Copyright in an Era of Information Overload: Toward the Privileging of Categorizers

Frank Pasquale

Environmental laws are designed to reduce negative externalities (such as pollution) that harm the natural environment. Copyright law should adjust the rights of content creators in order to compensate for the ways they reduce the usefulness of the information environment as a whole. Every new work created contributes to the store of expression, but also makes it more difficult to find whatever work one wants. "Search costs" have been well-documented in information economics. Copyright law should take information overload externalities like search costs into account in its treatment of alleged copyright infringers whose work merely attempts to index, organize, categorize, or review works by providing small samples of them. They are not "free riding" off the labor of copyrightholders, but rather are creating the types of navigational tools and filters that help consumers make sense of the ocean of expression copyrightholders have created.

The new scholarship of cultural environmentalism sets the groundwork for a better understanding of the *costs,* as well as the *benefits,* of copyrighted expression. Any bit of expression that signals something to one who wants exposure to it may constitute noise to thousands of others. By modeling information overload as an externality imposed by copyrighted works, this article attempts to provide a new economic justification for more favorable legal treatment of categorizers, indexers, and reviewers. Information overload is an unintended negative consequence of copyright law's success in incentivizing the production and distribution of expression. If courts grant content owners the right to veto categorizers' efforts to make sense of given fields of expression, they will only exacerbate the problem. Designed to promote the "progress of the arts and sciences," copyright doctrine should privilege the efforts of those who make that progress accessible and understandable. Categorizers fill both those vital roles.

Copyright in an Era of Information Overload: Toward the Privileging of Categorizers

Frank Pasquale

Copyright in an Era of Information Overload: Toward the Privileging of Categorizers

*Frank Pasquale**

I. INTRODUCTION

What to read? or watch? or listen to? These are hard questions, not because of any scarcity of expression, but rather because of its abundance. Over 100,000 books are published in the United States each year, thousands of movies and CD's are released, and the amount of textual, musical, and visual works on the internet continues to rise exponentially. Whose work can we trust? And who knows what of it will rank among the best that has been thought and said—or even provide a few moments levity?[1]

---

*Associate Professor, Seton Hall Law School. Many thanks to my hosts at the St. John's University Distinguished Speakers Colloquium and the Seton Hall Law School Faculty Retreat for giving me an opportunity to present on this topic. Erik Lillquist, Gaia Bernstein, Thomas Healy, Marina Lao, Nelson Tebbe, Eric Goldman, Brett Frischmann, James Grimmelmann, and Charles Sullivan provided valuable comments on the paper. I also wish to thank participants at the May Gathering on Methodology in Legal Scholarship (at the University of Virginia) for their comments. The Intellectual Property Scholars Conference of 2006 (at Berkeley) also provided an excellent venue for discussion. Thanks also to Mohammed Azeez, Dean Murray, Scott Sholder, and Matthew Tuttle for excellent research assistance.

[1] As Matthew Arnold put it, the purpose of culture is to preserve "the best that is known and thought in the world." Matthew Arnold, *The Function of Criticism at the Present Time*, in THE NORTON ANTHOLOGY OF ENGLISH LITERATURE 2147, 2156 (M. H. Abrams ed., 5th ed. 1987) (1865). Copyright law is one of the most important legal tools for regulating culture in the United States. *See* Guy Pessach, *Copyright Law as Silencing Restriction on Noninfringing Materials: Unveiling the Scope of Copyright's Diversity Externalities*,

76 S. CAL. L. REV. 1067, 1079 (2003) (discussing how "copyright regimes deal with cultural and political resources").

Admittedly, a bulging bookshelf or surfeit of films prompts an existential crisis in only the most sensitive souls. Most of us, most of the time, drift along a well-trod path of filters and recommenders. The *New York Review of Books* may be a trusted guide to "must-reads" (or "must-avoids"). A favored movie or music critic might act as Beatrice (or Virgil) in our daunting quest for information, entertainment, or a fresh perspective on current events.[2] As Richard Caves observed in his classic analysis of the "creative industries," "buffs, buzz, and educated tastes" are indispensable tools for making sense of the world of media around us.[3]

Such tastemakers have become all the more important, and varied, as content offerings proliferate.[4] They provide the *metadata* (i.e., data about data) essential to finding the expression one wants. A website like "Rotten Tomatoes" can quickly aggregate reviews of a movie and present them concisely. Amazon invites anyone to review the books it sells. The iTunes music store posts customer reviews of the podcasts it offers. Search engines complement all these efforts by quickly assembling digital information regarding a query.[5] Such categorizers are on the verge of becoming even more

---

[2] In the *Divine Comedy*, Beatrice, one of the blessed, sends Virgil to guide Dante through Hell and Purgatory. Beatrice herself guides him in Heaven. DANTE ALIGHIERI, I THE DIVINE COMEDY: INFERNO 37 (Canto II, lines 55-70) (trans. John Sinclair, 1961) (As Virgil explains to Dante, "a lady [Beatrice] called me, so blessed and so fair that I begged her to command me.").

[3] RICHARD CAVES, CREATIVE INDUSTRIES: CONTRACTS BETWEEN ART AND COMMERCE 175 (2000) (describing the gatekeeping role of various entities in recommending (and discounting) works).

[4] For accounts of the accelerating pace of digitization of data, *see* PHILLIP EVANS AND THOMAS WURSTER, BLOWN TO BITS: HOW THE NEW ECONOMICS OF INFORMATION TRANSFORMS STRATEGY (1999); NICHOLAS NEGROPONTE, BEING DIGITAL (1995); DON TAPSCOTT, THE DIGITAL ECONOMY: PROMISE AND PERIL IN THE AGE OF NETWORKED INTELLIGENCE (1996).

[5] For a more complete list of categorizers, see discussion of the "value of categorizers" in Part IV below, beginning at page 61.

effective guides to online content; for example, as Google aims to index books, and new technologies of sampling provide ever more sophisticated ways for online reviewers to illustrate their posts and podcasts. The rise of these metadata providers suggests that the problem of information overload is beginning to solve itself. As more and more services rate and organize content, there is less reason to think one has missed some particularly compelling, delightful, or important work.

Unfortunately, copyright litigation has begun to stifle this development. Content owners are beginning to demand license fees not merely for works themselves, but also for any fragments of them. The Motion Picture Association of America has already shut down a site that illustrated the information it provided about movies with trailers.[6] Major publishers have sued Google, insisting that the search engine license any "snippets" from books that it deems relevant to a search query.[7] A small search engine had to fight a long legal battle merely to defend its practice of putting tiny, "thumbnail" reproductions of an artist's landscapes in its database.[8] Claiming absolute rights over the content they own, many copyrightholders appear to demand nothing less than perfect control over any fragment or sample of their works.

Many copyright theorists have documented how such fine-grained control would harm society,[9] and perhaps even

---

[6] *See* discussion of Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 342 F.3d 191 (3rd Cir. 2003), *infra,* text accompanying notes 215 to 223.

[7] *See* discussion of Google Print and Google Library, *infra,* text accompanying notes 39 to 90.

[8] *See* discussion of Kelly v. Arriba Soft, 336 F.3d 811 (9th Cir. 2003), *infra,* text accompanying notes 208 to 212.

[9] Larry Lessig and Yochai Benkler have both documented the flood of creativity that new computing technologies have helped unleash, and proposed legal rules to protect such innovation. LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999); Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 414-29 (1999). Lastowka and Hunter's work has illustrated the extraordinary importance of amateur production of content. Dan Hunter and F. Gregory Lastowka, *Amateur-to-Amateur*,

copyrightholders themselves.[10] These theorists have closely tied their encouragement of new creativity and "remix culture" to proposals for copyright reform. In order to make the "raw material" of innovation more available to improvers" seeking to work with extant innovation,[11] copyright reformers aim to reduce the scope, strength, and duration of exclusive property rights in information. They have offered a number of compelling justifications for their position, focusing on promotion of innovation, diversification of content providers, equality of access, and virtue-creating effects of production (as opposed to mere consumption) of content.[12]

---

46 WM. & MARY L. REV. 951, 952 (2004) (arguing that "the creation, selection, production, dissemination, promotion, sale, and use of expressive content . . . are undergoing revolutionary decentralization and disintermediation [and] . . . are increasingly being performed by individuals and disaggregated groups" rather than large corporations").

[10] *See* Frank Pasquale, *Breaking the Vicious Circularity*: Sony*'s Contribution to the Fair Use Doctrine*, 55 CASE W. RES. L. REV. 777, 789 (2005) (discussing how the film industry's failed effort to outlaw the VCR ultimately redounded to its benefit); William Fisher, *Don't Beat Them, Join Them*, N.Y. TIMES June 25, 2004, at A23 (claiming that one of the film industries' biggest failures in litigation led to one of its biggest revenue streams).

[11] For a definition of improvers, see Mark A. Lemley, *The Economics of Improvement in Intellectual Property*, 75 TEX. L. REV. 989, 1016-1020 (1997). Lemley notes that patent law traditionally does a much better job of protecting improvers than copyright law. Id., at 1022 (noting lack of doctrine analogous to "blocking patents" in copyright law), and 1023 (In the case of "improvers who have made a major contribution to social value, for example a work in which the new material predominates over infringing material . . . . patent law offers the possibility of complete immunity from infringement under the reverse doctrine of equivalents. Copyright law has no corresponding doctrine.").

[12] *See* LAWRENCE LESSIG, THE FUTURE OF IDEAS: THE FATE OF THE COMMONS IN A CONNECTED WORLD 120 (2002) (praising "innovation from the internet"); Yochai Benkler and Helen Nissenbaum, *Commons-Based Peer Production and Virtue, available at* http://www.nyu.edu/projects/nissenbaum/ (last visited Feb. 25, 2006).

Unfortunately, most of the proposed justifications have not been compelling to legislators or courts. Though their rationales for gradually strengthening copyright protection have been varied, they boil down to a common perception of unlicensed uses as free-riding.[13] "All this new creativity is great," leading copyrightholders admit. "But why permit it at my expense? Why not get a license like everyone else?" On this view, reductions of intellectual property rights are takings, to be compensated like any other transfer of property from private hands for public purposes.[14] The copyrightholder is always an innocent who has contributed something original to the store of knowledge, and those using any part of its work without a license are unfairly refusing to pay for the unalloyed benefit the work has conferred on society.[15]

---

[13] *See* Mark Lemley, *Property, Intellectual Property, and Free Riding*, 83 TEX. L. REV. 1031, 1039 (2005) (documenting "courts and scholars . . . preoccupied with the problem of "free riding.'"). As Lemley explains, "If the goal of creating property rights is to equate private and social costs and benefits by having the property owner internalize the social costs and benefits, those who 'free ride'--obtain a benefit from someone else's investment--are undermining the goals of the property system." *Id.* Lemley articulates a number of compelling reasons why law should not strive to internalize all positive externalities of intellectual property to its owner. *Id.*, at 1048 ("If I plant beautiful flowers in my front lawn, I don't capture the full benefit of those flowers--passers-by can enjoy them too. But property law doesn't give me a right to track them down and charge them for the privilege, though owners of property once tried unsuccessfully to obtain such a right."). Lemley provides several reasons why content owners should not be compensated for all positive externalities arising from their expression, and this article proposes another: the *negative* externalities content owners also produce due to information overload.

[14] U.S. courts have established that there can be a "taking" of patents or trade secrets. See Monsanto; Philip Morris v. Reilly. The recent French decision limiting recent legislation regarding iPods also suggested a takings rationale for preserving IP rights against legislative alteration. Were such a rationale adopted in the U.S., the much-observed "ratchet effect" of IP expansionism would move from empirical regularity to constitutionalized norm.

[15] The Clinton Administration's White Paper on internet policy

How can this view be challenged? Some cyberlaw theorists have argued that the social benefits of looser intellectual property protections greatly outweigh the costs of reduced protection.[16] This is likely true, but given valuation difficulties, it's hard to *prove* its truth in the economic patois that now dominates intellectual property policy.[17] This article

---

expressed this view, and it has animated copyright policy in legislation such as the Digital Millennium Copyright Act of 1998, which greatly expanded copyrightholders' rights to control digital uses of their works. *See* INFORMATION INFRASTRUCTURE TASK FORCE, INTELLECTUAL PROPERTY AND THE NATIONAL INFORMATION INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS (1995) [hereinafter *White Paper*] ("Some participants have suggested that the United States is being divided into a nation of information 'haves' and 'have nots' and that this could be ameliorated by ensuring that the fair use defense is broadly generous in the NII context. The Working Group rejects the notion that copyright owners should be taxed--apart from all others--to facilitate the legitimate goal of 'universal access.'") James Boyle has observed the fatuousness of the Working Group's assumption of a baseline of absolute copyrightholder control over expression. Boyle, *A Politics of Intellectual Property: Environmentalism for the Net*, 47 DUKE L. J. 87 (1997) ("Of course, given the goals of copyright law, it would have made just as much sense if the argument had been reversed, taking the fair use rights of users and consumers as the baseline.").

[16] Glynn Lunney, *Fair Use and Market Failure:* Sony *Revisited*, 82 B.U. L. REV. 975, 978 (2002) ("Only where the copyright owner has demonstrated by the preponderance of the evidence that the net benefit to society will be greater if a use is prohibited, should a court conclude that a use is unfair.").

