rule 2(B).)

Consider an example:

> A group of music publishers, including the songwriter Sammy Cahn, has filed a lawsuit seeking to halt the sale in the United States of new digital audio tape recorders, contending that they contribute to copyright infringement.
>
> The suit, filed as a class action on behalf of all music copyright holders, contends that the machines are "inaugurating a new era in unauthorized home taping" of records and compact disks, depriving songwriters and music publishers of royalties they would receive from sales of such records and disks.
>
> The lawsuit was filed Monday in the Federal District Court in Manhattan against the Sony Corporation, which last month became the first company to begin selling digital tape recorders to American consumers.

This excerpt shows both rules 2(A) and 2(B). The first two paragraphs clearly qualify under 2(A); copyright holders are worried about the ease of making perfect copies and have taken action. The first and third paragraphs also count under 2(B), since they discuss the potential impact of these legal threats on Sony's ability to sell the devices.

Keep an eye out for allusions to the *Sony* decision, which held that the makers of VCRs could not be held liable for infringing copying by consumers. The debate over DAT was often explicitly linked to this decision, and such comparisons will generally meet this rule. For example:

> The action promises to raise anew the controversial subject of home taping and sets the stage for a replay of the so-called Betamax case. In that case, also involving Sony, the Supreme Court in 1984 ruled that video taping of television programs off the air for home

viewing did not constitute copyright infringement and that Sony was not violating the law in selling its Betamax videocassette recorder.

"It's deja vu," said Charles Ferris, the counsel to the Home Recording Rights Coalition, a lobbying group that supports sales of the digital tape recorders.

The lawsuit names only Sony, which is based in Japan, and its American subsidiaries, and not any consumers who would go out and buy a digital tape recorder . The suit also does not seek to stop sales of the older analog cassette recorders.

All three paragraphs are relevant. The first two (the second when read in context) connect this suit to the Betamax case—in other words, to the debate over whether copyright should restrict new technologies. The third paragraph is an application of rule 2(B); the suit would stop digital, but not analog recorders, and users are not threatened. In short, this this statement highlights the extent to which the legal restrictions would (or would not) restrict technology and its uses.

Here is another example of Rule 2(B):

Digital audio tape recorders have been kept off the market for two years by a copyright dispute between record companies and stereo equipment manufacturers. With that dispute apparently resolved, the new product faces what could be an even more daunting barrier: consumer acceptance.

This paragraph states that the copyright dispute harmed DAT manufacturers' ability to sell their wares, thus qualifying as relevant.

Here is one final example of both rules in play:

SCMS is intended to prohibit DAR [digital audio recorder] devices from recording "second-generation" digital copies from "first-generation" digital copies containing audio material over which copyright has been asserted via SCMS. It does not generally restrict the ability of such devices to make "first-generation" digital copies from "original" digital sources such as prerecorded commercially available compact discs, digital transmissions or digital tapes.

Currently, the predominant type of DAR device offered for sale in the United States is the DAT recorder… Additional types of DAR devices and interface formats are being or may be developed The standards and specifications in this Technical Reference Document are not intended to hinder the development of such new technologies but require, in accordance with Section 1021(a)(1)(A)-(C) of Subchapter C of the Act, that they incorporate the functional characteristics of SCMS protection. …

The first paragraph discusses SCMS as a tool for addressing the problem of unauthorized copying (Rule 2(A)), and the second reassures the reader that the copyright problem will not prevent new technological development (Rule 2(B)). Thus they are both relevant—though note that the second paragraph is also relevant based on mentioning the AHRA explicitly.

3. Discusses the role of the Serial Copy Management System (SCMS) in restricting copies

SCMS was the specific technology that was included in the final legislation. All consumer-level digital audio recording devices, such as DAT, must now include this system, which permits the user to make unlimited copies of a digital source (e.g., a CD or a DAT tape) but not copies of copies.

Several paragraphs discuss SCMS, but not all call it by name. Under this rule, discussion of SCMS or any similar technology that limits digital reproduction of audio recordings, whether named explicitly or not, is relevant. For example:

> But recording and stereo-equipment companies from the United States, Europe and Japan have now reached an agreement, to be formally announced this week, that will clear the path for the machines to be sold in this country. Under that agreement, the digital tape machines will be designed to allow the taping of compact disks but to prevent the subsequent copying of the digital tapes, say people familiar with the plan.

This agreement is to implement SCMS, which coders should recognize based on the primer on this debate.

Here is another example:

> SCMS is intended to prohibit DAR devices from recording "second-generation" digital copies from "first-generation" digital copies containing audio material over which copyright has been asserted via SCMS. It does not generally restrict the ability of such devices to make "first-generation" digital copies from "original" digital sources such as prerecorded commercially available compact discs, digital transmissions or digital tapes.

This paragraph discusses the anti-copying effect of SCMS, so it is relevant.

4. Describes the bill's supporters or opponents with some sense that there is a *debate* afoot

For instance, one congressperson introducing a hearing has the following paragraph:

> After that, we will hear from the supporters of S. 2358, Mr. Jason Berman, President of the Recording Industry Association of America; …

He continues listing supporters from there. Because the debate over a bill is often a matter of creating the perception of support or opposition, this is a small contribution to the strong copyright side. As such, it is relevant.

As another example, consider the following:

> Legislation intended to strengthen international copyright protections in cyberspace has run into an unexpected enemy: some of the nation's best computer security experts.

> The proposed bill would ratify two international treaties negotiated under the authority of the World Intellectual Property Organization, both of which have the backing of some of the nation's largest media companies and software publishers.

The first paragraph identifies computer security experts as opposing the bill, and the second identifies the bill's supporters. These are therefore both relevant.

Introducing a sponsor of a bill will generally count as identifying a supporter and therefore be relevant.

Do not include a paragraph MERELY because it:

5. Discusses the economic impact of digital audio recording devices such as DAT *without* reference to the devices' impact on copyright industries.

    For instance, a paragraph that merely discusses digital recorders as an economic good that may have certain effects on the economy—for instance, reviving consumer spending on retail electronics—is not included.

6. Previews the witnesses who are about to speak or simply identifies those who spoke

    Here is a longer excerpt from the congressperson above who was introducing a hearing:

    > I look forward to hearing from all the witnesses that we have gathered this morning to speak on this bill. First, we will hear from the Registrar of Copyrights, Mr. Ralph Oman.

    > After that, we will hear from the supporters of S. 2358, Mr. Jason Berman. …

    > Then we will listen to the testimony of Mr. Philip Greenspun, President of Isosonics Corporation; Mr. Edward Murphy, President and Chief Officer of the National Music Publishers Association; …

    As described above, the paragraph identifying Mr. Berman and company as *supporters* is relevant. Merely listing witnesses, as the speaker does in the first and third paragraphs above, is *not* relevant.

7. Discusses the technological characteristics of digital recording devices or SCMS, as detached from its capacity from stopping copies.

    Unless they also meet one of the rules for inclusion detailed above, do not include documents discussing the technical advantages of digital audio recording (e.g., higher audio fidelity) or debating the merits of various formats on technical grounds.

    For instance, one irrelevant paragraph is as follows:

    > Digital tape recorders record music in the same way as information is stored in computers, as a series of ones and zeroes. The advantage of digital tape recorders is that they can make essentially perfect copies of a compact disk, which also uses digital technology.

    This paragraph is purely technical. Thus, it is excluded from this study. Here are excerpts from two additional example paragraphs from the same document:

Title I of the DMCA

Include if it meets ANY of the following:

1. Discusses the implementation of the anticircumvention provisions of the WIPO Copyright Treaty or the WIPO Performances and Phonograms Treaty or any of the bills toward this effect

In 1995, Sen. Orrin Hatch (R-UT) introduced S. 1284, the NII Copyright Protection Act of 1995. That same year, Rep. Carlos Moorhead (R-CA) introduced H.R. 2441, a similar bill with the same title. In 1997, Rep. Howard Coble (R-NC) introduced H.R. 2281, the WIPO Copyright Treaties Implementation Act. In 1998, Sen. Hatch introduced S. 2037, the Digital Millennium Copyright Act of 1998.

Keep an eye out for paragraphs mentioning these by name, though doing so is not required for inclusion. Discussing them generally in a way, in such a way that the bill is obviously the object of discussion, is adequate.

IMPORTANT AND TRICKY: Several of these bills contain provisions that are totally unrelated to the implementation of the WIPO treaties and the DRM regulations that eventually became Title I of the DMCA. For instance, the 1995 bills (H.R. 2441 and S. 1284) and Sen. Hatch's 1997 bill (S. 2037) also contained provisions dealing with completely separate topics; particularly common is the issue of online service provider liability for end users' behavior. This is not relevant.

To help clarify how to handle documents discussing each of these bills, see the following table:

Table B.2: Relevance of Paragraphs Discussing Multiple Topics Included in DMCA

| | | Topics included | | Include paragraph if it discusses | |
| | | DRM | | Only Whole | Specifics |
Year	Bill #	Circumvention	ISP Liability	Bill	
1995	S. 1284	Yes	Yes	Yes	Maybe
1995	H.R. 2441	Yes	Yes	Yes	Maybe
1997	S. 2037	Yes	Yes	Yes	Maybe
1997	H.R. 2281	Yes	No	Yes	Yes

If a paragraph expresses nonspecific/categorical support for or opposition to any of the four bills, include it as relevant. Even if an author merely muddles about the importance of getting the bill right, include it as relevant.

This rule means that most mentions of the WIPO treaties will be relevant. For instance, consider this excerpt:

> In the Information Age, the concept of copyright and intellectual property law is a keystone to developing electronic commerce. … Indeed, as technology changes and converges, the law must do so as well, and that is the foundation and core mission of the WIPO treaties.