[17] James Boyle, *Cruel, Mean or Lavish?: Economic Analysis, Price Discrimination and Digital Intellectual Property*, 53 VAND. L. REV. 2007, 2009 (2000) ("Information economics [offers] us plot-lines and econo-dramas, readymade images of types of dysfunction in information markets that sharpen our perceptions of potential risks and benefits. Unfortunately, it tends to offer them in antagonistic and mutually annihilating pairs."). I have attempted to address some of the valuation difficulties in *Toward an Ecology of Intellectual Property*, 8 YALE JOURNAL OF LAW AND TECHNOLOGY 78 (2006)); see also GHOSH, CODE (estimating the value of OSS with reference to

proposes another tack, analogizing information overload in the cultural environment to pollution of the physical environment.[18]

Environmental laws force polluters to pay for the ways they reduce the usefulness of air, water, and soil. Information law should adjust the rights of content creators in order to compensate for the ways they reduce the usefulness of the information environment as a whole. Every new work created contributes to the store of expression, but also helps make it more difficult to find whatever work a particular user needs or wants. The "search cost" of finding a needed work has been well-documented in the literature of information economics.[19] Copyright law should take negative externalities like search costs into account in its treatment of alleged copyright infringers whose work merely attempts to index, organize, categorize, review, or provide small samples of work generally.[20] They are not simply "free riding" off the labor of

---

the price of proprietary software that accomplishes the same or similar functions.

[18] James Boyle pioneered this approach in an article recommending that intellectual property law reformers learn from strategies adopted by environmentalist scholars and activists. Boyle, *Environmentalism for the Net, supra* note 15, at 109. Several other scholars have extended and developed this metaphor. *See,* e.g., Brett Frischmann, *An Economic Theory of Infrastructure and Commons Management*, 89 MINN. L. REV. 917 (2005); Peter K. Yu, *Intellectual Property and the Information Ecosystem*, 2005 MICH. ST. L. REV. 1; Dennis D. Hirsch, *Is Privacy Regulation the Environmental Law of the Information Age?, in* PRIVACY AND TECHNOLOGIES OF IDENTITY 239 (Katherine Strandburg and Daniela Stan Raicu, eds., 2006); Frank Pasquale, *Toward an Ecology of Intellectual Property*, supra n. 17.

[19] WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 167 (2003). Though Landes and Posner here concentrate on trademark law, the point is generalizable to any situation where a given product (information or otherwise) is being sought out by a potential purchaser or user.

[20] As Demsetz has argued, "property rights develop to internalize externalities when the gains of internalization become larger than the cost of internalization." Harold Demsetz, *Toward a Theory of*

copyrightholders, but rather are creating the types of navigational tools and filters that help consumers make sense of the ocean of data copyrightholders have created.[21]

By modeling information overload as an externality imposed by copyrighted works, this article attempts to provide a new economic justification for more favorable copyright treatment of a group of users collectively deemed "categorizers." Though categorizing is but one small part of what indexers, samplers, and search engines do, this synecdochic designation participates in the very phenomenon it is used to describe. For often the part is very revealing of the whole, and categorizers' efforts to reveal the whole via samples and snippets deserve far more solicitude from the law than they currently receives.

The argument proceeds as follows. Part II describes how conflicts between copyrightholders and those who categorize their content have complicated our understanding of fair use. The recent suit against the "Google Print" project has crystallized the legal issues at stake: 1) whether categorizers can provide small samples of copyrighted works to illustrate the categorizations made, and 2) whether a categorizer can copy an entire work digitally in order to prepare such samples.[22] Though doctrines protecting fair use and "intermediate copying" may protect such indexing activities, a series of court decisions limiting fair use have made their applicability questionable. Few areas of law are more unsettled.

Stepping back from the doctrine, Part III explains the

Property Rights, 57 AM. ECON. REV. PAPERS & PROC. 347, 348 (1967).

[21] As Pessach has observed, "Recent scholarly work has emphasized copyright's 'dynamic effect,' that is, the ongoing influence of expansive copyright protection toward an enclosure of the creative commons, and diminishment of cultural diversity." Guy Pessach, *Copyright Law as Silencing Restriction*, *supra* note 1, at 1067.

[22] Elisabeth Hanratty, *Google Library: Beyond Fair Use?*, 2005 DUKE L. & TECH. REV. 10, 14 (suggesting that Google's book digitization and indexing projects may run afoul of current copyright laws).

Formatted: Font: Not Bold

Formatted: Font: Not Bold

Formatted: Font: Not Bold

role of categorizers in the information ecosystem. While past legal scholarship has celebrated categorizers' creativity and utility, this article focuses on information overload as a negative condition that necessitates it. Just as the production of physical goods burdens the natural environment, the production of copyrightable expression imposes costs on the cultural environment. These information overload externalities include the increased "search cost" of finding the particular piece of expression one most wants, increased anxiety, and loss of solidarity via a fragmented public sphere.

The classic economic response to physical pollution is a "Pigouvian tax," designed to internalize the cost of emissions to their source.[23] Such a tax would be impossible in the cultural environment, because information overload is not an artifact of any particular act of creation but rather of the creative process *overall*.[24] The more practical method of addressing information overload is to empower the categorizers who can help us make sense of the "blooming, buzzing confusion" of the information society.

Part IV proposes a way of adjusting copyright doctrine to accomplish this goal. Because categorization projects are so necessary to counteract the negative effects of information overload, they deserve positive recognition in the *first* fair use

---

[23] *See* Dennis W. Carlton and Glenn C. Loury, *The Limitation of Pigouvian Taxes as a Long-Run Remedy for Externalities*, QUARTERLY J. ECON. 95 (1980): 559, 562 (1980) (defining Pigouvian taxes as efforts to internalize the costs of externalities to their creators, and discussing their disadvantages).

[24] Moreover, the old adage that "one man's trash is another's treasure" is commonly thought to be more true of cultural than physical products. I have resisted this sort of relativism in some venues; see, e.g., *Is Bach Better than Britney*, PRAWFSBLAWG, May 9, 2006, *available at* http://prawfsblawg.blogs.com/prawfsblawg/2006/05/is_bach_better_.html; *The Strange Romance of IP Expansionism and Aesthetic Relativism*, MADISONIAN THEORY, Aug. 4, 2006 *(available at* http://madisonian.net/archives/2006/08/04/the-strange-romance-of-ip-expansionism-and-aesthetic-relativism/). However, I recognize it as a persistent feature of what Rorty calls postmodern bourgeois liberal democracy.

factor, which focuses on the "purpose or character of the use."[25] Traditional analysis of whether the use is commercial and transformative has extremely limited utility in the categorization context. Courts can short-circuit these endlessly manipulable formal distinctions by recognizing categorization as a *per se* pro-defendant finding in the first fair use factor. Courts should also immunize initial digital copies of works *used for generating* such samples.

Information overload is an unintended but serious consequence of copyright law's success in creating incentives the production and distribution of expression. If courts grant content owners the rights to veto categorizers' efforts to make sense of given fields of expression, they will only exacerbate the problem. Designed to promote the "progress of the arts and sciences,"[26] copyright doctrine should privilege the efforts of those who make that progress accessible and understandable. Categorizers fill both those vital roles.

II. DILEMMAS OF CATEGORIZERS

Categorizers, reviewers, and indexers have existed for a some time..[27] But the legal questions they raise have become increasingly urgent as new technologies advance their effectiveness. Without digital technology, one could usually only find a book by subject if it were so relevant to the search that the "subject" words in a card catalog happened to match one's search. Now, digitized textual searches can make the entire book a de facto index card. Before web access, the only way to watch a film review actually illustrated by clips was to

[25] 17 U.S.C. § 107 (2004).

[26] U.S. CONST. art. I, § 8, cl. 8 (Congress has the power "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.") (cite articles glossing this; Dotan Oliar, etc.).

[27] JAMES F. ENGLISH, THE ECONOMY OF PRESTIGE 1 (2005) (noting that one type of categorizing, the "custom of . . . selecting outstanding individuals from various fields of cultural endeavor and presenting them with special tokens of esteem . . . dat[es] back at least to the Greek drama and arts competitions in the sixth century B.C.").

watch Gene Shalit or some other noted reviewer with a television show—which may in turn be owned by the financial backers of the movies reviewed. Now there is no technological barrier to reviewers putting up clips to graphically illustrate the picks and pans they dish out.

However, there are many legal barriers. Section 106 of the Copyright Act grants copyrightholders six exclusive rights—all of which may be violated by the would-be reviewer.[28] Any copy of the film made in order to isolate the clips violates the owner's exclusive right to copy.[29] The clip itself may be deemed a "derivative work, which only the owner may prepare."[30] Placing it on a website may be termed "distribution," or even a "public performance," depending on how many individuals have access to the site.[31] Even if the clip

---

[28] 17 U.S.C. § 106 (2004) (listing the right to reproduce the copyrighted work in copies or phonorecords; to prepare derivative works based upon the copyrighted work; to prepare derivative works based upon the copyrighted work; to perform the copyrighted work publicly (applies to literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works); to display the copyrighted work publicly (applies to 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); and, to perform the copyrighted work publicly by means of a digital audio transmission (which applies to sound recordings)).

[29] *See* 17 U.S.C. § 106(1) (2004) (Such copying infringes the owner's exclusive "Reproduction Right.")

[30] The exclusive right to prepare a derivative work based on a copyrighted work is codified in 17 U.S.C. §106(2). "A 'derivative work' is a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a 'derivative work.'" 17 U.S.C. § 101 (2004).

[31] The "distribution" right is codified in 17 U.S.C. § 106(3) (2004), and the "public performance" right is codified in 17 U.S.C. § 106(4) (2004).

has no negative impact on the market for the film, the copyrightholder can still sue for statutory damages—which range as high as $150,000 for a willful infringement.[32]

Regardless of these deterrents, thousands of individuals still post and comment on movie clips, texts, music, images, and other copyrighted works. To the extent they comment on the original, they have a decent shot at a "fair use" defense.[33] Fair use is copyright's "safety valve," permitting a wide range of uses unauthorized by copyrightholders.[34] To the extent the user's commentary is more voluminous than the clip or sample involved, the fair use defense is stronger.[35]

But as automated categorizers, such as search engines, have begun to enter the field, the limits of fair use are being tested. Search engines' ranking of cached content in response to a search inquiry is a "comment" on the content; as one court recently held, rankings are a form of expression protected under

---

[32] "[T]he copyright owner may elect, at any time before final judgment is rendered, to recover, instead of actual damages and profits, an award of statutory damages for all infringements involved in the action, with respect to any one work, for which any one infringer is liable individually, or for which any two or more infringers are liable jointly and severally, in a sum of not less than $ 750 or more than $ 30,000 as the court considers just." 17 U.S.C. § 504(c)(1) (2004). "In a case where the copyright owner sustains the burden of proving, and the court finds, that infringement was committed willfully, the court in its discretion may increase the award of statutory damages to a sum of not more than $150,000." 17 U.S.C. § 504(c)(2) (2004).

[33] "[T]he fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright." 17 U.S.C. §107 (2004).

[34] "[T]he "fair use" defense allows the public to use not only facts and ideas contained in a copyrighted work, but also expression itself in certain circumstances. . . . The fair use defense affords considerable latitude for scholarship and comment." *Eldred v. Ashcroft*, 537 U.S. 186, 219-220 (2003).

[35] *See* 17 U.S.C. § 107(3) (2004) (calibrating the fair use defense to the "amount and substantiality" of the work used).

the First Amendment.[36] Nevertheless, a wide array of content owners—ranging from book publishers to sports broadcasters to news services—have complained that Google's initial copy of their content into its databases, and subsequent provision of fragments of that content in response to search queries, is a violation of their copyrights.

Given the paucity of comment they offer, search engines pose the copyright issues raised by categorizers in the starkest form.[37] A long review encompassing a small film clip would strike many as a classic fair use (though the law of fair use is so unclear that even that conclusion cannot be made with certainty). But if a categorizer's only contribution consists of organizing and ranking content, should *that* excuse an infringement of copyright?[38]

---

[36] Search King Inc. v. Google Tech., Inc., No. CIV-02-1457-M, at 3 (W.D. Okla. Jan. 13, 2003) ("Two questions remain. First, are PageRanks constitutionally protected opinions? Second, if PageRanks fall within the scope of protection afforded by the First Amendment, is the publication of PageRanks *per se* lawful under Oklahoma law, thereby precluding tort liability premised on the intentional and even malicious manipulation of PageRanks by Google? The Court answers both questions in the affirmative.").

[37] In this respect, the dilemmas facing search engines and bond raters raise similar First Amendment issues. See Gregory Husisian, *What standard of care should govern the world's shortest editorials?: An analysis of bond rating agency liability*, 75 CORNELL L. REV. 411 (1990). In a fair use dispute, the search engine would likely call its ranking system a type of "editorializing" or comment on the items ranked and organized. However, it is unclear whether ranking in itself is "transformative" enough a use to merit a favorable fair use finding, or even a positive "first factor" finding.

Formatted: Font: Not Bold

Formatted: Small caps

Formatted: Small caps

[38] I have discussed the nature of the ranking in connection with the *SearchKing* case, in *Rankings, Reductionism, and Responsibility*, 55 Cleveland St. L. Rev. 777 (2006). In *SearchKing*, Google successfully resisted an action sounding in defamation and unfair competition by characterizing its page rankings as "opinion" rather than "fact." While I am wary of giving search engine rankings First Amendment protection in this way, I think that rankings contain enough expressive content to merit favorable copyright treatment analogous to that already enjoyed by reviewers.

As the rest of Part II demonstrates, that legal question is deeply contested. Since the search engine Google is now directly confronting legal challenges usually only hypothetically posed to categorizers, I focus the discussion on it. The Author's Guild, major publishers, and Agence France Press have all claimed Google's current and planned services infringe their copyrights.[39] The rest of this part examines the strength of each side's claims, setting up a discussion in Part III on which side *deserves* to be vindicated.

A. Case Study: The Google Print Project

Sergey Brin has said that the perfect search engine would be like the "mind of God."[40] Hubris aside, the comment reveals much about the aspirations of general purpose search engines. Their business model is predicated on their being the first source of information that "searchers" seek out when they need to find a site whose URL they do not know, or any resource they can't locate by themselves. Searchers will only trust a given search engine as an all-purpose portal if they can be reasonably assured that it has indexed the relevant information. If, for example, you are searching for "resorts near Cancun," and you know that with a given search engine

[39] *Complaint*, The Author's Guild v. Google, Inc., paragraph 3 (filed in S.D.N.Y.) (on file with author) (alleging that, "[b]y reproducing for itself a copy of those works that are not in the public domain . . . Google is engaging in massive copyright infringement."); *Complaint*, Agence France Press v. Google, Inc., paragraph 28 (filed in D.D.C., Mar. 17, 2005) (on file with author) (alleging that, "without AFP's authorization, Defendant is continuously and willfully reproducing and publicly displaying AFP's photographs, headlines, and story leads on its Google News web pages.").

[40] Quoted in Siva Vaidhyanathan, *A Risky Gamble with Google*, CHRON. HIGHER ED., Dec. 2, 2005, at B8. According to a recent BBC Program Inside the World of Google, Brin has also speculated (perhaps facetiously) about the feasibility of a brain implant which would directly report users' intentions to the search engine. BBC Documentary Archive, Inside the World of Google, interview with Susan Wojcicki.

Formatted: Font: Not Bold

only lists American sites, you'd be sure to avoid that one.[41]

Although the Cancun example is fanciful (given the international reach of the main general-purpose search engines operating in the U.S.), it does highlight the importance of *comprehensiveness* to a search engine.[42] For some time, search engines have jockeyed to claim that they have indexed the most websites.[43] Nevertheless, search engines have also conceded to individual site-owners' demands by not indexing sites that have a small programming script ("robots.txt") at the top of the "source pages."[44] This opt-out strategy has worked well in the online context because Digital Millennium Copyright Act immunizes "information service providers" from copyright liability for caching websites.[45]

Similar express immunities do not apply to books, but Google has nevertheless attempted to apply this opt-out approach to the texts it is indexing for its "Google Library" project. The quest for comprehensiveness has taken search engines beyond online sources and into the print world; all the major general-purpose search engines have begun scanning

---

[41] China adopts an analogous policy, having constructed "what is known in academic circles as the great firewall of China," which keeps government-censored sites from reaching its citizens. *See* JOHN BATTELLE, THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE 204 (2005).

[42] Danny Sullivan, *Search Engine Size War Erupts*, SEARCH ENGINE WATCH, Nov. 11, 2004, *available at* http://blog.searchenginewatch.com/blog/041111-084221. Database comprehensiveness benefits searchers because it increases the odds that searchers with esoteric search objectives will have a successful search.

[43] Danny Sullivan, *End Of Size Wars? Google Says Most Comprehensive But Drops Home Page Count*, SEARCH ENGINE WATCH, Sept. 27, 2005, *at* http://searchenginewatch.com/searchday/article.php/3551586.

[44] *See generally* Google, *Need to remove content from Google's index?*, *available at* http://www.google.com/webmasters/remove.html. Other large general-purpose search engines follow similar policies.

[45] 17 U.S.C. § 512(b) (2004).

books into an online database.[46] However, only Google is committed to copying copyrighted books into a database and making them searchable.[47] (In time, searches for "resorts near Cancun" might not just generate links to relevant websites, but also snippets of text from relevant books like *Fodor's Mexico*.) Google is permitting copyright owners to keep their books out of the database, provided they notify Google of their objections.[48] This "opt-out" approach has provoked the ire of the Author's Guild and major publishers, who sued to enjoin the Google project.[49]

Google has partnered with five major libraries in order to build a massive digital library based on their holdings.[50] To

---

[46] Elinor Mills, *Microsoft to offer book search,* Oct. 25, 2005, (*available at* http://news.com.com/Microsoft+to+offer+book+search/2100-1025_3-5913711.html (last visited Feb. 23, 2006).

[47] *See*, *e.g.*, Edward f, *Google Adds Library Texts to Search Database,* N.Y. Times, Nov. 3, 2005, at C1.

[48] See Oren Bracha, *Standing Copyright Law on Its Head? The Googlization of Everything and the Many Faces of Property*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=931426 (last visited Oct. 8, 2006) (using "the Google Print Library case in order to examine the role of opt-out arrangements in copyright law in general and in the context of digital libraries in particular . . . [and] argu[ing] that the choice between opt-in and opt-out is always a context-specific policy determination and that the digital-library context makes a compelling case for an opt-out regime.").

Field Code Changed

[49] Complaint, *supra* note 39, at 4 (asserting that "Google knew or should have known that the Copyright Act, 17 U.S.C. § 101 et seq. ("the Act"), required it to obtain authorization from the holders of the copyrights in these literary works before creating and reproducing digital copies of the Works for its own commercial use and for the use of others. Despite this knowledge, Google has unlawfully reproduced the Works and has announced plans to reproduce and display the Works without the copyright holders' authorization. Google intends to derive revenue from this program by attracting more viewers and advertisers to its site.").

[50] *See* Siva Vaidhyanathan, *A Risky Gamble with Google*, *supra* note 40, at B10 ("The University of Michigan at Ann Arbor has agreed to let Google scan its entire collection — some 7.8 million

date, the University of Michigan, Harvard University, Stanford University, the New York Public Library, and Oxford University have agreed to participate.[51] Each library partner has control over which of its books are scanned, and is to receive a copy of the books pursuant to conditions specified in its contract with Google.[52] Google plans to add over fifteen million library volumes to its electronic index (and is expected to spend $150 million dollars (approximately $10 per book) in compiling its digital library.[53] Users will be able to enter a search term and receive a results list of everywhere that term appears in Google's database.[54]

---

works — and Stanford University says it is keeping open the possibility of including "potentially millions" of its more than eight million volumes.").

[51] Contract Between Google Tech., Inc., and The University of Michigan, on file with author. The Author's Guild is not only concerned about Google's copy of the libraries' holdings, but also with the digital copy they plan to give to each of the libraries. At the Yale Search Conference, Author's Guild representative Paul Aiken fretted that universities might be so bold as to permit libraries to *make course reserve readings available online*. Presentation of Paul Aiken, Yale Information Society Project Conference on Search Engines, Dec. 3, 2005. Given the coursepack cases discussed below, Mr. Aiken likely has a legal basis for demanding compensation for such a use.

[52] *See* Susan Wojcicki, *Google Print and the Author's Guild*, Official Google Blog, *available at*

http://googleblog.blogspot.com/2005/09/google-print-and-authors-guild.html (last visited Dec. 20, 2005).

[53] John Markoff & Edward Wyatt, *Google is Adding Major Libraries to its Database*, N.Y. TIMES, Dec. 14, 2004, at A1.) For comparative information on costs of digitization, see *The Library of Amazonia*, Wired (describing process of creating Amazon's "Look-Inside-The-Book" feature).

[54] Even the project's critics acknowledge the enormous gains in access to knowledge the project promises. See Siva Vaidhyanathan, *A Risky Gamble with Google*, *supra* note 40, at B10 ("The dream of a perfect research machine seems almost within our reach. Google . . . announced late last year that it would digitize millions of bound books from five major English-language libraries. It plans to make available online the full text of public-domain books (generally those

Google has set forth different terms of distribution for the materials that will be included in Google Print, depending on their copyright status. Materials in the public domain will basically be made wholly available. If a page from a public domain work contains a search term entered by a user, she will be able to view the entire page that contains the search term and can also read through or print the entire book.

Copyright owners can also make their books available in this way, but very few have chosen to do so.[55] More likely, a copyright owner submitting a work to Google print will choose Google's second distribution option: to permit a searcher to view a full page of the text surrounding their search results and also a few pages on either side of the results. The result page will also provide links to online book retailers who sell the text.

For copyrighted works not recommended to Google by their publisher, a far more constricted result will appear. The user will only be able to view the bibliographic information and a few short sentences of text around their search term.[56] The results page will provide links to other information on the

---

published before 1923, plus government works and others never under copyright) and excerpts from works still in copyright."). These new technologies promise to transform many information-related industries. See, e.g., Deanna Barmakian, *Better Search Engines for Law*, 92 LAW LIBR. J. 399 (2000).

[55] One example is Larry Lessig's *Free Culture*, *available at* http://www.free-culture.cc/ (including a PDF of the book, information on where to buy a paper copy, reviews, a wiki, and audio versions). Though this strategy may not maximize revenues for a copyrightholder, there are clear advantages for an author trying to disseminate ideas. *See Lessig Interview with Frank Ling*, BERKELEY GROKS SCIENCE PODCAST (April 28, 2004), *available at* http://www.ugcs.caltech.edu/~frank/BerkeleyGroks_Lessig.htm (describing near-instantaneous creation of audio-version, wiki-version, and translations of the book after it was made available online.).

[56] *See* Google, *What you'll see when you search on Google Book Search*, *available at* http://print.google.com/googleprint/screenshots.html#excerpt (last visited Feb. 20, 2006) (explaining that "The Snippet View, like a card catalog, shows information about the book plus a few snippets – a few sentences of your search term in context.").

web regarding the search terms, links to retailers who sell the book (or used book sellers in case the work is out-of print), and a link suggesting places to find the book at the user's local library.[57] In addition, "To further protect the copyright holders, Google disables the user's print, save, cut and copy functions on the text display pages so that the user is limited to reading the information on the screen."[58]

B. Indeterminate Legal Analysis

Nevertheless, Google's project has provoked objections from leading copyrightholders and raises two important issues for copyright law. First, courts must decide whether the initial archival copy, necessary to the creation of the index, violates copyright law. Second, the status of whatever "snippets" the search engine generates in response to search inquiries is also at issue. The plaintiffs want Google to license each of these uses; Google claims each is a fair use. The following statutory text of § 107 governs the controversy:

> Notwithstanding the provisions of sections 106 and 106A [enumerating copyright holders' rights], the fair use of a copyrighted work ... 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;

[57] *Id.* ("These links aren't paid for by those sites, nor does Google or any library benefit if you buy something from one of these retailers. . . . To enforce limits on how many pages a user can view, we do connect some information -- your Google Account name -- with the books and pages that you've viewed.").

[58] Elizabeth Hanratty, *Beyond Fair Use?*, *supra* note 21, at 12.

> (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.[59]

A rather complicated caselaw has developed around each factor. Sections 1 and 2 below analyze how each might apply to the two stages of the Google Print project.[60]

1. The Initial Archival or Indexed Copy

To create a searchable database of books, Google first needs to scan each paper book into a digital file. Only a complete copy can serve the tool's purpose: to permit every word of every text to be a search query. Generally speaking,

---

[59] 17 U.S.C. § 107 (2004). For a rough guide on how these factors usually play out in the caselaw, *see* Eric Goldman, *Fair Use Cheatsheet, available at* http://www.ericgoldman.org/Courses/ipsurvey/fairusecheatsheet.pdf

[60] I have split the analysis into the two stages of the project, but even this initial step is contestable. While the entire searchable database of Google Library may be deemed a "transformative" use of the works involved, there is no guarantee that a court will focus on that stage of the process. It may well focus on the initial, archival copies Google makes, and deem those "nontransformative." A first factor finding that this essential aspect of the Google Library project is both commercial and nontransformative would seriously undermine any fair use defense Google might have. On the other hand, a court may elide the "initial archival copying" analysis by deeming that incidental or intermediate copying. *See* 17 U.S.C. § 117(c) (2004) (providing that it is not infringement for the owner of a machine to make a copy of a computer program if the copy is made automatically by virtue of the activation of a machine that contains a licensed copy of the computer program, for repair and maintenance purposes); Sega v. Accolade, 977 F.2d 1510, 1527 (9th Cir. 1993) ("[W]here disassembly is the only way to gain access to the ideas and functional elements embodied in a copyrighted computer program and where there is a legitimate reason for seeking such access, disassembly is a fair use of the copyrighted work, as a matter of law.").

the first fair use factor (relating to the "purpose and character of the use") has two dimensions: commerciality and transformativity.[61] Noncommercial uses are favored, as are those that add to, revise, or reconfigure the work.[62] An archival copy *itself* serves no commercial purpose—only the result it generates does. However, archival copies do not have transformative qualities; their very purpose is to faithfully reproduce the content they index. So the first fair use factor appears to be a "wash."