The next paragraph discusses the bills specifically, but a discussion of the treaties similar to the above (whether similarly pro-WIPO and thus strong copyright or not) will generally be relevant.

Many of the relevant paragraphs will take no particular stand on the issue. Some will discuss a bill's progress through Congress. Others will make comments like the second paragraph in the following:

> The anti-circumvention provision would also enable copyright owners to dictate nonnegotiable terms in contracts for how others could use their information -- terms now governed by the Federal Copyright Act, not by contract law.

> "This raises some fundamental jurisdiction issues" between state law, which often governs contracts, and Federal law, said Mark Radcliffe, an Internet law specialist and a partner at the law firm of Gray, Cary, Ware & Freidenrich in Palo Alto, Calif.

The first paragraph has enough loaded language, especially "dictate nonnegotiable terms," that it is clearly part of the back-and-forth, he-said she-said coverage of the debate. The second paragraph is also relevant, however, as it simply raises (in a far more neutral way) the question of conflict between copyright and contract.

2. Discusses any proposal to ban or impede the circumvention of DRM

   Even if a paragraph does not mention one of the above bills by name, title, or author, include it as long as it discusses the same basic idea. This could be a ban on circumvention of DRM, a ban or limit on tools to circumvent DRM, etc.

   One particularly important proposal was the set of recommendations made by the Clinton Administration's Working Group on Intellectual Property Rights. One document highlights this group's role with the following excerpt:

   > We do not know who will be able to have actual knowledge of the existence of a copyrighted work on the Internet or control over its dissemination. And so we must cautiously wade into this pool, testing the waters by carefully changing our successful copyright system only as needed.

   > The bill before us today is the product of recommendations made by the Working Group on Intellectual Property Rights of the Administration's Information Infrastructure Task Force. The working group held a hearing in November 1993. They then drafted a so-called green paper and circulated it widely for comment, and subsequently held 4 additional days of hearings in Chicago, Washington, DC, and Los Angeles.

   > A final report was issued in September 1995, completing 2 ½ years of study and analysis of each of the major areas of intellectual property law. H.R. 2441 and its sister bill pending in the Senate, S. 1284, will represent the collective input of Congress, the administration and private industry to best protect copyrighted works on the Internet.

   The first is not quite relevant; even though it serves as a nice segue into the second paragraph, the speaker is not saying anything directly tied to the DMCA. The second paragraph directly discusses the green paper, and the third ties this proposal to the bills. The second paragraph best highlights this rule: proposals to restrict DRM circumvention are relevant.

3. Mentions the debate about whether to prevent the circumvention of DRM or the development, distribution, or sale of circumvention devices

4. Participates in or describes the debate over the pros and cons of circumventing DRM

5. Describes the bill's supporters or opponents with some sense that there is a *debate* afoot

Do not include a paragraph MERELY because it:

6. Discusses portions of the DMCA other than Title I, the anticircumvention provisions

   In particular, do not include a paragraph merely because it discusses these other titles:

- Title II of the DMCA sets up a notice-and-takedown procedure for copyright holders to contact internet service providers and demand the removal of copyrighted content from the internet; this limits the legal liability of online service providers.

- Title IV of the DMCA includes six miscellaneous provisions, including one that sets up a new system for webcasters to pay royalties on music broadcast over the internet.

- Title V provides copyright protection for the design of vessel hulls.

As these ideas began to meld together during the legislative process, hearings began mixing the debates around the various proposals; one important point of this code is to separate these debates.

7. Discusses the WIPO treaties or other international concerns

This is tricky; most paragraphs discussing WIPO will actually be relevant for one of the reasons described above. For instance, if a paragraph calls for the implementation of the portion of the WIPO treaties dealing with DRM circumvention, it is included. This does not mean that any discussion of the treaties is relevant, because the treaties also cover other topics. Much like the bills discussed in these hearings, one of the heated points of exchange is whether internet service providers (ISPs) should be liable for their customers' behavior.

Examples of topics that would not lead to a paragraph being marked as relevant include:

A.   Provisions stipulating that copyright shall apply online in the same way that it applies to physical media such as CDs and books

B.   Concerns about liability for online service providers

C.   Discussion of the treaty process or the resulting treaty *without* reference to implementing legislation in the US

8. Discusses other policy issues involving digital media and copyright or copyright in general

For example, consider the following paragraph:

> Lawmakers also deleted a controversial provision of the bill that would have granted broad new intellectual property rights to the owners of databases -- including such things as telephone books and sports statistics. In addition, Congress provided protection from prosecution for telephone and Internet companies that merely transmit materials pirated by others.

The first sentence discusses the debate over whether copyright should cover the raw data in databases, which is not relevant. Likewise, the second sentence describes a different part of the DMCA—which, as covered above, is not relevant.

Broadcast (Audio and/or Video) Flag

Include a paragraph if it Meets ANY of the Following:

1. Describes the proposed broadcast flag regulations or legislation

Any document that discusses the proposed regulations or legislation in meaningful detail will be included regardless of whether that document discusses the arguments for or against such proposals.

For instance, consider the following:

> The MPAA began its legislative push on Capitol Hill shortly after a May 6 ruling by the U.S. Court of Appeals for the District of Columbia.
>
> In that decision, the court reversed a Federal Communications Commission order that required makers of consumer-electronics devices capable of receiving broadcast digital TV signals to recognize a "broadcast flag" -- code that allows content owners to place limits on redistribution of digital content streams. The rule was to apply to devices manufactured on or after July 1, 2005.

The first paragraph describes the policy process surrounding the bill, and the second paragraph provides some more detail about what the proposal would do. Both are thus relevant.

2. Describes the potential social or economic impact of the broadcast flag and/or circumvention of the broadcast flag

The debate over the value of the flag as a technology is part of the debate over whether or not to require it.

Also, be sure to look for mentions of the Broadcast Protection Discussion Group (BPDG), the inter-industry group (similar to the DVD Copy Control Association) that developed the flag standard. Documents discussing the impact of the flag standard as developed by the BPDG are to be included under this rule.

3. Describes the potential copyright implications of the adoption of digital radio or TV broadcasting in an obvious allusion to the debate over whether or not to impose a broadcast flag

The most common example of this is the question of whether fans should continue to be able to record programming from radio or TV to be replayed later. For instance, any document with an extended statement to the effect that the adoption of digital and/or HD radio or television broadcasts forces us to reconsider the balance of copyright law—e.g., we cannot allow fans to record digital content that is of such pristine quality and so easily reproduced—should be included.

Likewise, any statement that takes the opposite stance, insisting that copyright should stay the same—e.g., consumers should retain the right to record digital radio in the same way they have recorded analog radio programming—is also to be included.

Consider this example:

> Those measures [broadcasters withholding content because of threats of online distribution of HD programs] are unlikely to be taken any time soon. First, advocates of the flag technology will try to circumvent the court's ruling through Congressional legislation. And despite the industry's alarm bells, Internet video piracy still is very much a nascent issue. At today's transmission speeds, it would take about 24 hours to send a one-hour show broadcast in HDTV over the Internet.

But broadband speeds will eventually increase, making it more practical to share digital broadcast files -- and thus more likely that it will be done illegally.

The first paragraph describes the case against worrying about copyright infringement based on digital broadcasting, while the second paragraph makes the case to consider worrying at a later date. These are both relevant.

4. Describes the bill's supporters or opponents with some sense that there is a *debate* afoot

Consider this example:

> Motion Picture Association of America Executive Vice President John Feehery on Wednesday confirmed that [Republican Representative Joe] Barton told the MPAA he doesn't support broadcast-flag provisions in his bill, but Feehery said the group hasn't determined its next course of action.
>
> "If that's what he thinks, that's what he thinks," said Feehery. "But we're continuing to educate members on the broadcast flag, and we're not sure where it will go."

The first paragraph states that the bill has supporters (MPAA) and opponents (Barton). The second is to be taken more as a reaffirmation of the MPAA's support for the flag—a more one-sided but still relevant addition to the article.

Do Not Include a Paragraph MERELY Because It:

5. Describes details that are not relevant to the debate about the social, economic, and political values at stake with the flag, its rollout, or home recording of digital broadcast

 Examples include:

A. The technical details of the technology

B. The legal subtleties of the law as it now stands

C. The FCC's (lack of) jurisdiction to impose such a mandate

D. Other details that do not otherwise shed light on the merits of the flag or its circumvention and/or the merits or costs of laws or policies that would require or limit its adoption.

None of these topics provide any insight into the costs or benefits of the flag or any flag-related policy changes.

6. Discusses other issues related to copyright in digital media

7. Discusses other issues related to digital broadcasting

Consider the following paragraph:

> The price of television sets will rise substantially in a couple of years because of an FCC ruling Thursday that mandates digital tuners be included in the sets.

425

The issue of digital tuners is a different policy related to the transition to digital TV.

DMCA Reform Bill

Include a Paragrph if it Meets ANY of the Following:

1. Discusses a proposed DMCA reform, including its practical, social, economic, or other implications.

   In 2003, Representative Zoe Lofgren (D-CA)introduced HR 1066, the BALANCE Act, which would have permitted circumvention of DRM for otherwise noninfringing purposes and would have permitted the design and marketing of tools for such circumvention.

   Also in 2003, Rep. Rick Boucher (D-VA) introduced a very similar bill, HR 107, the Digital Media Consumers' Rights Act of 2003. In 2005-06, it was HR 1201, which may be under-described as merely a bill that required labeling of copy-protected media such as compact discs. Any document mentioning these bills by name or number will generally be included, subject to the length requirement of 4 relevant sentences, but any mention of DMCA reform will be included, even without the corresponding bill number.