Inquiries into the second fair use factor, the "nature" of the copyrighted work, usually have two prongs as well—first, whether the work is fact or fiction,[63] and second, whether the work is published or unpublished.[64] Courts are less concerned about unauthorized uses of factual, published works than they are about fictional, unpublished works, and adjust fair use jurisprudence accordingly. Thus, it appears that this factor slightly favors Google: though some of the books are fictional, some are factual, and all are published.

The third fair use factor, "amount and substantiality," cuts against Google, since Google plans to copy each work in its entirety.[65] Finally, the fourth fair use factor, the effect on

---

[61] In the last Supreme Court case addressing fair use, the court suggested that transformativity analysis is more important than commerciality analysis. Campbell v. Acuff-Rose, 510 U.S. 569, 579 (1994) ("The central purpose of [first factor analysis] is to see . . . whether the new work merely supersede[s] the objects of the original creation, or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is transformative.").

[62] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

[63] The copying of factual works, including factual elements of creative works (such as the table of contents and index) is likely to be deemed fair use. Harper & Row Publ. Inc. v. Nation Enter., 471 U.S. 539, 563 (1985).

[64] For an overview of how courts have treated various uses under the second factor, see WILLIAM F. PATRY, THE FAIR USE PRIVILEGE IN COPYRIGHT LAW xv (2d ed., 1995).

[65] Copying an entire work militates against a finding of fair use unless mitigating circumstances are found. Infinity Broad. Corp. v.

the market, is entirely indeterminate.[66] A court might find that Google's failure to pay licensing fees for the right to archive the books is a grievous financial loss to the copyrightholders.[67] Or a court might find that such a licensing market is not "traditional, reasonable, or likely to be developed,"[68] and that

---

Kirkwood, 150 F.3d 104, 109 (2nd Cir. 1998). However, making a copy of a television program for home viewing at a later time entails copying the entirety of a creative work but has been found to be fair use. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 449-51 (1984).

[66] Jonathan Band's analysis suggests that Google has a better chance at "winning" this factor than is suggested here. *See* Band, *The Authors Guild v. The Google Print Library Project*, *available at* http://islandia.law.yale.edu/isp/search_papers/band.pdf ("Without question, the Print Library Project will increase the demand for some books. The project will expose users to books containing desired information, which will lead some users to purchase the books or seek them out in libraries (which in turn may purchase more copies of books in high demand). It is hard to imagine how the Library Project could actually harm the market for certain books, given the limited amount of text a user will be able to view.") Band also reasons that the "Google Publisher" option offered to copyrightholders (which provides them a share of any advertising revenue generated by the project) would circumvent the licensing demands mentioned below.

[67] The reasoning is obviously circular: "[A] potential market, no matter how unlikely, has always been supplanted in every fair use case, to the extent that the defendant, by definition, has made some actual use of plaintiff's work, which use could in turn be defined in terms of the relevant potential market. In other words, it is a given in every fair use case that plaintiff suffers a loss of a potential market if that potential is defined as the theoretical market for licensing the very use at bar." MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 13.05(A)(4) (2004). Nevertheless, it has guided a number of leading fair use decisions. *See* Frank Pasquale, *Breaking the Vicious Circularity*, *supra* note 11, at 777. (cataloging and criticizing such decisions as directly contradicting the Supreme Court's method of fourth factor analysis exemplified in *Sony v. Universal Studios*.).

[68] This is the legal standard for determining the legitimacy of licensing evidence in fourth factor analysis. *See* Matthew Africa, *The Misuse of Licensing Evidence in Fair Use Analysis: New*

the archival copies, standing alone, pose no threat to the commercial interests of copyrightholders.[69] Given the equivocal nature of the other three factors, the futility of the fourth factor's analysis makes fair use analysis of the initial archival copies a black box.[70]

Since there is no direct precedent for Google's service, it is difficult to apply caselaw here. Some commentators have argued that Google's archival, indexed copies are prohibited under a 2000 district court opinion, *UMG Recordings, Inc. v.*

Technologies, New Markets, and the Courts, 88 CAL. L. REV. 1145, 1168-69 (2000).

[69] *See* Jonathan Band, *The Authors Guild v. The Google Print Library Project*, *supra* note 60 ("Without question, the Print Library Project will increase the demand for some books. The project will expose users to books containing desired information, which will lead some users to purchase the books or seek them out in libraries (which in turn may purchase more copies of books in high demand). It is hard to imagine how the Library Project could actually harm the market for certain books, given the limited amount of text a user will be able to view.") Band also reasons that the "Google Publisher" option offered to copyrightholders would circumvent the licensing demands mentioned below. I provide doctrinal support for Band's broad reading of fourth factor, "effect on the market" analysis in a recent article. See Pasquale, *Breaking the Vicious Circularity*, *supra* note 67, at 790. Nevertheless, a strict application of rules developed in the coursepack cases would suggest that publishers could *still* claim a negative market effect because a fair use finding would deny them the chance to charge a licensing fee for snippets.

Formatted: Font: Not Bold

[70] This should not be surprising; as David Nimmer has suggested, "had Congress legislated a dartboard rather than the particular four fair use factors embodied in the Copyright Act, it appears that the upshot would be the same." David Nimmer, *'Fairest of Them All' and Other Fair Use Fairy Tales*, 66 L. & CONTEMP. PROBS. 263 (2003). *See also* MARJORIE HEINS AND TRICIA BECKLES, WILL FAIR USE SURVIVE? FREE EXPRESSION IN THE AGE OF COPYRIGHT CONTROL (2005) 16-25 (discussing various focus groups' and lawyers' complete uncertainty about whether certain uses of copyrighted work would count as fair).

MP3.com, Inc.[71] In that case, a company with an apparently forward-looking business model, MP3.com, copied 80,000 CD's in order to permit users who could *prove they already owned certain CD's* to "space-shift" their content—i.e., listen to streamed music via MP3.com's website, rather than lugging around their CDs.[72] The district court rejected every aspect of MP3.com's fair use defense, a decision that ultimately led to a settlement of $53.4 million indamages paid to Universal Music Group.[73]

MP3.com stands as a chilling landmark of copyright formalism.[74] For instance, the Recording Industry Association of America has used *MP3.com*-like reasoning to argue that owners of CD's have no right to rip them onto their personal MP3 players. Though the RIAA has long held the position that "it's perfectly lawful to take a CD that you've purchased, upload it onto your computer, put it onto your iPod," it and other large content holders recently made it clear that they believe they can revoke that right at any time and for any reason:

> Nor does the fact that permission to make a copy in particular circumstances is often or even routinely granted, necessarily establish that the

---

[71] **UMG Recordings, Inc. v. MP3.com, Inc., 92 F. Supp. 2d 349 (S.D.N.Y. 2000). *See* Hanratty, *Beyond Fair Use?*, *supra* note 22, at 10.**

[72] MP3.com would stream the music to the user once they had validated their ownership of the relevant CD. UMG Recordings, Inc. v. MP3.COM, Inc., 92 F.Supp. 2d 349, 350 (S.D.N.Y. 2000).

[73] Amy Harmon, *Deal Settles Suit Against MP3.com*, N.Y. Times, Nov. 15, 2000.

[74] For the leading contemporary jurisprudential definition of formalism, *see* Frederick Schauer, *Formalism*, 97 YALE L. REV. 509, 510 (1988) (defining formalism as "the way in which rules achieve their 'ruleness' [by] . . . screening off from a decisionmaker factors that a sensitive decisionmaker would otherwise take into account."). The fair use test codified in 17 U.S.C. 107 is more a standard than a rule, left deliberately open-ended in order to permit contextual judgment to trump mechanical pigeonholing of cases.

Formatted: Font: Not Bold

Formatted: Font: 14 pt

> copying is a fair use when the copyright owner withholds that authorization. [A policy permitting copies is] simply a statement about authorization, not about fair use.[75]

If content owners succeed in preventing iPod owners from ripping their own music to their own personal music players, it's hard to imagine Google achieving the right to copy content en masse.

Unless, of course, courts overcome the formalism of *MP3.com* and look at the initial copy *in the context of* Google's larger purposes in creating an index of books. Both courts and Congress have recognized the legitimacy of intermediate copying in contexts where a spare copy was necessary to a noninfringing goal.[76] For example, computer repairers can make an identical but non-infringing copy of a program on a hard drive to perform maintenance,[77] and programmers can

---

[75] Recording Industry Association of America, et al., *Joint Reply Comments*, DMCA RULEMAKING ON EXEMPTIONS TO PROHIBITION ON CIRCUMVENTION OF COPYRIGHT PROTECTION SYSTEMS FOR ACCESS CONTROL TECHNOLOGIES, Feb. 2, 2006, *available at* http://www.copyright.gov/1201/2006/reply/11metalitz_AAP.pdf.

[76] Paul Ganley, *Google Book Search: Fair Use, Fair Dealing and the Case for Intermediary Copying*, 1, *available at* http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID875384_code549519.pdf?abstractid=875384&mirid=1 (last visited Feb. 20, 2006) (proposing, based on American law, a "specific defence for 'intermediary copying premised on the 'temporary copies' exception . . . of the Copyright, Designs, and Patents Act of 1988 . . . and alternatively a new defence of 'fair dealing for informational purposes'. . . .").

Field Code Changed

[77] In 1993, the Ninth Circuit refused to find a repairman's unauthorized "copying" of a program to a computer's hard drive a fair use, despite the necessity of doing so merely to turn the computer on. *MAI Systems Corp. v. Peak Computer, Inc.*, 991 F.2d 511 (9th Cir. 1993). Five years late, the Digital Millennium Copyright Act of 1998 legislatively overruled that part of the holding via 17 U.S.C. § 117(c) (providing that it is not infringement for the owner of a machine to make a copy of a computer program if the copy is made automatically by virtue of the activation of a machine that contains a

make a non-infringing copy of a program in order to reverse engineer it.[78]

2. Snippets

An intermediate copying, fair use defense can only succeed if the final product—a database providing "snippets" as samples of books—is noninfringing. This is an extraordinarily difficult question in copyright law.[79] Google has argued that the snippets are a fair use of the books from which they derive. But copyrightholders may argue that the

---

licensed copy of the computer program, for repair and maintenance purposes).

[78] *Sega v. Accolade*, 977 F.2d 1510 (9th Cir. 1993). The rule of *Sega* is often put as follows: "[W]here disassembly is the only way to gain access to the ideas and functional elements embodied in a copyrighted computer program and where there is a legitimate reason for seeking such access, disassembly is a fair use of the copyrighted work, as a matter of law." *Id.* at 1527-28. Beyond that bright line rule, *Sega* may well also stand for a judicial willingness to permit copying of small portions of competitor's software in order to promote interoperability.

[79] Unfortunately, the formalism of the statutory fair use factors makes much of the analysis of the initial archival copy planned by Google applicable to the "snippets" that search queries would generate. The fourth factor licensing analysis is indeterminate, and the second, "nature of the work" query is slightly in Google's favor, given the published status of all the works. The third factor shifts to Google's favor, since the "snippets" are only a tiny fraction of the work as a whole (generally, the three or four lines above and below the search term.) The first factor may shift to Google's favor as well, since the generation of snippets is far more transformative than the mere copying of texts. Yet the commercial dimensions of the project are far more evident here, since Google will be selling advertising on pages that feature the snippets. So the first factor, and brute fair use analysis generally, appears indeterminate here as well. For a commentary on the formalism and ultimate incoherence of the factors, *see* Michael Madison, *A Pattern-Oriented Approach to Fair Use*, 45 WM. & MARY L. REV. 1525, 1550 (2005) (describing "conflicts and complications in the statutory text" that governs fair use determinations).

snippets themselves are copyrighted *works.*[80] Consider a query for "poetry about plums" that returns an anthology by William Carlos Williams. The entire 28-word poem, *This is just to say*, might show up as a snippet.[81] It is hard to argue against the conclusion that this snippet is a reproduction of copyrighted work.

Though such a result might merely lead one to exclude short poetry from the digitized databases, run of the mill protection of "microworks" raises other problems.[82] Though "[w]ords and short phrases such as names, titles, and slogans" clearly are not copyrightable,[83] clever advocacy has eroded

---

[80] Hughes, *Size Matters (or Should) in Copyright Law*, 74 FORDHAM L. REV. 574, 576 (citing 17 U.S.C. 101, 102) ("The Copyright Act defines a 'collective work,' a 'work made for hire,' 'literary works,' a 'joint work,' and 'a work of visual art.' But the law runs silent on the foundational concept on which these definitions are built."). Hughes notes that "Under the fair use doctrine, the smaller the amount copied, the fairer the copying. Courts have also deployed a "*de minimis*" copying rule separate from, and antecedent to, any fair use analysis. The de minimis rule expressly allows the copying of small and insignificant portions of the plaintiff's work. However, neither of these doctrines is an adequate device because each takes the work as its starting point to measure the amount of copying." *Id.* Many courts have been unwilling to reject claims for control of very small "works." See, e.g., Thomas Cotter, *Memes and Copyright*, 80 TUL. L. REV. 331, 332 n.3 (2005) (citing Mattel, Inc. v. Goldberger Doll Mfg. Co., 365 F.3d 133, 134-37 (2d Cir. 2004) (noting that "the United States Court of Appeals for the Second Circuit vacated a judgment that Barbie's eyes, nose, and mouth were uncopyrightable standard features.")).