   For instance, consider the following excerpt:

   > In turn, the Business Software Alliance and another technology group, the Computer Systems Policy Project, said they would not support legislation that seeks to clarify and bolster the rights of people to use copyrighted material in the digital age, which the recording industry has opposed as unnecessary.

   > One such proposal, made by Representative Rick Boucher, Democrat of Virginia, and supported by several consumer groups and some technology companies, including Intel and Gateway, would seek to relax some of the restrictions imposed on users by the 1998 Digital Millennium Copyright Act. The statute makes it a crime for a consumer to make a backup copy or use a clip of video from a DVD, for instance.

   > "As a matter of first order we believe the marketplace should address these issues," said Ken Kay, executive director of the Computer Systems Policy Project, whose group represents Dell Computer, Intel, Hewlett-Packard and others.

   These paragraphs are all relevant. Read in light of the second paragraph, the end of the first paragraph clearly means the Boucher bill, a proposal "to clarify and bolster the rights of people to use copyrighted material in the digital age." The second paragraph names Boucher and alludes to his proposal. In the third, context indicates that Kay is expressing his opposition to the Boucher bill.

   Consider another example that highlights the combination of this rule with the guideline to read in context:

   > With talk of preemptive war all the rage on Capitol Hill, it seems that such posturing has extended into the world of digital copyright law.

426

> On Thursday Rep. Rick Boucher (D-Va.) and Rep. John Doolittle (D-Calif.) introduced the Digital Media Consumers Rights Act to preserve specific fair-use rights to copy digital works as well as "circumvention" rights to bypass copy protections. With no chance of passage this year, the bill's introduction prepares the ground for battle in the next session of Congress.

Read in light of the second paragraph, the first is clearly referring to the debate over Boucher's bill, describing its introduction as a preemptive political move.

2. Describes costs/benefits of DRM-specific portions of DMCA (17 USC §§ 1201-1204), especially the anti-circumvention provision of DMCA (§ 1201)

The debate over the DMCA is highly relevant; implicitly, it sets up the debate over reform proposals.
As part of this rule, include paragraphs that make (or describe) arguments about whether the Copyright Office's triennial rulemaking is properly determining exemptions to the law against circumventing DRM protections.

3. Participates in or describes the debate over the pros and cons of circumventing DRM.

The point of this rule is to include documents that are part of the debate over whether and when circumvention is sometimes valuable. Documents that meet this rule without also mentioning the law are very rare, but this rule may help a document with an otherwise brief discussion of the law *per se* to meet the overall length requirement.

In addition to the impact on fair uses generally, pay attention for debates around issues such as interoperability and research. Much of the debate about this portion of the DMCA revolves around the (in)ability of technologists to conduct research or to create products that integrate properly with systems that use DRM technologies.

4. Describes the bill's supporters or opponents with some sense that there is a *debate* afoot

One article states:

> "It's just time," said a beaming Gary Shapiro, president of the Consumer Electronics Association. "Consumers have been pushed up against the ropes. This is the first time in 20 years in which consumers are going on the offense rather than on the defense."

> Content owners, meanwhile, rolled their eyes.

Even though the 6-word second paragraph is informal and brief, it counts as relevant because it counts a group as being opposed to the bill; it is a small but nontrivial contribution to the argument against the bill.

The portrayal of sides may also be rather two-sided:

> Of course, Boucher is badly outgunned. The Republicans rule the House of Representatives, and while Boucher sits on the right subcommittee, he's only the third-most-senior Democrat. What's more, his Democratic counterparts in the Senate oppose him.

This is actually doubly relevant. It is implicitly talking about the bill—in terms of whether it will pass—and it is also counting up the supporters and opponents.

Do Not Include A Document MERELY Because It:

5.   Discusses other copyright-related issues

> For instance, other policy proposals are sometimes discussed. Consider the following:
>> The entertainment industry seeks to squash what it sees as rampant and illegal copying of digital content and, consequently, supports a bill introduced in July by Rep. Howard Berman (D-Calif.).
>> The Berman bill would give copyright owners the legal right to disrupt the unauthorized use of their copyrighted works on peer-to-peer (P2P) networks and exercise other content controls. (Berman's office declined to comment on Boucher's bill.)
>
> These paragraphs are not relevant. The Berman bill is a different policy debate, and the parenthetical mention of Boucher's bill is only by way of somebody declining to say anything about it.

Coding Paragraphs for Rhetorical Valence

Next, coders ask, "By itself, does this paragraph argue for stronger copyright, stronger fair use, or some neutral position?" As with documents, a code of neutral for paragraphs could mean the paragraph advances no position or that it attempts to present or juxtapose both positions.

IMPORTANT: Do not code a paragraph as neutral if it includes arguments from one camp only to refute or rebut them. As with identifying relevant paragraphs, consider the context before and after the paragraph to identify its purpose. Here is a hypothetical example:

> The motion picture industry claims that peer-to-peer trading is a threat to their ability to make money. They cite billions of dollars lost every year to this kind of noncommercial piracy, and they cannot imagine going on with things the way they are headed. They say they'll be out of business by 2015 at the current rate.

> I will concede that it's a threat to their *current* business model, but that's no more profound than saying that the car was a threat to the horse carriage industry. They need to adapt. People will pay for legitimate peer-to-peer downloads, and Congress has no obligation to prevent the music industry from going through these growing pains—especially at the expense of technological innovation. We need to keep the lawyers out of the laboratories.

In this example, *both* paragraphs would be coded as being strong fair use. As with relevance, keep the immediate context in mind in coding paragraphs for valence.

APPENDIX C: ON USING GOOGLE FOR ACADEMIC RESEARCH

As described in Chapter IV, Google provides less-than-satisfactory results compared with searches in bounded, organized databases such as LexisNexis. This should be unsurprising. An organized database has relatively consistent metadata, so documents can be sorted according to date, publication, author, and other criteria. Online documents are generally not organized in similar fashion *within* sites, so expecting database-like consistency in results when searching *across* sites is unrealistic. This is not to dismiss the value that Google offers to internet researchers. Quite the contrary, it is an easy and powerful means to learn a good deal about the web, but researchers are still grappling with exactly what they can expect to learn. The purpose of this Appendix (which is best read in conjunction with the section of Chapter 4 describing the process of testing recall and precision for Google searches) is to provide a more detailed explanation of its strengths and weaknesses. This, in turn, will suggest specific research strategies that can leverage its strengths and minimize the effect of its weaknesses.

I begin by offering a basic overview of the architecture of Google, including its PageRank algorithm and what is known about how it applies PageRank to sorting the results of searches for specific keywords. Second, I examine the strengths and weakness of using Google to identify relevant, high-quality websites. Third, I discuss some of the limitations I have encountered attempting to increase search recall as part of the search for documents on specific topics. Finally, I suggest some of the research strategies that can help leverage Google's strengths and limit its weaknesses.

PageRank and Search Results

Google's major innovation in search technology is the introduction of PageRank

(Langville & Meyer, 2006). Developed by Sergey Brin and Larry Page in the late 1990's,

PageRank "uses derived reputations from Web page interlinking to decide which search

results are most relevant" (Masum and Zhang, 2004). For instance, imagine an internet of

four websites, A, B, and C. A links to B, B links to A and C, and C links to A. See Figure

C.1:

Figure C.1: Counting Inlinks
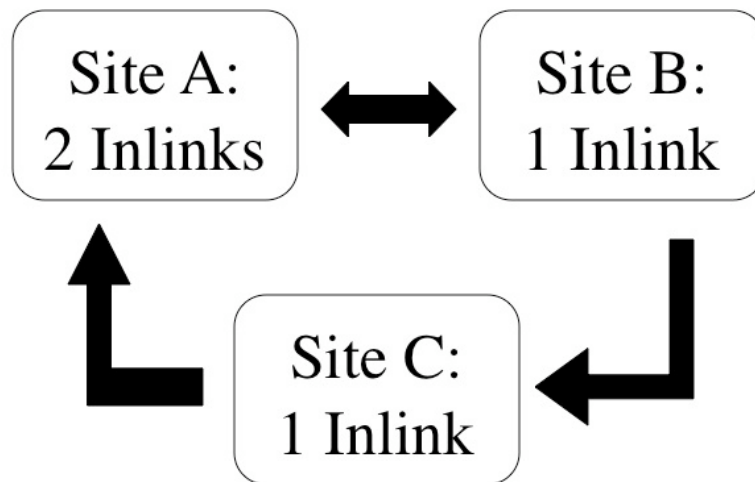
Site A is the only one of the three sites with incoming links from the other two sites; thus,

it has the highest PageRank. Further, links from more highly-ranked pages are worth

more. Thus, site B is ranked higher than site C, since its incoming link is from A, the highest-ranked site.