[81] The entire poem is 28 words long. *See* William Carlos Williams, *This is Just to Say*, quoted in its entirety in CAMILLE PAGLIA, BREAK, BLOW, BURN 134 (2005) (apologizing for finishing off the cold plums in the refrigerator).

[82] I borrow the term "microworks" from Hughes. Hughes, *Size Matters*, *supra* note 80, at 576.

[83] 37 C.F.R. § 202.1(a) (2004). This rulemaking by the the Copyright Office stated that "[w]ords and short phrases such as names, titles, and slogans; familiar symbols or designs; mere variations of typographic ornamentation, lettering or coloring; [and]

extant judicial resistance to the copyrightability of abridgments, samples, or small portions of works.[84]

Under the fair use doctrine, it usually is the case that the less of a work is copied, the more likely the use is fair.[85] However, the fourth "effect on the market" component of fair use cases has sometimes devolved into a judicial insistence that any use that can be paid for, should be paid for.[86] The

---

mere listing of ingredients or contents" are not subject to copyright. *Id.*

[84] *See* Hughes, *Size Matters*, *supra* note 80, at 577 (observing that the "creeping protection of 'microworks'" occurs primarily because fair use and de minimis doctrine (permitting the copying of small portions of plaintiffs' work) "each take[] the work as its starting point to measure the amount of copying," leaving open the possibility that the court will define the work narrowly and characterize the plaintiff as copying the whole of it.). A similar issue is known as the "denominator problem" in takings law; as the Supreme Court acknowledged, "Because our test for regulatory taking requires us to compare the value that has been taken from the property with the value that remains in the property, one of the critical questions is determining how to define the unit of property 'whose value is to furnish the denominator of the fraction.'" Keystone Bituminous Coal v. DeBenedectis, 480 U.S. 470, 497 (1987) (quoting Frank Michelman, *Property, Utility, and Fairness: Comments on the Ethical Foundations of Just Compensation Law*, 80 HARV. L. REV. 1165, 1192 (1967)).

[85] *See* 17 U.S.C. § 107(3) (calibrating fair use protection to "amount and substantiality" of the work used). *But see* Los Angeles News Serv. v. KCAL-TV Channel 9, 108 F.3d 1119, 1123 (9th Cir. 1997) (noting that even taking a small portion of the work may lead to a negative third factor finding if it is the "heart of the work"); Religious Technology Center v. Lerma, 908 F. Supp. 1362, 1367 (E.D. Va. 1997) (same); Sundeman v. Seajay Soc'y, 142 F.3d 194, 205 (4th Cir. 1998) (mentioning "heart of the work" doctrine); L.A. News Serv. v. Reuters Tv Int'l, 149 F.3d 987, 994 (9th Cir. 1998) (same).

[86] *See* Wendy J. Gordon, *Market Failure and Intellectual Property: A Response to Professor Lunney*, 82 B.U. L. REV. 1031, 1032 (2002) (criticizing this development as a misinterpretation of her seminal article on the topic, *Fair Use as Market Failure: A Structural and Economic Analysis of the* Betamax *Case and Its Predecessors*, 82 COLUM. L. REV. 1600, 1601 (1982) (which asserted that "the courts

"coursepack" cases, for instance, required that instructors at universities get copyright permission for reproducing *portions* of books in materials photocopied for classes.[87] On one reading, these cases require a copier to license any portion of a work for which a market is "'traditional, reasonable, or likely to be developed."[88] Given the extent to which the internet reduces the transaction costs of micropayment systems,[89] publishers might argue that a new branch of the extant Copyright Clearance Center would easily set up a marketplace for snippets.[90] Mere token efforts by copyrightholders to set up

---

and Congress have employed fair use to permit uncompensated transfers that are socially desirable but not capable of effectuation through the market")).

[87] *See*, e.g., Bernard Zidar, *Fair Use and the Code of the Schoolyard: Can Copyshops Compile Coursepacks Consistent with Copyright?*, 46 EMORY L.J. 1363, 1364 (suggesting that the "prevailing analysis by which fair use is determined effectively denies fair use protection to all commercial copyshops").

[88] Princeton Univ. Press v. Mich. Document Servs., 99 F.3d 1381, 1407 (6th Cir. 1997) (en banc) (quoting American Geophysical Union v. Texaco, 60 F.3d 913, 930-31 (2d Cir. 1994), cert. denied, 116 S.Ct. 594 (1995)).

[89] Micropayments are "small digital payments of between a quarter and a fraction of a penny." Clay Shirky, *Fame v. Fortune: Micropayments and Free Content*, first published September 5, 2003 on the "Networks, Economics, and Culture" mailing list, *available at* http://www.shirkey.com/writings/fame_vs_fortune.html. Internet services like BitPass, FirstVirtual, Cybercoin, Millicent, Digicash, Internet Dollar, and Pay2See have served as micropayment systems.

[90] For example, in *Princeton Univ. Press v. Michigan Document Services*, the Sixth Circuit ruled that publishers' development of the Copyright Clearance Center made it reasonable for them to demand that universities license even brief excerpts of copyrighted works included in coursepacks. *See, e.g.*, Princeton Univ. Press v. Mich. Document Servs., 99 F.3d 1381, 1386-87 (6th Cir. 1997) (en banc) (holding that defendant's photocopying of plaintiff's copyrighted work was not a fair use because it harmed the reasonable potential market value of the copyrighted works); Am. Geophysical Union v. Texaco Inc., 60 F.3d 913, 930-31 (2d Cir. 1995) (same). For a complete collection of documents relating to the coursepack case, *see* Stanford University Libraries, *Copyright and Fair Use:* Michigan Document Services *and Coursepacks*, *available at*

a licensing market for "snippets," then, could fatally undermine Google's argument for unlicensed use. And many publishers have made more than token efforts, brokering deals with Amazon's similar "Look Inside the Book" program (though they have refused to disclose the terms of the deal).[91]

Those unfamiliar with copyright arcana might find it odd that the mere ability to charge for snippets should have anything to do with a legal requirement to do so.[92] As Gideon Parchomovsky argues,

> [T]he ability to charge by itself cannot possibly determine legal rights. A hoodlum might have the ability to charge protection fees, and yet no one would argue that this in itself gives him a right to do that . . . . Absent an underlying theory of rights, the ability to charge is normatively meaningless.[93]

---

http://fairuse.stanford.edu/primary_materials/cases/michigan_document_services/index.html (last visited Feb. 27, 2006). This site is one excellent example of the value of the work of "legal categorizers" on the internet.

[91] Gary Wolf, *The Great Library of Amazonia*, 11.12 WIRED 76, Oct. 23, 2003, *available at* http://www.wired.com/wired/archive/11/12/amazon_pr.html (last visited Feb. 27, 2006) (Amazon "created an unrivaled digital archive of more than 120,000 books" by "negotiat[ing] contracts with hundreds of publishers."). Google aims to digitize millions of titles held by five major university libraries, and argues that given the millions of "orphan works" (with no clear copyright ownership) and the unreasonableness of some publishers, there is no way to attain a comprehensive index via negotiations.

[92] As I have expressed elsewhere, I think that both *MP3.com* and *Princeton Univ. Press* are misguided as a matter of copyright policy, ultimately undermining the constitutional purpose of intellectual property protection. There is no sound economic rationale for taking the position that every use (outside a narrow band of "classic" fair uses) that *can* be paid for *should* be paid for. Frank Pasquale, *Breaking the Vicious Circularity*, *supra* note 11, at 781..

[93] Gideon Parchomovsky, *Fair Use, Efficiency, and Corrective Justice*, 3 LEGAL THEORY 347, 359 (1997).

Given the utilitarian contours of U.S. copyright law,[94] content owners try to supply such an "underlying theory" by arguing that the better they perfect their control over the use of their works, the better they can maximize the development of future works.[95] Their opponents, known variously as the Open Access Movement,[96] Free Culture Movement,[97] or the Copyleft,[98] argue

---

[94] *See* PAUL GOLDSTEIN, COPYRIGHT'S HIGHWAY 26 (1994) (contrasting utilitarian American approach to copyright with natural rights perspective that is more common internationally).

[95] Jane C. Ginsburg, *Legal Protection of Technological Measures, Protecting Works of Authorship; International Obligations and the US Experience*, 29 COLUM. J. L. & ARTS 11, 13 (2005) (stating that that "[t]he US experience to date indicates that legal protection for technological measures has helped foster new business models that make works available to the public at a variety of price points and enjoyment options, without engendering the 'digital lockup' and other copyright owner abuses that many had feared."); June M. Besek, *Anti-Circumvention Laws and Copyright: A Report from the Kernochan Center for Law, Media, and the Arts*, 27 COLUM. J. L. & ARTS 385, 486 (2004) (finding that "[s]ection 1201 has been successful in stimulating new means of distribution and promoting consumer choices with respect to a variety of works, particularly sound recordings, motion pictures and television programming, and literary works.").

[96] Public Library of Science, *Open Access*, at http://www.plos.org/about/openaccess.html (last visited Feb. 26, 2006) (defining open access publications as those for which "[t]he author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.")

[97] *See Free Culture Manifesto*, at http://freeculture.org/manifesto.php (last visited Feb. 26, 2006) ("We won't allow the content industry to cling to obsolete modes of distribution through bad legislation. We will be active participants in a free culture of connectivity and production, made possible as it never was before by the Internet and digital technology, and we will fight to prevent this new potential from being locked down by corporate and legislative control.").

that such a strategy tends only to enrich dominant players, and that an alternative, more open-access policy, would maximize expression.[99]

This article takes a different normative tack in advocating for copyright reform. Instead of arguing that more access to works for categorizers would increase the *amount* of expression, I take the position that such a policy would reduce the costs of information overload generated by the abundance of works. I expand and develop my critique of cases like *MP3.com* and *Princeton University Press* in Parts III and IV below. The view that "every use that can be paid for, should be paid for" is parasitic on an assumption that *every copyrighted work* somehow contributes positively to the store of expression. As our understanding of information overload externalities grows, such an assumption is becoming increasingly naïve.

III. FROM MAXIMIZING TO OPTIMIZING EXPRESSION

Copyright law may not only permit content owners to scuttle search engines' quest for an authoritative index of copyrighted expression, but may also chill the efforts of smaller categorizers who want to sample works for illustrative purposes. This result is unacceptable in an age of data proliferation, as *metadata* (data about data) becomes an ever more important resource.[100] Some balance is needed before

---

[98] Richard Stallman, *Copyleft*, at http://www.gnu.org/copyleft/ (last visited Feb. 26, 2006) ("Copyleft is a general method for making a program or other work free, and requiring all modified and extended versions of the program to be free as well."). The copyleft tries to assure that open access versions of software remain open access by trying to require any future users and developers of such software to make their products open access as well.

[99] *See*, e.g., HENRY C. MITCHELL, THE INTELLECTUAL COMMONS: TOWARD AN ECOLOGY OF INTELLECTUAL PROPERTY 137 (2005) (decrying dominant role of content industries in setting policies and arguing for the restoration of "user rights and social utility as factors in IP regimes.").

[100] As David Weinberger comments on the Google Print controversy, "Despite the present focus on who owns the digitized

aggressive litigation strategies permit content owners to leverage control over copyrighted works into veto power over any project that ranks, reviews, organizes, or even refers to them.[101]

Unfortunately, extant academic commentary on copyright tends to obscure the importance of categorizers by elevating the value of all copyrighted work indiscriminately. Section A below explores this rhetoric, concluding that normative discourse on copyright needs to balance accounts of how to maximize expression with frank recognition of expression's costs. Section B suggests one way of recognizing expression's costs, by analogizing the problem of information

---

content of books, the more critical battle for readers will be over how we manage the information about that content--information that's known technically as metadata." Weinberger, *Crunching the Metadata*, BOST. GLOBE, Dec. 15, 2005, at B1. For a definition of metadata, *see Metadata, available at* http://www.webopedia.com/TERM/M/metadata.html ("Metadata describes how and when and by whom a particular set of data was collected, and how the data is formatted"). For a broader definition, *see Metadata, available at* http://en.wikipedia.org/wiki/Metadata (last visited Jan. 21, 2006) ("Metadata . . . literally 'data about data,' is information that describes another set of data. A common example is a library catalog card, which contains data about the contents and location of a book: It is data about the data in the book referred to by the card. Other common contents of metadata include the source or author of the described dataset, how it should be accessed, and its limitations. Another important type of data about data is the links or relationship among data."). I am aware that the "wikipedia" is not an authoritative source, but its treatments of technical subjects are sufficiently clear and reliable to stand here as a placeholder for a later, stabler source of definitions. Moreover, to the extent they reflect *current* public understanding of terms, they may well prove superior to dictionary definitions written long ago.