The idea behind this is that each link is an indicator of authority. In this sense, it uses the same fundamental logic as the IssueCrawler's web graphing software: websites with more inlinks carry more weight. PageRank is even more sophisticated, though, in determining the relative importance of each link. Langville (2006) elaborates:

A hyperlink from my homepage to your page is my endorsement of your page. Thus, a page with more recommendations (which are realized through inlinks) must be more important than a page with a few inlinks. However, similar to other recommendation systems such as bibliographic citations or letters of reference, the status of the recommender is important. … [Further], weights signifying the status of a recommender must be lowered for recommenders with little discrimination. In fact, the weight of each endorsement should be tempered by the total number of recommendations made by the recommender. (pp. 27-28)

The value of an inlink is determined by the linking page's importance (or PageRank) divided by the number of outlinks. This is illustrated in Figure C.2 (345Kai & Stannered, 2007):
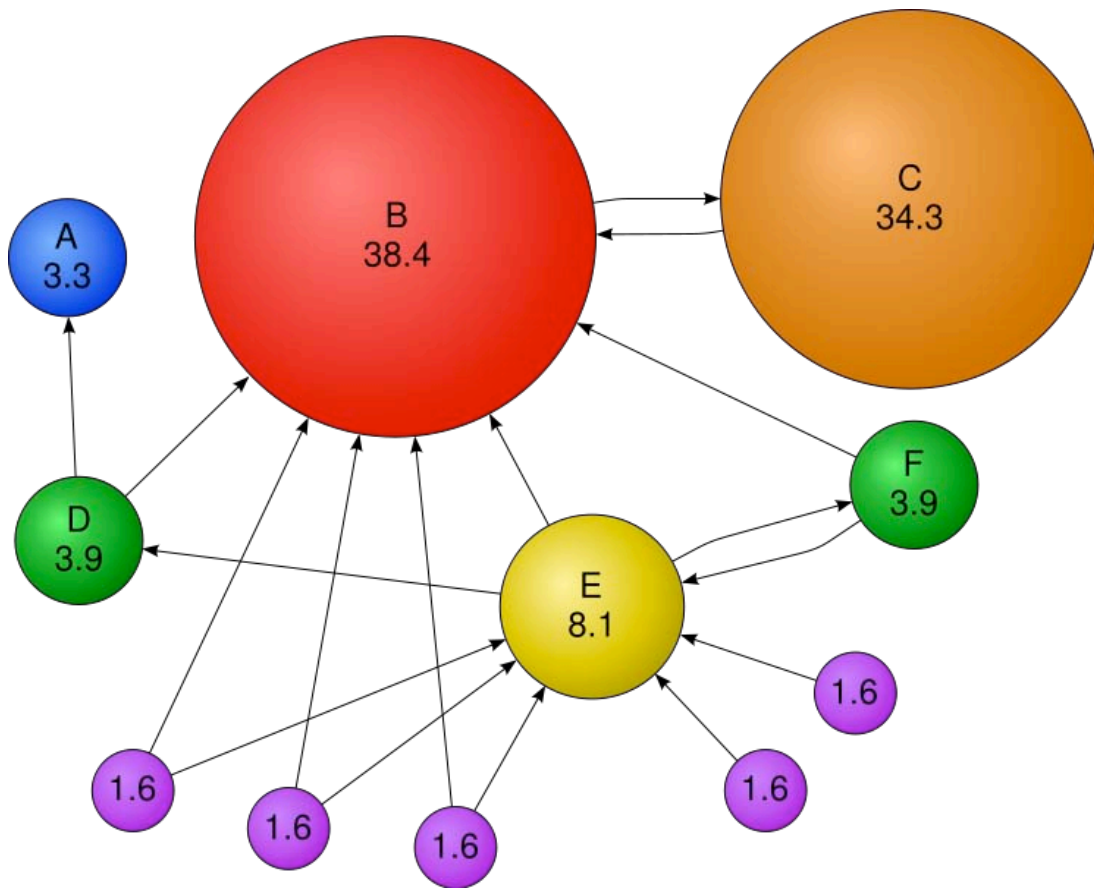
The total PageRank distributed among the 11 sites adds up to 100, with each page's

number reflecting the odds that a viewer will browse to that page. Page B is the highest-

ranked page with 38.4. Page C has the second-highest PageRank, 34.3, illustrating the

importance of each link's weight. Because Page C's only incoming link is also the only

outgoing link from the highest-ranked page—Page B—its PageRank is also quite high.

Page E has more incoming links, but each is from a low-visibility website, so its total

PageRank is only 8.1. In short, the PageRank algorithm assumes the chance that a viewer

will land on any given page is determined primarily by the number and weight of incoming links.[103]

In this way, Google calculates the PageRank value of every webpage in its index. While Google keeps the exact figures of each site's PageRank in strict confidence, it does provide a rough approximation via the Google Toolbar. This toolbar estimate is a number from 0 to 10, with 10 representing the highest PageRank and 0 representing the lowest. This summary of a site's actual PageRank score "is a transformed function (conjectured to be logarithmic or distributional)" (Griffiths & Christensen, 2005). Further, PageRanks "are known to be power-law distributed" (Griffiths & Christensen) and the 0-to-10 PageRank scores seem to reflect this distribution; very few sites are at the top of the index, and very many are at the bottom.

For researchers using Google PageRank in any systematic fashion, the Google Toolbar provides a frustrating graphic representation of PageRank; it is represented by the length of a small green bar, and users must hover the cursor over the green bar to get a numeric value. Firefox users can solve this by installing the Greasemonkey JavaScript manager (Boodman, 2008) and then installing the PageRank script (Shinya, 2007a); this changes one's Google search results pages to include a site's PageRank. A similarly useful Greasemonkey script, by the same programmer, is Result Numbers (Shinya, 2007b). This shows each page's rank for a given search, so researchers do not need to

---

[103] This graph also assumes an 85% "damping factor," or reduction in the weight of each link, reflecting an estimated 15% chance of landing on any arbitrary page in the network (*PageRank*, 2009). Thus, even the pages with no incoming links have a small but positive PageRank. The author is in no position to verify the specific numbers in this illustration, but it accurately conveys the idea of PageRank.

count by hand—especially handy if one changes one's Google preferences to return 100 results per page.

In determining the results for any given search query, Google combines the calculation of PageRank with estimations of a site's relevance to the specific query. Unlike the formula for PageRank, which is public knowledge, the means by which Google determines relevance to a given query is a closely guarded secret. The company describes it thus:

> Google combines PageRank with sophisticated text-matching techniques to find pages that are both important and relevant to your search. Google goes far beyond the number of times a term appears on a page and examines dozens of aspects of the page's content (and the content of the pages linking to it) to determine if it's a good match for your query. (Google, 2008d)

While the specifics are kept quiet, one particular aspect is commonly known: Google is particularly interested in the language used to describe a link to another webpage, often called the link label or anchor text. For instance, suppose I change my webpage to include the following:

> Are you looking for the best communication school ever? Come to Annenberg at Penn.

My site will display the phrase "best communication school ever" as a link that takes users to the Annenberg at Penn website. Google takes this as my recommendation that the website would be a good result for those searching for the phrase. In contrast, consider the following:

I am happy to be attending the best communication school ever, <a

href="http://www.asc.upenn.edu">Annenberg at Penn.

This would make a strong recommendation for the Annenberg site as a good result for

"Annenberg at Penn" and a weaker recommendation for the site as a good result for the

phrase "best communication school ever".[104] Google thus works in large part by counting

all of the incoming links for a given phrase or similar phrases, weighted by the linking

site's PageRank and number of outgoing links.

Google's Strengths and Weaknesses for Scholarly Research

Compared to the benchmark of the centrally planned databases, especially those

developed by librarians for library patrons, internet search engines such as Google are

both more and less powerful. Most obviously, the public internet is exponentially larger

than any database. Also, except for the development and implementation of

communication standards for the transmission of data, planning and development are

radically decentralized. The lack of gatekeepers and low barriers to entry ensure that the

entire industrialized world can speak online but make impossible the guarantee of

information quality and the provision of consistent metadata. Thus, compared to a

---

[104] Tatum (Tatum, 2005b) describes how this came to be exploited by "Google bombs" that willfully associated a phrase with an unlikely target website for political or comic effect. For instance, thousands of web pages linked the phrase "miserable failure" to the official White House biography page for President George W. Bush. The company has since taken steps to prevent mischievous use of this feature (D. Sullivan, 2007). Top results for "miserable failure" are now about the prank and Google's algorithmic countermeasures.

centrally planned database, any internet search engine provides access to far more information that is less consistently reliable and more difficult to sort and analyze.

Google has been extremely successful for a confluence of reasons, but the core of its success is its superior technology for identifying relevant, high-quality search results (Battelle, 2005). Through the techniques described above, Google leverages the collective intelligence of the entire internet to decide which pages are the best results for any given search (Masum & Zhang, 2004). Mental health researchers have even found a significant association between PageRank and other measures of website quality (Griffiths & Christensen, 2005). Searchers consistently return to Google because it performs well, and this means generally putting the best results toward the top of the list.

While these strengths are valuable, the tool also has shortcomings. Professionally designed databases permit end users to search for documents meeting certain parameters, including or eliminating each document based on criteria such as keywords, author, and date. Google's advanced search menu (Google, 2008a) provides similar features, but they do not measure up to the same standards of quality for at least two reasons. First, there is no standard style of metadata by which websites identify themselves. Thus, criteria such as date are unreliable; Google's estimate of a website's date may reflect recent minor changes to what is essentially an older document. Second, like the internet generally and thus like all search engines, Google has a bias toward the recent past (Hellsten et al., 2006).

Third, Google does not obey Boolean commands in a straightforward fashion. Closed databases reliably retrieve documents with certain search terms and omit those

without; if one searches LexisNexis for all *New York Times* articles containing the phrase "digital rights management," it will do exactly that. In contrast, Google allows the rest of the web to speak for a site. Thus, it sometimes includes documents that do not have a search phrase and excludes documents that do have the phrase. The very nature of the search business ensures that this technology has been developed to satisfy the roughly 95% of searchers who do not and will not use advanced search features (Battelle, p. 25). Thus, precisely tuned Boolean searches are not their core strength. Research conducted in efforts to develop a method to validate the recall and precision of search terms (Stryker et al., 2006) used in Google searches of specific websites help illustrate some of these problems.