[101] An anti-commons develops when fragmented ownership causes high transaction costs that stunt the development of a resource. Michael Heller, *The Tragedy of the Anticommons: Property in Transition from Marx to Markets*, 111 HARV. L. REV. 621 (1998). Just as Garrett Hardin observed that insufficient propertization could lead to overuse of resources, anticommons theorists show how excessive propertization can lead to underuse of resources.

overload to pollution in the environmental context. If copyrightholders insist on characterizing robust fair use exceptions as a "tax" or "taking" of potential licensing fees, it is just as appropriate to characterize those exceptions as compensation to society for the information overload they have helped create.

A. The Maximizing Paradigm

Full accounting for information overload externalities would balance any putative sacrifice in licensing fees that fair use for categorizers would cause. Unfortunately, current normative discourse on copyright tends to occlude this possibility. Scholars on both sides of copyright disputes generally assume that more expression is always better—that copyright's constitutional purpose (to promote the progress of the arts and sciences) is primarily about the creation of incentives for expression.[102] Copyright expansionists point to the incentive effects of increased legal protection of works.[103] Promoters of an expansive public domain claim that perfect control by extant owners will excessively raise the cost of the "raw materials" used in future information creation and thereby prevent more expression than such control promotes.[104] Such assumptions, which tend to ignore or downplay the *costs* of expression, unfairly disadvantage categorizers in the normative debates surrounding their uses of copyrighted work.

---

[102] Jane C. Ginsburg, *Legal Protection of Technological Measures*, *supra* note 95, at 37 ("The technological measures that reinforce legal control may enable and encourage authorial entrepreneurship, because authors may be able to rely on these measures to secure the distribution of and payment for their works, and new business models may therefore emerge.").

[103] *Id.*

[104] JAMES BOYLE, SHAMANS, SOFTWARE, AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY 37 (1996) ("[P]erfect information is one of the elements of the perfect market. If information can be commodified, then a host of transaction costs are introduced into information flow and a limited monopoly is granted in the midst of a system supposedly premised on competition.").

Much scholarship in copyright can be characterized as an effort to reconcile the curiously dual roles of copyrighted expression in market economies. The entire field of intellectual property is premised on the idea that the commodification of information will create incentives for its production. However, one of the aspirational conditions of any market is "perfect information;" i.e., consumers' and producers' full knowledge of the utility of products.[105] To the extent such information is protected by copyright law, it is less than universally available. The conflict is particularly acute in the sale and licensing of copyrightable expression, an "experience good" whose value often cannot be fully ascertained without some exposure to it. Costless exposure to the product can often substitute for the product itself—an insight formalized as Arrow's "paradox of disclosure."[106]

Another janiform role of information in the economy lies in its status as both "finished good" and "raw material" for future creation.[107] Legal rules that raise the price for, say, copyrighted musical lyrics, may incentivize more lyricists, but

---

[105] George Stigler, *The Economics of Information*, in FOUNDATIONS OF THE ECONOMIC APPROACH TO LAW 259, 263 (Avery Weiner Katz ed., 1998); George Akerlof, *The Market for Lemons: Qualitative Uncertainty and the Market Mechanism*, in FOUNDATIONS OF THE ECONOMIC APPROACH TO LAW 265, 268 (Avery Weiner Katz ed., 1998); Joseph E. Stiglitz, *Information*, in THE CONCISE ENCYCLOPEDIA OF ECONOMICS (David R. Henderson, ed. 2005), *available at* http://www.econlib.org/library/Enc/Information.html (last visited Jan. 10, 2006) ("Many of the central theories and principles in economics are based on assumptions about perfect information."). For the legal implications of this assumption, *see* Boyle, *Environmentalism for the Net*, *supra* note 15, at 92 ("Barriers to the free flow of information lead to the inhibition of innovation [and] inadequate circulation of information.").

[106] *See* Kenneth Arrow, *Economic Welfare and the Allocation of Resources for Invention*, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY 124 (Kenneth Arrow, ed., 1962).

[107] JAMES BOYLE, SHAMANS, SOFTWARE, AND SPLEENS: LAW, *supra* note 98, at 38. ("[O]ne important use of 'fair use' law is to make sure that future creators have available to them an adequate supply of raw materials.").

ultimately raise the price of recordings. For this reason, the recording industry has long (and successfully) lobbied Congress to compulsorily license musical lyrics and compositions at a low, flat rate.[108] Of course, the industry is considerably less willing to recognize that its own music is "raw material" for DJ's, filmmakers, and other creative workers.[109]

Many scholars have tried to develop proposals that reduce the "commodification/perfect information", and the "finished good/raw material" tensions.[110] Reliance on information economics has enhanced the validity of this work generally over the past two decades. However, the verdict of information economics is frequently equivocal in particular disputes, given the two tensions discussed above. Two schools—the "copyleft" and "copyright expansionists"—have each developed rival prescriptions for maximizing expression in the midst of these tensions.

---

[108] *See* Theresa Bevilacqua, *Time to Say Good-Bye to Madonna's American Pie: Why Mechanical Compulsory Licensing Should be Put to Rest*, 19 CARDOZO ARTS & ENT'T L. J. 285, 285 (2001) (noting that "anyone who desires can make an arrangement of an existing work, record the arrangement, and sell it."); 17 U.S.C. § 115 (2004).

[109] Joel Rose, *Copyright Laws Severely Limit the Availability of Music*, ALL THINGS CONSIDERED, Jan. 9, 2006 (reporting on Library of Congress study indicating that "over 70 percent of American music recorded before 1965 is not legally available in the United States.") According to a recent report on the "clearance culture," "in the music industry, the practice of requiring a license for even the smallest sample is entrenched." MARJORIE HEINS AND TRICIA BECKLES, WILL FAIR USE SURVIVE? FREE EXPRESSION IN AN AGE OF COPYRIGHT CONTROL 5 (2005). The contrast with the recording industry's own treatment of composers and lyricists could not be more stark. Bevilacqua, *Compulsory Licensing*, *supra* note 102, at 296 ("The RIAA [has] argued the compulsory license [for lyrics and compositions has] to be retained because the threat of an industry-wide monopoly might resurface if authors were allowed to grant exclusive licenses.").

[110] *See, e.g.*, William W. Fisher III, *Reconstructing the Fair Use Doctrine*, 101 HARV. L. REV. 1659 (1988); WILLIAM W. FISHER III, PROMISES TO KEEP: LAW AND THE FUTURE OF ENTERTAINMENT 47 (2004) (recommending compulsory licensing scheme for all works).

Scholars and policymakers who favor more expansive copyright protections also draw on analogies between real and intellectual property. The most famous is the tragedy of the commons. Just as a common pasture may be overused if the entire community can use it to graze their cattle, so too unowned or "orphan" works might be mis- or overused.[111] Similarly, potential investors will balk at committing resources to a project whose benefits they cannot reliably and exclusively appropriate or sell. Just as private ownership of land has proven to be an engine of economic growth,[112] private ownership of copyrightable works is the key to their development.[113] Thus, expansionists argue that IP laws should expand the scope and term of IP rights, and grant rights of control over earlier stages of the development of information goods.[114]

A number of critical IP scholars have attempted to refute these models by emphasizing the benefits of more open access to works.[115] The Free Culture movement has countered

---

[111] Garrett Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968).

[112] HERNANDO DESOTO, THE OTHER PATH 8 (1990) (theorizing secure private ownership of land as the linchpin of economic development).

[113] See, William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 487 (2003) (arguing that untrammeled copying or use of a celebrity's image could "could prematurely exhaust the celebrity's commercial value, just as unlimited drilling from a common pool of oil or gas would deplete the pool prematurely."); *but see* Stewart Sterk, *Intellectualizing Property: The Tenuous Connections Between Land and Copyright*, 83 WASH. U. L. Q. 417, 421 ("Real property rights operate to avoid the 'tragedy of the commons'--a problem that does not arise with intellectual works--because once created, those works, unlike land, are non-rivalrous public goods.")

[114] *See* Jane C. Ginsburg, *Legal Protection of Technological Measures*, *supra* note 95, at 12; Bruce Kuhlik and Richard Epstein, *Is There a Biomedical Anticommons?*, REGULATION 54 (Summer 2004) (answering no).

[115] See Donna Wentworth, *What Does 'Copyfight' Mean?*, *available at* http://copyfight.corante.com/archives/2005/07/30/what_does_copyfig

each of these contentions with arguments that limits on "real world" property rights should be extended to IP.[116] Scholars have developed theories of the "tragedy of the anticommons"[117] or "comedy of the commons."[118] They argue that an open-access regime, or "low IP equilibrium," can be far more productive than a situation where early innovators "lock up" certain fields and initiate escalating arms races toward

---

ht_mean.php (last visited Feb. 26, 2006) ("The copyfight is the battle to keep intellectual property tethered to its purpose, understanding that when IP rights are pushed too far, they can end up doing exactly the opposite of what they're intended to do."). The Copyfight blog has a long list of "copyfighters" on the left side of the page. *Id.*

[116] Michael Carrier, *Cabining Intellectual Property Through a Property Paradigm*, 54 DUKE L.J. 1, 3 (2004) (arguing that "property is not as absolute as it is often claimed to be," surveying "fifty doctrines of property law," "distilling restrictions centered on development, necessity, and equity" from these restrictions, and "import[ing] these categories of limits into" intellectual property law.)

[117] *See*, e.g., Dan Hunter, *Cyberspace as Place and the Tragedy of the Digital Anticommons*, 91 CAL. L. REV. 439, 441 (2003) ("Anticommons property occurs when multiple parties have an effective right to prevent others from using a given resource, and as a result no one has an effective right of use."); Michael S. Mireles, *An Examination of Patents, Licensing, Research Tools, and the Tragedy of the Anticommons in Biotechnology Innovation*, 38 U. MICH. J. L. REFORM 141 (2004) (arguing that fragmented intellectual property rights hamstring innovation); *but see* Bruce Kuhlik and Richard Epstein, *Is There a Biomedical Anticommons?*, *supra* note 108, at 54. ("Without ample patent protection, no combination of first-mover advantages or altruism will generate the capital sums needed. Reducing the patentees' right to exclude or its power to price is a partial repeal of the patent grant with mischievous social consequences.").

[118] *See, e.g.*, Carol Rose, *The Comedy of the Commons: Custom, commerce, and Inherently Public Property*, 53 U. CHI. L. REV. 711 (1986) ("There is . . . an extensive academic and judicial discussion of the possibility that certain kinds of property ought to be public."); DAVID BOLLIER, SILENT THEFT: THE PRIVATE PLUNDER OF OUR COMMON WEALTH 34-40 (2002) (celebrating the nonrivalry of consumption of information resources).

acquiring more and more IP rights.[119]

Despite their opposing conclusions on most particulars, scholars in favor of open access share some common assumptions with copyright expansionists. Both schools emphasize that the decision to make any particular *stage* of intellectual property production commodifiable will draw investment to that stage, while diverting resources away from earlier or later stages. For example, if very short, basic sequences of programming code are copyrightable, software firms would invest a great deal in developing (and copyrighting) those sequences. If only larger structures (or collections) of code are copyrightable, then investment is shifted toward those structures. Copyright expansionists tend to argue that commodification should be permitted at earlier stages of production, since competition in "innovation markets" is often *for*, and not *within*, a market.[120] The Free Culture movement tends to argue that only later stages of production should be commodifiable, in order to ensure a larger "public domain" of materials is open for everyone's use and development.

Neither side tends to acknowledge the *costs* of producing ever more copyrightable expression, or to investigate deeply the

[119] See Kal Raustiala and Christopher Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design*, *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=878401 (last visited Feb. 26, 2006) ("Why, when other major content industries have obtained increasingly powerful IP protections for their products, does fashion design remain mostly unprotected--and economically successful? . . . . We argue that the fashion industry counter-intuitively operates within a low-IP equilibrium in which copying does not deter innovation and may actually promote it.").

[120] *See* Richard J. Gilbert and Steven C. Sunshine, *Incorporating Dynamic Efficiency Concerns in Merger Analysis: The Use of Innovation Markets*, 63 ANTITRUST L. J. 569, 575 (1995) (suggesting a "causal connection between market concentration and the pace of technological innovation"); *but see* Robert Hoerner, *Innovation Markets: New Wine in Old Bottles*, 64 ANTITRUST L.J. 49 (1995).

quality and *kind* of expression produced.[121] However, some scholars have hinted at a recognition of this problem by applying environmental economics as a way of balancing the competing roles of information in the market. Rather than *maximizing* resource yield, environmental law frequently focuses on *optimizing* it in order to preserve a robust natural ecosystem.[122] Similar insights are beginning to inform a new movement of "cultural environmentalism" aimed at improving the quality, diversity, and organization of copyrighted expression.

B. An Ecology of Expression

Environmental economists try to balance the commodification of environmental resources with their importance as a general "background" in which all other activity (including market exchange) takes place.[123] The valuation of ecosystem services helps quantify the monetary value of this "background" role by estimating the full costs of pollution.[124] For example, the

[121] Here copyright may find some guidance from patent law; see, e.g., Lemley and Lichtman, *Gold-Plated Patents* (proposing a method of identifying and legally valorizing particularly useful patents); BEN KLEMENS, MATH YOU CAN'T USE (suggesting that many pharmaceutical patents ought to be given better treatment than many software patents, since the former are often based on far more costly research.).