Problems Using Google

The problems described above came to life as I attempted to refine the search terms used to retrieve documents related to the broadcast flag and DMCA reform debates. In particular, Google's failure to follow straightforward Boolean commands was unnerving. My search for documents relevant to the DMCA reform debate and the broadcast flag debate helps to illustrate this problem. An overview of this process was detailed in Chapter 4, but here is a more detailed description of the problems I encountered.

As detailed in Chapter 4, the Stryker method (Stryker et al., 2006) requires the estimation of a search term's recall and precision. Recall is the proportion of relevant documents retrieved; if a database contains 100 relevant documents, then a search retrieving 80 relevant documents has a recall of .80. Recall is calculated by comparing

the results of the targeted, "closed" search term with those of a much broader "open" search. The open search should be designed so that every potentially relevant document will be included.

Precision is the proportion of retrieved documents that are relevant. If the same search that retrieves 80 relevant documents also retrieves 40 irrelevant documents (120 total), that search has a precision of .67. Precision is calculated using only the results of the closed search term.

I began with an educated guess about what might constitute a good closed search term for identifying documents relevant to the DMCA reform debate:

copyright (dmca OR "digital millennium copyright act") (boucher OR encrypt!

OR 1201 OR hack! OR DRM OR "digital rights management") site:house.gov

This contains four parameters: the term copyright, the title of the DMCA, any of a set of words related to the debate, and the host domain House.gov. This retrieved 130 documents, and 39 of them were relevant. In order to provide a baseline estimate of the total number of relevant documents, from which I could calculate a recall rate of this search (Stryker et al., 2006), I also conducted to following open search:

copyright site:house.gov

This retrieved 793 documents. Google estimated that there are actually 8,140 relevant documents, but it will only return 1,000 or less documents for any given search (Google, 2008c). In this case, it retrieved only 793. Of these, 46 documents were relevant to the DMCA reform debate, and this included 38 of the 39 retrieved by the closed search term. In other words, the closed term's recall was 38 / 46, or .826. This exceeded the required

threshold of .80, but in an effort to improve this figure, I began manipulating the search

results. Essentially, I was trying to pick up several of the 8 documents missing from the

closed search. Several searches could retrieve 2 to 4 of them, but of the 8 documents, 4

were never retrieved by any closed search term. One document in particular illustrates

how sharply Google departs from simple Boolean retrieval of documents based on their

contents: a letter from Rep. John T. Doolittle (Doolittle, n.d.) to his House colleagues

introducing H.R. 107 (*Digital Media Consumers' Rights Act of 2003*, 2003).

The Doolittle letter does not contain the phrases "DMCA" or "Digital Millennium

Copyright Act." Thus, its absence from the closed search results suggests others on the

web were not linking to it using this phrase. It does contain three of the other search

terms: copyright, Boucher, and encryption.[105] This suggested that the requirement of

including a direct reference to the DMCA was excluding documents that referred to the

bills by number: H.R. 107 in the 108th Congress (referring to 17 U.S.C. § 107, which sets

out the conditions for fair use) and H.R. 1201 in the 109th (referring to 17 U.S.C. § 1201,

which contains the DMCA's anticircumvention provisions). It also excluded documents

that used the name of the bills' consistent lead author, Rep. Rick Boucher (D-VA), in

place of explicit references to the DMCA or the bills' numbers. Other documents

contained terms such as "digital locks" and "keys" more often than or in place of more

---

[105] Root extenders are unnecessary in Google. A search for "encrypt" will include results
for closely related words such as "encrypted" and "encryption" (see Google, 2008b). In
many closed databases, one would need to ask that these be included via a special
character such as an exclamation point. The Doolittle letter is retrieved (and the top
result) for this search:
     doolittle "hr 107" encrypt site:house.gov

technical terms such as encryption. All this suggested the need for a broader search.

Attempts to find an improved search term included the following modified search:

> copyright (("HR 107" OR "HR 1201") OR ((dmca OR "digital millennium
>
> copyright act" OR boucher) (encrypt OR 1201 OR hack OR DRM OR "digital
>
> rights management" OR lock OR key))) site:house.gov

Like the initial search, this contains four parameters: the term copyright, direct or indirect

reference to the bills under consideration, one or more of a larger set of terms, and the

host domain. This term is substantially more inclusive than the first attempt at a closed

search term.[106]

This search retrieved 3 of the 8 missing documents. Google automatically

excludes what it deems to be duplicate documents, but users can opt to see the full list;

doing so picked up an additional missing document, or 4 out of 8 total. Under a strict

Boolean retrieval of documents based on search terms, the search should retrieve the

Doolittle letter (n.d.). It meets all four terms: it contains the word "copyright," the bill

number *and* Boucher's name, the word "encryption," and the host domain House.gov. In

fact, the Doolittle letter was not retrieved by any of the modified searches, even though it

met the strict Boolean criteria for each and even though Google obviously thought it

important enough to include in the

---

[106] In a strictly logical sense, this set does not quite include all results from the initial search; by moving the term "Boucher" from the third parameter to the second, it would exclude any documents that mentioned the DMCA and Boucher but elided any of the terms from the new, larger set of terms from the debate such as "encrypt." Practically speaking, this set of documents is empty; none of the test documents from any of the search results mentioned both Boucher and the DMCA without also using one or more of the larger set of terms.

In the open search, "copyright site:house.gov," the letter was document number 263 out of over 8,000 results; thus, it had a high estimate of relevance and quality for the term "copyright." By forgetting the operator "site:house.gov" in another search, "copyright 'HR 107'," I accidentally searched the entire web, and the Doolittle letter ranked number 19 out of about 15,400. Yet the addition of terms further narrowing the field to a set of documents related to this very debate, on the letter's host domain, excluded the letter. It was not in any of the results for any of the closed searches, each of which yielded less than 500 documents; this document was not retrieved despite a total lack of a "1,001 problem" as described in Chapter 4. It was not retrieved despite meeting the exact Boolean specifications. This same basic problem occurred with three other documents; despite the inclusion of additional search terms to increase the set with an eye toward their inclusion, these documents remained out of each list of search results.

These refined searches proved unsatisfactory for an additional reason: each dropped as many relevant documents as it picked up. That is, while the new searches retrieved up to 4 of the original 8 missing documents, none could do so without losing at least as many relevant documents from the original closed search. As with the Doolittle letter and the other three seemingly unretrievable documents, these newly lost documents also met the Boolean specifications of the much broader searches. In other words, despite in effect asking Google for "all those documents plus a few more," Google simply returned a substantially overlapping set of documents that did not include all the relevant documents from the first search.

Two additional caveats merit mention here. First, results and document rankings are particularly fluid. The above searches were all conducted from July through November 2007, less than a year after the elections at the end of the 109th Congress. By January 2008, the total results for identical searches of House.gov produced less than half the number of results across the board. The initial attempt at a closed search term above retrieved 130 documents on July 25, 2007. On November 9, 2007, it retrieved 98. On January 25, 2008, it retrieved just 43. On October 20, 2007, the alternative closed search term described above retrieved 381 documents; by January 25, 2008, it retrieved just 120. Document rankings are even more volatile over time. Within a week, a day, or an hour, a document can go from nonexistent within a given search's results to within the top ten or vice versa. As long as an issue stays current, the number of relevant results for a given search tends to stay very similar over time, but this relative stability is only in aggregate.

As a second caveat, the order in which searchers enter words can affect results. While LexisNexis would interpret "copyright OR dmca" to be coextensive with "dmca OR copyright," Google does not. This is another component of their stepping away from a transparent Boolean retrieval mechanism. By doing so, Google has confounded any effort to extract completely predictable results. One cannot follow the ordinary means of expanding or contracting search results with an eye toward increasing recall (Stryker et al., 2006). This compounds the problems with unreliable metadata, the ever-shifting nature of the web, and the bias toward the recent past. In short, researchers must be cautious about the differences between Google results and those that would result from a Boolean search of a database.

Google Strategies for Researchers

While these problems are substantial, they are not insurmountable. Several

strategies are in order. First, researchers are particularly well advised to test recall and

precision in a formal rather than an informal manner. The tools for doing so are outlined

in Chapter 4. As the search experiences documented here illustrate, informal tests using

several similar closed search terms will almost certainly underestimate the number of

relevant documents not retrieved by such a method. Yet one need not discard this very

useful tool because of its relative inability to reach very high recall percentages. Through

leveraging its strength in putting relevant results near the top (estimated via logistic

regression) and, if necessary, combining multiple searches—again, see Chapter 4—one

can get a lot of mileage out of this free tool.

Second, researchers must develop methods for countervailing the internet's short

memory. This means different things for different research projects, but some general

outlines are in order. First and foremost, researchers must save the results of specific

searches; these snapshots will not be the same in a week. One can save the HTML code.

Also, if one has the software and the inclination, one can also create companion PDF

files.[107] This research project was greatly facilitated by saving both. Especially with the

Greasemonkey scripts described above, a PDF of the search results is incredibly useful if

one needs to know a document's search placement or PageRank.

---

[107] For this purpose, Mac and PC users both have no-cost options. Mac OS X now
includes PDF creation as part of the print menu for all programs. While I have not tried
them, there are also freeware programs for PC users that do the same thing.

For researchers looking at multiple websites or the results of searches for multiple search terms, combating the web's short memory also means conducting successive searches as close to each other as possible. This minimizes the problem of the short memory of the search engine. If one is dealing with a population of websites that tend to remain relatively stable over a period of months, one can then retrieve the documents—generally a more time-consuming process—with somewhat more leisure. Despite retrieving documents from the websites studied here over the months in the timeframe described above, almost all are still online despite Google's much more volatile search results. Even if a document goes offline in the interim, one can often recover its contents by using Google's cache or the Internet Archive (*Internet archive*, n.d.). If one is comparing Google results with less forgetful media archives, researchers would also do well to provide statistics on the mean and median dates of documents collected and control for date in all statistical analyses.

Finally, as in the exploratory stages of any research project, researchers do well to experiment wildly and take careful notes on the outcomes. Google is particularly disobedient, and it is very secretive about how it translates queries into results. Asking it very similar questions with many subtle variations can massage slightly better results out of it. If one goes in knowing the strengths and weaknesses, this experimentation can itself become part of learning about the online communication being studied.

REFERENCES

345Kai, & Stannered. (2007, November 25). *File:PageRanks-Example.svg*. Retrieved

February 5, 2009, from http://en.wikipedia.org/wiki/File:PageRanks-Example.svg

A bill to provide fair compensation to artists for use of their sound recordings, S.379,

111th Congress Cong. (2009).

*American Library Association et al. v. Federal Communications Commission and United

States of America*, 406 F.3d 689 (U.S. App. D.C. 2005).

Anderson, C. (2004, October). *The long tail*. Retrieved March 18, 2009, from

http://www.wired.com/wired/archive/12.10/tail.html

Anderson, N. (2008, October 3). *Key senators oppose DRM, ISP filtering in secret ACTA

treaty*. Retrieved February 10, 2009, from

http://arstechnica.com/old/content/2008/10/key-senators-oppose-drm-isp-

filtering-in-secret-acta-treaty.ars

Audio Broadcast Flag Licensing Act of 2006, H.R. 4861, 109th Cong. (2006).

Audio Home Recording Act, Pub. L. No. 102-563, 17 U.S.C. §§ 1001-1010 (1992).

Baker, C. E. (2002). *Media, markets, and democracy*. Cambridge ; New York:

Cambridge University Press.

Barabási, A.-L. (2003). *Linked: How everything is connected to everything else and what

it means for business, science, and everyday life*. New York: Plume.

Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business

and transformed our culture*. New York: Portfolio.
445

Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and instability in American politics*. Chicago: University of Chicago Press.

Benefit Authors without Limiting Advancement or Net Consumer Expectations (BALANCE) Act of 2003, H.R. 1066, 108th Cong. (2003).

Benkler, Y. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.

Bennett, W. L., & Entman, R. M. (2001). *Mediated politics: Communication in the future of democracy*. Cambridge: Cambridge University Press.

Bennett, W. L., & Manheim, J. B. (2001). The big spin: Strategic communication and the transformation of pluralist democracy. In W. L. Bennett & R. M. Entman (Eds.), *Mediated politics: Communication in the future of democracy* (pp. 279-298). Cambridge: Cambridge University Press.

Bimber, B. A. (2003). *Information and American democracy: Technology in the evolution of political power*. Cambridge, UK ; New York: Cambridge University Press.

Boodman, A. (2008). Greasemonkey (Version 0.7.20080121.0) [Computer software].

Boyer, P. J. (1988, September 18). Sony and CBS Records: What a romance! *The New York Times, sec. 6,* p. 34.

Bridgeport Music, Inc. v. Dimension Films, et al., 410 F.3d 792 (6th Cir. 2005).

Brown, A. E., & Stewart Jr., J. (1993). Competing advocacy coalitions, policy evolution, and airline deregulation. In P. A. Sabatier & H. C. Jenkins-Smith (Eds.), *Policy*

change and learning: An advocacy coalition approach (pp. 83-103). Boulder, Colo.: Westview Press.

Browne, W. P., & Paik, W. K. (1993). Beyond the domain: Recasting network politics in the Postreform Congress. *American Journal of Political Science, 37*(4), 1054-1078.

Business Software Alliance. (2008, December 17). *BSA Presents Cyber Champion Award to Congressman John Conyers*. Retrieved February 10, 2009, from http://www.bsa.org/country/News%20and%20Events/News%20Archives/en/2008/en-12172008-cyberchamps-conyers.aspx

Cahn v. Sony Corp., No. 90 Civ. 4537 (S.D.N.Y. filed July 9, 1990).

Cappella, J. N., & Jamieson, K. H. (1996). News frames, political cynicism, and media cynicism. *Annals of the American Academy of Political and Social Science, 546*, 71-84.

Center for Public Integrity. (2006). *Well connected: Tracking the broadcast, cable & telecommunications industry*. Retrieved May 25, 2006, from http://www.publicintegrity.org/telecom/

Center for Responsive Politics. (2009). *How we classify contributions*. Retrieved March 8, 2009, from http://www.opensecrets.org/industries/slist.php

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis *Statistics in Medicine, 19*, 3127-3131.

Cigler, A. J., & Loomis, B. A. (2007). Introduction: The changing nature of interest

    group politics. In A. J. Cigler & B. A. Loomis (Eds.), *Interest group politics* (7th

    ed., pp. 1-33). Washington, DC: CQ Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.).

    Hillsdale, NJ: L. Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Communications Opportunity, Promotion, and Enhancement Act of 2006, H.R. 5252,

    109th Congress Cong. (2006).

Communications, Consumer's Choice, and Broadband Deployment Act of 2006, S 2686,

    109th Congress Cong. (2006).

Consumer Broadband and Digital Television Promotion Act, S 2048, 107th Congress

    Cong. (2002).

Copyright Act of 1790, Pub. L. No., 1 Stat. 124 (1790).

Copyright Act of 1976, Pub. L. No. 94-553, 17 U.S.C. § U.S.C. 101 *et seq.* (1976).

Crutchfield, W. (1986, October 24). Next home stereo advance: Digital tape cassettes in

    1987. *The New York Times,* p. A1.

Davis, A. (2002). *Public relations democracy: Public relations, politics and the mass

    media in Britain*. New York: Manchester University Press.

Dean, J., Anderson, J. W., & Lovink, G. (2006). *Reformatting politics: Information

    technology and global civil society*. New York: Routledge.

Digital Audio Tape Recorder Act of 1987, Senate Committee on Commerce, Science, and

    Transportation, 100th Sess. (1987).

Digital Audio Tape Recorder Act of 1990, Senate Subcommittee on Communications of

the Committee on Commerce, Science, and Transportation, 101st Sess. (1990a).

Digital Audio Tape Recorder Act of 1990, S. 2358, 101st Cong. (1990b).

Digital audio tape recorders, House Subcommittee on Commerce, Consumer Protection,

and Competitiveness, 100th Sess. (1987).

Digital Future Coalition. (n.d.). *A description of the Digital Future Coalition*. Retrieved

March 14, 2008, from http://www.dfc.org/dfc1/Learning_Center/about.html

Digital Media Consumers' Rights Act of 2003, H.R. 107, 108th Cong. (2003).

Digital Media Consumers' Rights Act of 2005, H.R. 1201, 109th Cong. (2005).

Digital Millennium Copyright Act, HR 2281, 105th Congress Cong. (1997).

Digital Millennium Copyright Act, S 2037, 105th Congress Cong. (1998a).

Digital Millennium Copyright Act, Pub. L. No. 105-304, (1998b).

Digital Transition Content Security Act, HR 4569, 109th Congress Cong. (2005).

Doolittle, J. T. (n.d.). *Preserving personal fair use of digital media: Support H.R. 107*,

from http://www.house.gov/hensarling/rsc/doc/Doolittle--Fair%20Use.doc

Dudley, G., & Richardson, J. (1999). Competing advocacy coalitions and the process of

'frame reflection': A longitudinal analysis of EU steel policy. *Journal of European

Public Policy, 6*(2), 225-248.

Electronic Frontier Foundation. (2002, February 6). *Security researchers drop scientific

censorship case: Government, industry claim DMCA not a threat to science*.

Retrieved January 13, 2009, from

http://w2.eff.org/IP/DMCA/Felten_v_RIAA/20020206_eff_felten_pr.html

449

Electronic Frontier Foundation. (2008). *RIAA v. the people: Five years later*. Retrieved February 21, 2009, from http://www.eff.org/files/eff-riaa-whitepaper.pdf

Electronic Frontier Foundation. (2009, February 17). *File sharing*. Retrieved February 21, 2009, from http://www.eff.org/issues/file-sharing

Elias, S., & Goldoftas, L. (1996). *Patent, copyright & trademark: A desk reference to intellectual property law* (1st ed.). Berkeley: Nolo Press.

Fair Copyright in Research Works Act, H.R.801, 111th Congress Cong. (2009).

Farrall, K. N. (2005, October 1). *Web graph analysis in perspective: Description and evaluation in terms of Krippendorff's conceptual framework for content analysis (version 1.0)*. Retrieved March 5, 2008, from http://farrall.org/papers/webgraph_as_content.html

Farrall, K. N. (n.d.). *Web graph sociology research initiative: Main page*. Retrieved November 12, 2006, from http://farrall.org/webgraph/home.html

Farrall, K. N., & Delli Carpini, M. X. (2004). *Cyberspace, the web graph and political deliberation on the internet.* Paper presented at the International Conference on Politics and Information Systems: Technologies and Applications, Orlando, Florida.

Feist Publications, Inc., v. Rural Telephone Service Co., 499 U.S. 340 (1991).

Felten, E. (2003, August 5). *Revisionism*. Retrieved January 8, 2009, from http://www.freedom-to-tinker.com/blog/felten/revisionism

Felten, E. (2006). *Happy endings*. Retrieved January 8, 2009, from http://freedom-to-tinker.com/blog/felten/happy-endings

450

Fishman, S. (2003). *The copyright handbook: How to protect & use written works* (7th ed.). Berkeley, CA: Nolo.

Freedom and Innovation Revitalizing U.S. Entrepreneurship Act of 2007, H.R.1201, 110th Congress Cong. (2007).

Gais, T. L., Patterson, M. A., & Walker, J. L. (1984). Interest groups, iron triangles, and representative institutions in American national government. *British Journal of Political Science, 14*(2), 161-185.