[122] *See*, e.g., H. Gary Knight, *International Fisheries Management: A Background Paper*, in THE FUTURE OF INTERNATIONAL FISHERIES MANAGEMENT 1, 41 n.25 (H. Gary Knight ed., 1975), at 2, 23 ("Optimum sustainable yield was established as the international management criterion for fisheries in the Convention on Fishing and Conservation of the Living Resources of the High Seas.").

[123] HERMAN E. DALY, STEADY-STATE ECONOMICS 89-90 (1991) ("Benefits and costs that do not register themselves as conscious short-run pleasure or pain at an individual level but that are organic, with interdependencies far exceeding market relationships, must be dealt with outside the market and must result in constraints on the market.").

[124] EBAN S. GOODSTEIN, ECONOMICS AND THE ENVIRONMENT 32 (1999) ("Economists define 'pollution' as a negative externality: a

"harvest" of many natural resources not only results in a finished product (such as a fish to eat), but also short-circuits that resource's capacity to yield more in the future (i.e., the eaten fish is not going to spawn).[125] Environmental economists have given us a much more sophisticated understanding of the trade-offs between commodification and preservation regimes.[126]

There have been many instructive translations of these ideas from the physical to the virtual realm. For example, Brett Frischmann emphasizes the analogies between information goods and physical infrastructures that enable markets to function.[127] Peter Yu's[128] and James Boyle's work similarly elevates open access to certain types of information as a *sine qua non* for successful markets.[129] Elinor Ostrom has applied her landmark work on physical commons to the information commons.[130] These scholars argue that information goods, like

---

human-made, unbargained for, negative element of the environment. Pollution is termed an externality because it imposes costs on people who are 'external' to the transaction between the producer and consumer of the polluting product.").

[125] *Id.*, at 41-2 (observing that "free-market forces do not provide the right incentives to insure that adequate precautions are taken to protect our environment.").

[126] Knight, *International Fisheries Management*, *supra* note 122, at 23.

[127] Brett Frischmann, *An Economic Theory of Infrastructure*, *supra* note 17, at .

[128] Peter K. Yu, *Intellectual Property and the Information Ecosystem*, *supra* note 17, at .

[129] Boyle, *Environmentalism for the Net*, *supra* note 15. A "Conference on Cultural Environmentalism" at Stanford is set to revisit Boyle's contribution in March, 2006. *Cultural Environmentalism at 10*, *available at* http://cyberlaw.stanford.edu/conferences/cultural/ (last visited Feb. 25, 2006) ("host[ing] a symposium to explore the development and expansion of the metaphor of 'cultural environmentalism' over the course of ten busy years for intellectual property law.")

[130] Charlotte Hess and Elinor Ostrom, *Ideas, Artifacts, and Facilities: Information as a Common Pool Resource*, 66 LAW & CONTEMP. PROBS. 111, 112 (2003) ("summariz[ing] the lessons learned from a large body of international, interdisciplinary research

wetlands or forests, tend to produce many positive externalities undervalued by markets.[131]

The new scholarship of cultural environmentalism sets the groundwork for a better understanding of the *costs*, as well as the *benefits*, of copyrighted expression. Any particular consumer good is both a finished product, potentially useful to some consumer, and potential trash, distracting the attention of shoppers from what they *really* want or need, or cluttering landfills of the future.[132] Similarly, any bit of expression that *signals* something to one who wants exposure to it may constitute *noise* to thousands of others. It is as foolish to hope for the maximization of copyrighted expression as it is to measure our welfare merely with reference to the amount of stuff we have.[133]

---

on common-pool resources in the past twenty-five years and consider its usefulness in the analysis of scholarly information as a resource.").

[131] *See* Mark Lemley and Brett Frischmann, *Spillovers*, forthcoming, Columbia L. Rev. (discussing positive externalities) (currently available on SSRN); Frank Pasquale, *Toward an Ecology of Intellectual Property*, *supra* note 17 (arguing that the positive complementary, network, and long-range effects of new technologies on the value of copyrighted works parallel the indirect, direct, and option values of biodiversity recently recognized by environmental economists). For a definition of positive externalities, see William B. Rubenstein, *A Positive Externalities Theory of the Small Claims Class Action*, 74 UMKC L. REV. 709 (2006) (defining an externality as "a cost or benefit that the voluntary actions of one or more people imposes or confers on a third party or parties without their consent").

[132] JAMES R. KAHN, THE ECONOMIC APPROACH TO ENVIRONMENTAL AND NATURAL RESOURCES 251, 254 (1995) (demonstrating that "we are generating an inefficiently high level of waste" because "the private cost of social waste does not equal its social cost."); but see Eric Goldman, A Coasean Approach to Marketing, forthcoming, Wisconsin L. Rev. 2006. My response to Goldman is posted as a commentary in the Conglomerate Blog Junior Scholars' Workshop (July 5, 2006).

[133] *See Grossly Distorted Picture: It's High Time that Economists Looked at More than Just GDP*, THE ECONOMIST, Feb. 9, 2005, *available at* http://www.economist.com/finance/PrinterFriendly.cfm?story_id=5504103 ("GDP . . . was never intended to be the definitive yardstick of

After developing the pollution analogy further, I make the case for considering information overload as an *externality* below.. Characterizing information overload as an externality advances legal recognition of the benefits of fair use for categorizers in two ways. First, the externality approach denies copyrightholders the rhetorical high ground cultivated by victims of eminent domain.[134] Fair use is less an easement or taking forced on an innocent to facilitate some utopian ideal of information dissemination, than a natural way of "cleaning up" the mess of expression so successfully encouraged by copyright law.[135] Second, the externality approach reframes *economically* and *neutrally* a perspective often consigned to the dustbin of *cultural* and *ideological* critique. Dominant IP policymakers may be frightened of a "Free Culture" movement, but can't shun the very economic methods originally used to justify strong copyright protection once they render more nuanced policy recommendations.

1. Information Overload as Externality

---

economic welfare [and] is not even the best gauge of the monetary aspects of living standards.").

[134] This rhetorical high ground is very effective in current efforts to limit eminent domain powers. *See* John M. Broder, *States Curbing Right to Seize Homes*, N.Y. TIMES, Feb. 21, 2006, at 1 ("In a rare display of unanimity that cuts across partisan and geographic lines, lawmakers in virtually every statehouse across the country are advancing bills and constitutional amendments to limit use of the government's power of eminent domain to seize private property for economic development purposes."); NYT article of July 30, 2006. For a comparison of fair use and takings doctrine, see Michael Carrier, *supra* note 84, at 3.

[135] Gregory Duhl provides a good comparison of fair use and eminent domain that confirms these points. Gregory M. Duhl, *Old Lyrics, Knock-Off Videos, and Copycat Comic Books: The Fourth Fair Use Factor in U.S. Copyright Law*, 54 SYRACUSE L. REV. 665, 730 (2004) (arguing that "courts [should] account for a copyright holder's lost licensing royalties in its analysis of market effect only if a primary or derivative market for licensing the original work exists, and only if the copyright holder is willing and able to exploit that market.").

Nuisance claims, anti-pollution laws, and taxation are all methods of "internalizing" the cost of harms like pollution to their creators. In his 1997 book *Data Smog*, David Shenk noticed that many producers of information were, like producers of physical goods, causing an externality: a miasmic haze of "information overload" that threatened to obscure the truth of any given matter in a cacophony of conflicting claims.[136] As early as 1991 James Boyle had noted the trend on a larger scale, calling for some "future Pigou [to] write an analysis of the . . . 'information pollution' we were creating . . . [because] our economics did not force us to internalize the consequences of our overproduction."[137] Answering Boyle's call, this article recognizes information overload externalities and proposes methods for courts to adjust copyright doctrine in order to help internalize them.[138]

A skeptic might concede that information overload is an annoying or troubling aspect of modern culture, but still question its characterization as an *externality*. A narrow economic approach may reserve the concept of externality for more concrete or quantifiable harms, dismissing an "overload externality" as excessively *subjective* or *normative*. On the other end of the methodological spectrum, humanists may find the *positivism* implicit in externality analysis suspect, insisting instead that the critiques of overload relate to the *kind of*

---

[136] DAVID SHENK, DATA SMOG, *supra* note 134, at 11. (condensing a variety of social scientific observations into thirteen "laws of data smog").

[137] James Boyle, *A Theory of Law and Information: Copyright, Spleens, Blackmail and Insider Trading*, 80 CALIF. L. REV. 1413, 1537 (1991).

[138] There was an early recognition of the problem of overload in Ira S. Nathenson, *Internet Infoglut and Invisible Ink: Spamdexing Search Engines with Meta Tags*, 12 HARV. J.L. & TECH. 43 (1998) (recognizing a "fundamental tension . . . among the public, webmasters, and those with proprietary rights, because each seeks to externalize its costs of information dissemination and retrieval."). Nathenson's article, along with many later pieces on spam, recognized the harms that could be caused by excess *marketing of* works. This articles focuses on information overload attributable to works themselves.

society we hope to create, rather than *harms to individuals* we hope to minimize. Though both the positivist individualism of orthodox economics and the normative holism of cultural studies might appear to pose obstacles to an externality-based understanding of information overload, they ultimately illuminate externality concept's potential to unite these two ways of thinking.[139] An economics without an understanding of how individualized transactions affect society as a whole is blind; but communitarian social critique left untranslated into the economic language of policy science is lame.[140] As empirical methods begin to complement (or even displace) mathematical modeling in economics,[141] the field is becoming increasingly relevant as it models competition and value-creation in specific realms of human experience.[142] Just as the

[139] I use the term "cultural studies" broadly to refer to all aspects of sociology, history, anthropology, political science, and social science generally which are not committed to the methodological individualism of economics. For more on the distinction between methodological individualism and holism, *see* Ernest Gellner, *Holism versus Individualism*, in READINGS IN THE PHILOSOPHY OF THE SOCIAL SCIENCES 254 (May Brodbeck, ed., 1968).

[140] This formulation follows Einstein's famous observation about science and religion ("Religion without science is blind; science without religion is lame."); see http://quoteworld.org/quotes/4186

[141] Richard A. Posner, *Rational Choice, Behavioral Economics, and the Law*, 50 STAN. L. REV. 1551 (1998); Tanina Rostain, *Educating Homo Economicus: Cautionary Notes on the New Behavioral Law and Economics Movement*, 34 LAW & SOC'Y REV. 973 (2000); Martha C. Nussbaum, *Flawed Foundations: The Philosophical Critique of (a Particular Type of) Economics*, 64 U. CHI. L. REV. 1197, 1199 (1997). These trends suggest that small communities of expertise are driving progress in economics, and the social sciences generally. *See* IAN SHAPIRO, THE FLIGHT FROM REALITY IN THE HUMAN SCIENCES 15 (2005) (encouraging social scientists to "confront the complexities inherent in the relational logics of . . . ideals.").

[142] *See* David Colander, et al., *Preface*, in THE CHANGING FACE OF ECONOMICS: CONVERSATIONS WITH CUTTING EDGE ECONOMISTS 18 (David Colander, et al., eds., 2004) (noting that "experimental economics is changing the way economists think about empirical

fields of chemistry, biology, and physics arose from "natural philosophy,"[143] today the fields of information economics, labor economics, behavioral economics, and health economics are developing in response to shortcomings in conventional microeconomic theory.[144] Two of the most important new challenges to conventional microeconomic theory are a) a growing recognition of the degree to which individual consumption decisions influence others' capacity to consume and produce[145] and b) the development of hermeneutical economics based less on quantification and modeling than on sophisticated interpretations of the *meaning* of economic exchanges for those participating in them.[146]

---

work" and "ecological economics is redefining how nature and the economy are viewed as interrelating").

[143] "[Natural philosophy is] the study of nature and the physical universe before the advent of modern science." *Natural Philosophy*, *available at* http://www.answers.com/topic/natural-philosophy; *see also* C.J. DIJKSTERHUIS, THE MECHANIZATION OF THE WORLD PICTURE 3 (1986); JOHN DUPRE, THE DISORDER OF THINGS: METAPHYSICAL FOUNDATIONS OF THE DISUNITY OF SCIENCE 7 (1995) (criticizing reductionism as philosophically naïve because "the dream of an ultimate and unified science is a mere pipe dream"); BERNARD PULLMAN, THE HISTORY OF THE ATOM IN WESTERN THOUGHT 12 (2001) (describing persistent lure of reductionist thought).

[144] For example, there are now distinctions between mainstream, orthodox, and heterodox economists. See David Colander, et al., *Preface*, *supra* note 142, at 8-9. Anita Bernstein has noted (with some skepticism) the resulting pluralism in law and economics scholarship. Bernstein, *Whatever Happened to Law and Economics*, 64 U. MD. L. REV. 303, 307-8 (2005) ("Stripped of its distinctive intellectual features, no longer able to give descriptions or policy recommendations that could not have come from sources outside the movement, law and economics now functions mainly as a faculty club with opaque, arbitrary criteria for membership.").