Galperin, H. (2004). Beyond interests, ideas, and technology: An institutional approach to communication and information policy. *The Information Society, 20*, 159-168.

Gandy Jr., O. H. (1982). *Beyond agenda setting: Information subsidies and public policy*. Norwood, N.J.: Ablex Pub. Co.

Gasaway, L. (2003). *When U.S. works pass into the public domain*. Retrieved October 10, 2006, from http://www.unc.edu/~unclng/public-d.htm

Gibson, R. K., Lusoli, W., & Ward, S. (2005). Online participation in the UK: Testing a 'contextualised' model of internet effects. *British Journal of Politics and International Relations, 7*, 561-583.

Gillespie, T. (2007). *Wired shut: Copyright and the shape of digital culture*. Cambridge, MA: MIT Press.

Gillmor, D. (2004). *We the media*. Sebastopol, CA: O'Reilly Media.

Givel, M. S., & Glantz, S. A. (2000). Failure to defend a successful state tobacco control program: Policy lessons from Florida. *American Journal of Public Health, 90*, 762-767.

Google. (2008a). *Advanced search*. Retrieved January 23, 2008, from

   http://www.google.com/advanced_search?hl=en

Google. (2008b). *Does Google support wildcard searches?* Retrieved January 25, 2008,

   from http://www.google.com/support/bin/answer.py?answer=3178&topic=352

Google. (2008c). *How can I see more than the first 1000 results?* Retrieved January 25,

   2008, from

   http://www.google.com/support/bin/answer.py?answer=484&topic=352

Google. (2008d). *Our search: Google technology*. Retrieved January 22, 2008, from

   http://www.google.com/technology/index.html

Govcom.org Foundation. (n.d.). *Issue Crawler*. Retrieved March 4, 2009, from

   http://www.issuecrawler.net/

Graber, D. (2005). Political communication faces the 21st Century. *Journal of

   Communication, 55*(3), 479-507.

Griffiths, K. M., & Christensen, H. (2005, November 15). *Website quality indicators for

   consumers*. Retrieved January 22, 2008, from http://www.jmir.org/2005/5/e55/

Hargittai, E., & Hinnant, A. (2008). Digital inequality: Differences in young adults' use

   of the internet. *Communication Research, 35*(5), 602-621.

Hellsten, I., Leydesdorff, L., & Wouters, P. (2006). Multiple presents: How search

   engines rewrite the past. *New Media & Society, 8*(6), 901-924.

Herman, B. D. (2002). *Theory jockey: A critical discourse analysis of the DJ in DJ

   culture.* Unpublished Thesis (MA), Colorado State University.

Herman, B. D. (2005). *ShoutingLoudly*. Retrieved March 18, 2009, from

    http://www.shoutingloudly.blogspot.com/

Herman, B. D. (2006a, November 22). *Copyright Office grants 6 exemptions for*

    circumventing TPMs. Retrieved January 7, 2009, from

    http://www.shoutingloudly.com/2006/11/22/copyright-office-grants-6-

    exemptions-for-circumventing-tpms

Herman, B. D. (2006b). Opening bottlenecks: On behalf of mandated network neutrality.

    Federal Communications Law Journal, 59, 107-160.

Herman, B. D. (2006c). Scratching out authorship: Representations of the electronic

    music DJ at the turn of the 21st century. *Popular Communication, 4*(1), 21-37.

Herman, B. D. (2008). Breaking and entering my own computer: The contest of

    copyright metaphors. *Communication Law and Policy, 13*(2), 231-274.

Herman, B. D., & Gandy Jr., O. H. (2006). Catch 1201: A legislative history and content

    analysis of the DMCA exemption proceedings. *Cardozo Arts & Entertainment*

    Law Journal, 24, 121-190.

Herman, B. D., Karpf, D., Tocci, J., & Tsui, L. (2006). *ShoutingLoudly*. Retrieved March

    18, 2009, from http://www.shoutingloudly.com/

Herman, E. S., & Chomsky, N. (2002). *Manufacturing consent: The political economy of*

    the mass media. New York: Pantheon Books.

Herring, S. C. (2004). *Content analysis for new media: Rethinking the paradigm*.

    Retrieved November 23, 2006, from

    http://ella.slis.indiana.edu/%7Eherring/newmedia.pdf

Hewitt, H. (2005). *Blog: Understanding the information reformation that's changing your world*. Nashville, TN: Nelson Books.

Higher Education Opportunity Act, Pub. L. No. 110-315, (2008).

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003, April). *"Googlearchy": How a few heavily-linked sites dominate politics on the web.* Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.

Hindman, M. S. (2009). *The myth of digital democracy*. Princeton: Princeton University Press.

Horrigan, J. B. (2008, July). *Home broadband adoption 2008: Adoption stalls for low-income Americans even as many broadband users opt for premium services that give them more speed* Retrieved March 2, 2009, from

http://www.pewinternet.org/pdfs/PIP_Broadband_2008.pdf

Hosein, I. (2004). The sources of laws: Policy dynamics in a digital and terrorized world. *The Information Society, 20*, 187-199.

Information Infrastructure Task Force: Working Group on Intellectual Property Rights, & Lehman, B. A. (1994). *Intellectual property and the National Information Infrastructure: A preliminary draft of the report of the Working Group on Intellectual Property Rights*. Washington, DC: Information Infrastructure Task Force.

Information Infrastructure Task Force: Working Group on Intellectual Property Rights, & Lehman, B. A. (1995). *Intellectual property and the National Information*

Infrastructure: The report of the Working Group on Intellectual Property Rights.
Washington, DC: Information Infrastructure Task Force.

Internet archive. (n.d.). Retrieved January 25, 2008, from
http://www.archive.org/index.php

Iron triangle. (2006, October 9). Retrieved October 15, 2006, from
http://en.wikipedia.org/wiki/Iron_triangle

Iyengar, S., & Simon, A. F. (2000). New perspectives and evidence on political
communication and campaign effects. *Annual Review of Psychology, 51*, 149-169.

Jackson, M. (1995). Commerce versus art: The transformation of fair use. *Journal of
Broadcasting and Electronic Media, 39*, 190-199.

Jamieson, K. H. (2000). *Everything you think you know about politics--and why you're
wrong* (1st ed.). New York: Basic Books.

John, P. (2001). Policy networks. In K. Nash & A. Scott (Eds.), *The Blackwell
companion to political sociology* (pp. xiv, 478). Malden, MA: Blackwell
Publishers.

Jones, B. D., & Baumgartner, F. R. (2005). *The politics of attention: How government
prioritizes problems*. Chicago: University of Chicago Press.

Kaplan, R. L. (2006). The news about new institutionalism: Journalism's ethic of
objectivity and its political origins. *Political Communication, 23*, 173-185.

Kelty, C. M. (2008). *Two bits: The cultural significance of free software*. Durham: Duke
University Press.

Kerbel, M. R., Apee, S., & Ross, M. H. (2000). PBS ain't so different: Public

    broadcasting, election frames, and democratic empowerment. *The Harvard*

    International Journal of Press/Politics, 5(4), 8-32.

Kingdon, J. W. (2003). *Agendas, alternatives, and public policies* (2nd ed.). New York:

    Longman.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.).

    Thousand Oaks, CA: Sage.

Lakoff, G. (2004). *Don't think of an elephant!: Know your values and frame the debate:*

    The essential guide for progressives. White River Junction, VT: Chelsea Green

    Pub. Co.

Landes, W. M., & Posner, R. A. (2004). *The political economy of intellectual property*

    law. Washington, DC: AEI Press.

Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond: The science of*

    search engine rankings. Princeton, NJ: Princeton University Press.

Lee, H. (2007). The audio broadcast flag system: Can it be a solution? *Communication*

    Law and Policy, 12, 405-476.

Lenz, G. S. (n.d.). *Learning and opinion change, not priming: Reconsidering the*

    evidence for the priming hyptothesis. Retrieved September 23, 2006, from

    http://www.princeton.edu/~glenz/Lenz_learning_not_priming.pdf

Lessig, L. (2001). *The future of ideas: The fate of the commons in a connected world* (1st

    ed.). New York: Random House.

Lessig, L. (2004). *Free culture: How big media uses technology and the law to lock down culture and control creativity*. New York: Penguin Press.

Lessig, L. (2007). *Required reading: The next 10 years*. Retrieved February 27, 2009, from

http://www.lessig.org/blog/2007/06/required_reading_the_next_10_y_1.html

Lessig, L. (2009). *NYPL REMIX event with Shepard Fairey and Steven Johnson*. Retrieved March 14, 2009, from http://blip.tv/file/1821209

Lessig, L., Fairey, S., & Johnson, S. (2009, February 26). *Remix: Making art and commerce thrive in the hybrid economy.* Paper presented at the New York Public Library, New York.

Litman, J. (2000). *Digital Copyright*. Amherst, NY: Prometheus Books.

Love, J. (2009, March 12). *Obama Administration rules texts of new IPR agreement are state secrets*. Retrieved March 19, 2009, from

http://www.huffingtonpost.com/james-love/obama-administration-rule_b_174450.html

Lubken, D. (2008). Remembering the straw man: The travels and adventures of Hypodermic. In D. W. Park & J. Pooley (Eds.), *The history of media and communication research: Contested memories* (pp. 19-42). New York: Peter Lang.

Macworld. (2004, March 15). *Indie record-labels' download joy*. Retrieved April 4, 2004, from http://www.macworld.co.uk/news/main_news.cfm?NewsID=8164

Marres, N. (2006). Net-work is format work: Issue networks and the sites of civil society politics. In J. Dean, J. W. Anderson & G. Lovink (Eds.), *Reformatting politics: Information technology and global civil society* (pp. 3-17). New York: Routledge.

Masum, H., & Zhang, Y. C. (2004). Manifesto for the reputation society. *First Monday, 9*(7).

McChesney, R. W. (1999). *Rich media, poor democracy: Communication politics in dubious times*. Urbana, IL: University of Illinois Press.

McChesney, R. W. (2004a). Making a molehill out of a mountain: The sad state of political economy in US media studies. In A. Calabrese & C. Sparks (Eds.), *Toward a political economy of culture: Capitalism and communication in the twenty-first century*. Lanham, MD: Rowman & Littlefield.

McChesney, R. W. (2004b). *The problem of the media: U.S. communication politics in the twenty-first century*. New York: Monthly Review Press.

McCombs, M. E. (2004). *Setting the agenda: The mass media and public opinion*. Cambridge, UK: Polity.

McLeod, K. (2001). *Owning culture: Authorship, ownership, and intellectual property law*. New York: P. Lang.

McLeod, K. (2005). *Freedom of expression®: Overzealous copyright bozos and other enemies of creativity* (1st ed.). New York: Doubleday.

Menell, P. S., & Nimmer, D. (2007). Legal realism in action: Indirect copyright liability's continuing tort framework and Sony's de facto demise. *UCLA Law Review, 55*, 1-62.

458

MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

Morelli, M. (2009, February 11). *Major labels struggle to survive*. Retrieved March 17,

2009, from http://beatcrave.com/2009-02-11/major-labels-struggle-to-survive/

Morgan, G. A., Leech, N. L., Gloeckner, G. W., & Barrett, K. C. (2007). *SPSS for*

introductory statistics: Use and interpretation (3rd ed.). Mahwah, NJ: Lawrence

Erlbaum.

Mossberger, K., Tolbert, C. J., & Gilbert, M. (2006). Race, place, and information

technology. *Urban Affairs Review, 41*(5), 583-620.

Napoli, P. M. (in press). Public interest media advocacy and activism as a social

movement. *Communication Yearbook, 33*.

Negativland. (1995). *Fair use: The story of the letter U and the numeral 2*. Concord, CA:

Seeland.

Netanel, N. (2008). *Copyright's paradox*. Oxford: Oxford University Press.

NII Copyright Protection Act of 1995, HR 2441, 104th Cong. (1995a).

NII Copyright Protection Act of 1995, S 1284, 104th Cong. (1995b).

Nimmer, D. (2000). A Riff on Fair Use in the Digital Millennium Copyright Act.

Univeristy of Pennsylvania Law Review, 148, 673-742.

PageRank. (2009, January 25). Retrieved February 5, 2009, from

http://en.wikipedia.org/w/index.php?title=PageRank&oldid=266365813

Pan, Z., & Kosicki, G. M. (2001). Framing as a strategic action in public deliberation. In

S. D. Reese, O. H. Gandy & A. E. Grant (Eds.), *Framing public life: Perspectives*

on media and our understanding of the social (pp. 35-65). Mahwah, NJ: Lawrence Erlbaum Associates.

Patry, W. (2006). *Grokster remand opinion*. Retrieved November 7, 2006, from http://williampatry.blogspot.com/2006/09/grokster-remand-opinion.html

Performance Rights Act, H.R.848, 111th Congress Cong. (2009).

Peters, M. (2006, November 17). *Letter from Marybeth Peters, Register of Copyrights, to James H. Billington, Librarian of Congress*. Retrieved January 7, 2009, from http://www.copyright.gov/1201/docs/1201_recommendation.pdf

Pickard, V. W. (2006). Assessing the radical democracy of Indymedia: Discursive, technical, and institutional constructions. *Critical Studies in Media Communication, 23*(1), 19-38.

Pollack, A. (1988, January 16). Move to end digital tape dispute. *The New York Times,* p. 35.

Pollack, A. (1990, July 11). Suit seeks to bar sale of new audio players. *The New York Times,* p. D5.

Rainie, L., & Horrigan, J. B. (2005). *A decade of adoption: How the internet has woven itself into American life*. Retrieved March 2, 2009, from http://www.pewinternet.org/pdfs/Internet_Status_2005.pdf

Reese, R. A. (2006). The temporal dynamics of "capable of substantial noninfringing uses". *Michigan Telecommunications and Technology Law Review, 13*, 197-224.

Reese, S. D. (2001). Prologue--Framing public life: A bridging model for media research. In S. D. Reese, O. H. Gandy & A. E. Grant (Eds.), *Framing public life:*

Perspectives on media and our understanding of the social (pp. 7-31). Mahwah, NJ: Lawrence Erlbaum Associates.

Reilly, D. W. (2006, October 2). *The telecom slayers*. Retrieved October 8, 2006, from http://www.salon.com/tech/feature/2006/10/02/slayers/print.html

Richtel, M. (2002, December 17). Russian company cleared of illegal software sales. *The New York Times,* p. C4.

Rodino-Colocino, M. (2006). Laboring under the digital divide. *New Media & Society, 8*(3), 487-511.

Rogers, R. (2004). *Information politics on the Web*. Cambridge, MA: MIT Press.

Rosenbloom, D. H. (1983). Public administrative theory and the separation of powers. *Public Administration Review, 43*(3), 219-227.

Sabatier, P. A., & Jenkins-Smith, H. C. (1993). The advocacy coalition framework: Assessment, revisions, and implications for scholars and practitioners. In P. A. Sabatier & H. C. Jenkins-Smith (Eds.), *Policy change and learning: An advocacy coalition approach* (pp. 211-235). Boulder, Colo.: Westview Press.

Sabatier, P. A., & Jenkins-Smith, H. C. (1999). The advocacy coalition framework: An assessment. In P. A. Sabatier (Ed.), *Theories of the policy process* (pp. 117-166). Boulder, CO: Westview Press.

Samuelson, P. (1999). Intellectual property and the digital economy: Why the anti-circumvention regulations need to be revised. *Berkeley Technology Law Journal, 14*, 1-49.

Samuelson, P. (2004). Intellectual property and the digital economy: Why the anti-circumvention regulations need to be revised. *University of Ottawa Law & Technology Journal, 1*, 1-21.

Schattschneider, E. E. (1960). *The semisovereign people: A realist's view of democracy in America* ([1st] ed.). New York: Holt Rinehart and Winston.

Shade, L. R. (2003). Digital divide. In S. Jones (Ed.), *Encyclopedia of new media: An essential reference to communication and technology* (pp. 138-139). Thousand Oaks, CA: Sage Publications.

Shapiro, E. (1991, July 12). Accord on digital taping now faces Congress debate. *The New York Times,* p. D1.

Shinya. (2007a). Google Pagerank [Computer software].

Shinya. (2007b). Result Numbers [Computer software].

Sigal, L. V. (1973). *Reporters and officials: The organization and politics of newsmaking*. Lexington, MA: D. C. Heath.

Somoano, M. L. (2006). *Bridgeport Music, Inc., v. Dimension Films*: Has unlicensed digital sampling of copyrighted sound recordings come to an end? *Berkeley Technology Law Journal, 21*, 289-309.

Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417 (1984).

Stemmler, M. (2009). *Lawrence Lessig: Live from the NYPL, Remix: Post event wrap-up*. Retrieved March 14, 2009, from http://www.nypl.org/blogs/2009/03/03/live-nypl-remix-post-event-wrap

Stets, D. (1996, April 11). Pump up the PC. *The Philadelphia Inquirer,* p. F01.

Stryker, J. E., Wray, R. J., Hornik, R. C., & Yanovitzky, I. (2006). Validation of database search terms for content analysis: The case of cancer news coverage. *Journalism and Mass Communication Quarterly, 83*(2), 413-430.

Sullivan, D. (2007, January 25). *Google kills Bush's miserable failure search and other Google bombs*. Retrieved January 22, 2008, from http://searchengineland.com/070125-230048.php

Sullivan, J. (1988, October 28). Rock the boat, Billy Bragg. *Boston Globe,* p. 51.

Tatum, C. (2005, October). *Deconstructing Google bombs: A breach of symbolic power or just a goofy prank?* Retrieved January 22, 2008, from http://www.firstmonday.org/issues/issue10_10/tatum/

Trippi, J. (2004). *The revolution will not be televised: Democracy, the Internet, and the overthrow of everything*. New York: Regan Books.

Vaidhyanathan, S. (2001). *Copyrights and copywrongs: The rise of intellectual property and how it threatens creativity*. New York: New York University Press.

Vaidhyanathan, S. (2004). The state of copyright activism. *First Monday, 9*(4).

Wallsten, K. (2007). Agenda setting and the blogosphere: An analysis of the relationship between mainstream media and political blogs. *Review of Policy Research, 24*(6), 567-587.

West, D. M., & Loomis, B. A. (1999). *The sound of money: How political interests get what they want* (1st ed.). New York: W.W. Norton.

WIPO Copyright and Performances and Phonograms Treaties Implementation Act of 1998, Pub. L. No. 105-304, 101-105 (1998).

463

WIPO Copyright Treaty, adopted by the Diplomatic Conference on Dec. 20, 1996, WIPO

   Doc. CRNR/DC/94 (1996).

Woodmansee, M., & Jaszi, P. (1994). *The construction of authorship: Textual*

   appropriation in law and literature. Durham ; London: Duke University Press.

Zaller, J. (1996). The myth of massive media impact revived: New support for a

   discredited idea. In D. C. Mutz, P. M. Sniderman & R. A. Brody (Eds.), *Political*

   persuasion and attitude change (pp. 17-78). Ann Arbor: University of Michigan

   Press.