[145] *See* JON ELSTER, STRONG FEELINGS: EMOTION, ADDICTION, AND HUMAN BEHAVIOR 106 (2006) (discussing "emotions as the object of social norms").

[146] See, e.g., Gavin Wright, Economic History as a Cure for Economics, in SCHOOLS OF THOUGHT: TWENTY-FIVE YEARS OF INTERPRETIVE SOCIAL SCIENCE 41, 42 (Joan Scoitt and Debra Klein,

Consider the role of millions of pounds of used computers dumped into landfills in lesser developed countries over the past decade.[147] The computers contain various heavy metals and contaminants that frequently pollute groundwater.[148] Degradation of natural resources via pollution is a classic example of a negative externality: a concrete harm (the dirt or damage caused by pollution) is being inflicted by one group on another without compensation. [149] To be sure, the burdens of information overload are less concrete than those of pollution.[150] Yet the harms information overload imposes, simply in terms of making sought-after information harder to find, are real:

---

eds., 2001); ECONOMICS AND HERMENEUTICS; Tyler Cowen, Good and Plenty; Deirdre McCloskey, If You're So Smart; Oz Shy, The Economics of Networks (section discussing modeling as secondary to narratives about economic reality).

[147] ELIZABETH GROSSMAN, HIGH TECH TRASH: DIGITAL DEVICES, HIDDEN TOXICS, AND HUMAN HEALTH (2006); Jim Puckett, *High-Tech's Dirty Little Secret: The Economics and Ethics of the Electronic Waste Trade*, in CHALLENGING THE CHIP: LABOR RIGHTS AND ENVIRONMENTAL JUSTICE IN THE GLOBAL ELECTRONICS INDUSTRY 225, 227 (Ted Smith, et al., eds., 2006) ("Free trade became a mechanism that allowed [hazardious materials] to be shunted to unsuspecting, disempowered communities . . . ."); Basel Action Network, *High-Tech Toxic Trash Exported to Africa*, *available at* http://www.minesandcommunities.org/Action/press772.htm ("[A] report entitled *The Digital Dump: Exporting High-Tech Re-use and Abuse to Africa*, exposes the ugly underbelly of what is thought to be an escalating global trade in toxic, obsolete, discarded computers and other e-scrap collected in North America and Europe and sent to developing countries by waste brokers and so-called recyclers.").

[148] *Id.*

[149] ROBERT D. COOTER, THE STRATEGIC CONSTITUTION 109 (2000).

[150] But note that even the latter are usually considered probabilistic harms, expressed as some odds that one exposed to pollution will develop a disease in response. See Lisa Heinzerling, *The Rights of Statistical People*, 24 HARV. ENVT'L L. REV. 189 (2000) (discussing probabilistic harms and cost-benefit analysis).

> [I]n a Babel of signals, we must listen to a great deal of chatter to hear one bit of information we really want. We discover that information can become noiselike when it is irrelevant or interferes with desired signals, so tending to defeat meaning--making it harder to extract meaning from information, just as it is hard to extract metal from low-grade ore.[151]

Or, one might say, clean water from increasingly polluted aquifers.[152] Given the directness of the analogy, it should not be surprising that mainstream economic theorists recognize the costs of information excess. For example, Landes and Posner note that "the demand for copies of a given work depends not only on the number of copies but also on the number of competing works. The more there are, the lower will be the demand for any given work."[153] Expression does not exist in a

---

[151] ORRIN KLAPP, OVERLOAD AND BOREDOM (1986)at 2. See also BART KOSKO, NOISE (2006) (discussing the prevalence and harms of noise). Low-income internet users are probably the worst affected by overload externalities. Commenting on the spam problem, Ray Everett-Church notes that "[W]hereas major corporations can afford to fight these cutting-edge cyberlaw battles, small mom-and-pop ISPs and their customers are left to suffer the floods. The harm inflicted is in many respects analogous to the effects on society from something like pollution. For example, it would be far cheaper for chemical manufacturers to dump their waste into rivers and lakes. However, those externalities--as economists call them--allow one party to profit at another's--or everyone's--expense." Ray Everett-Church, *Why Spam is a Problem*, E-OTI, at http://www.isoc.org/oti/articles/0599/everett.html (last visited Feb. 20, 2006).

[152]

[153] LANDES AND POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW, *supra* note 18. Posner has also applied a broad understanding of externalities in cultural disputes. *See Posner-Becker Blog*, July 24, 2005, *available at* http://www.becker-posner-blog.com/archives/2005/07/ ("Economics focuses on the consequences of social action. . . . [T]here is no difference from an economic standpoint between physical and emotional harm; either one lowers the utility of the harmed person.").

vacuum, but rather influences and is influenced by previous and expected expression.

A skeptic might object that these mutual influences are too various or tenuous to be identified scientifically. However, as Donald Herzog observes, "the criteria economists actually use to identify externalities . . . come from moral and political theory, not their own views about utility maximization."[154] This is a long and accepted practice in economics, permitting serious consideration of, say, smoking or jackhammer noise as a nuisance, but not unusually colorful clothing or disagreeable manners. Economists developed the concept of "externality" when it became apparent that many aspects of the production of physical goods either used or despoiled resources that the goods' producers never paid for.

Despite the inevitably normative dimension of externality labeling and measurement in economics, it is still widely accepted as an integral part of a *neutral and universalistic* language of policy justification. This is important because First Amendment concerns for content-neutrality may scuttle "first-order" efforts to deal with information overload directly by making certain types of expression more costly.[155]

But see Richard Epstein, *Externalities Everywhere? Morals and the Police Power,* 21 HARV. J. L. & PUB. POL'Y 61 (1997) (warning against overly expansive conception of externalities).

[154] Don Herzog, *Externalities and Other Parasites*, 67 U. MICH. L. REV. 895 (2000) (noting that "economists are opportunistic about invoking externalities[: t]hey do so not whenever we find people with preferences about others' preferences and actions; they actually do so in ways closely tracking the traditional harm principle of liberal theory.").

[155] *See, e.g.*, Jack Balkin, *Populism and Progressivism as Constitutional Categories*, 104 YALE L.J. 1935 (1995) (critiquing Cass Sunstein's constitutional theory for positing a "division between high and low culture -- with the former seen as essential to the repair of the system of democratic deliberation and the latter identified with mass culture and particularly with the culture of television."). Though courts have given their imprimatur to a number of copyright policies that have suppressed expression, they do so largely because the copyright laws are facially neutral with respect to different kinds of expression.

Though courts have accepted many copyright enforcement actions that suppressed expression, they are unlikely to countenance an effort to use copyright policy to reduce trivial, untrue, or immoral expression.[156] However, understanding information overload as an *externality* arising from information production could lead to content-neutral policy prescriptions acceptable to all. This understanding is focused merely on the amount, and not the kind, of information produced, leading to "second-order" solutions to overload externalities.

2. Reducing Search Costs: Copyright at the School of Trademark

We can further familiarize the idea of modeling information overload as an externality by turning to an area of intellectual property law well-acquainted with the problem of search costs: trademark law.[157] In the blooming, buzzing

[156] A first-order response to the problem—reducing legal incentives to create information—would require one to "take sides" in the debate between the copyleft and IP expansionists mentioned above. Ironically, it would tend to support the policy prescriptions of the loser (i.e., if we finally decided that copyright expansionists had proven that privatization is the best method for maximizing information, we'd be inclined to endorse more forms of public access—and vice versa). I am not prepared to take sides on this "first order" question here; suffice it to say, a great deal more empirical research is necessary to such a determination, which would likely only be valid for particular *sectors* of the information economy, and not for copyrightable expression as a whole. For more on sectoral analysis in copyright, see Michael Carroll, *One for All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM. U. L. REV. 845, 849 (2006) (claiming that "uniformity cost is the central problem that intellectual property law must manage," recognizing that "uniformity cost rises with the growing economic importance of, and variation among, information-centric industries," and recommending copyright law sensitive to the unique competitive environments in different industries).

[157] WILLIAM LANDES AND RICHARD POSNER, THE LAW AND ECONOMICS OF INTELLECTUAL PROPERTY LAW 168 (2003) ("The value of a trademark to the firm that uses it to designate its brand is

confusion of modern markets, trademarks are source-designators,[158] providing consumers with easily apprehended symbols indicating the provenance of the goods and services they purchase.[159] Leading articles on trademark law suggest

---

the saving in consumers' search costs made possible by the information that the trademark conveys or embodies about the quality of the firm's brand."); Scandia Down Corp. v. Euroquilt, Inc., 772 F.2d 1423 (7th Cir. 1985) ("Trademarks help consumers to select goods. By identifying the source of the goods, they convey valuable information to consumers at lower costs. Easily identified trademarks reduce the costs consumers incur in searching for what they desire, and the lower the costs of search the more competitive the market. A trademark also may induce the supplier of goods to make higher quality products and to adhere to a consistent level of quality") (cited in Dinwoodie & Janis, *Confusion Over Use: Contextualism in Trademark Law*, forthcoming, 2007).

[158] GRAEME B. DINWOODIE AND MARK D. JANIS, TRADEMARKS AND UNFAIR COMPETITION: LAW & POLICY 13 (2004) ("Historically, in the United States, trademark protection has existed as an important aspect of a larger body of law, namely unfair competition law: It is unfair competition to pass off your goods as those of another producer by using a trademark confusingly similar to that of the other producer."). *Accord*, *J. Diane Prods.* v. *Swan Cosmetic Research Labs., Inc.*, 1985 U.S. Dist. LEXIS 17173 (N.D. Ill. 1985).

[159] As Barnes explains,

> According to conventional theory, when mark holders invest in trademarks, they create goodwill that leads to greater profits from sales of their goods and services. Mark holders invest in trademarks through maintaining consistent quality and characteristics of their products and services and advertising that informs consumers about their products qualities or, at least, informs consumers about their products' availability. Consumers benefit because they can rely on familiar marks to locate satisfactory goods and services. Competition is enhanced because new entrants to a market can also invest in trademarks that will attract new customers to the qualities and characteristics of their own goods. Competition lowers prices. Everyone benefits. Investing in and protecting the signals conveyed by trademarks benefits suppliers and consumers.

that its primary purpose is to reduce consumers' "search costs," to make it easier for them to find what they want.[160] For example, "Starbucks" as a mark for cafés helps consumers navigate the marketplace to find (or avoid) a known consumption experience. Trademarks thus increase the *salience* of particular products and services, elevating them above the run of things by associating them with particular words, images, and prior experiences.[161] Like trademarks, search engine results help consumers navigate through a maze of digital data. They are "search cost reducers" par excellence, rapidly ordering a plethora of data to promote the discovery of URL's related to one's inquiry.

The degree of similarity of marks and rankings emerges most vividly in the way the prerogatives of search engines and

---

David (Jake) Barnes, *A New Economics of Trademarks* 1 n. 1, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=886045.

[160] Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 HOUSTON L. REV. 777 (2004); *see also* Stacey L. Dogan & M ark A. Lemley, *A Search Costs Theory of Trademark Defenses*, in TRADEMARK LAW AND THEORY: A HANDBOOK OF CONTEMPORARY RESEARCH (Dinwoodie and Janis eds. 2007) (forthcoming); *TY Inc.* v. *Ruth Perryman*, 306 F.3d 509, 510 (7th Cir. 2002) ("The consumer who knows at a glance whose brand he is being asked to buy knows whom to hold responsible if the brand disappoints and whose product to buy in the future if the brand pleases").

[161] Lemley & Dogan, supra n. *, at 781; *The New Kids on the Block* v. *Gannett Satellite Info. Network, Inc.*, 971 F.2d 302, 306 (9th Cir. 1992) ("In economic terms, trademarks reduce consumer search costs by informing people that trademarked products come from the same source. The benefit of the brand name is analogous to that of designating individuals by last as well as first names, so that, instead of having to say 'the Geoffrey who teaches constitutional law at the University of Chicago Law School – not the one who teaches corporations,' you can say 'Geoffrey Stone, not Geoffrey Miller".); *Union Natl. Bank of Tex., Laredo, Tex.* v. *Union Natl. Bank of Tex., Austin, Tex.*, 909 F.2d 839, 844(5th Cir. 1990) ("The idea is that trademarks are 'distinguishing' features which lower consumer search costs and encourage higher quality production by discouraging free-riders.").

trademark holders have come into conflict in recent years. For example, search engines want the right to display whatever their algorithms deem "most relevant" in response to a given mark—including links to those who compete against the mark's owner.[162] These disputes show how trademarks and search rankings are coming to enter the same cultural space—that of ordering an array of options for consumers.[163]

Trademark law has long acknowledged the type of concerns about information overload and attention scarcity that have driven the analysis of categorizers' positive externalities (and content providers' negative, overload externalities) in Sections A and B above.[164]Marks are designed to help consumers choose products and services. Marks help consumers order experiences and learn via impressions of given

---

[162] *See Google Inc. v. Am. Blind & Wallpaper Factory, Inc*., 2004 U.S. Dist. LEXIS 27601 (D. Cal. 2004) (Am. Blind & Wallpaper Factory, Inc. (ABWF) claimed that Google's advertising program infringed on their trademarks by giving links to competitor's websites when users searched for terms that were trademarked by ABWF); *See also Reed Elsevier, Inc.* v. *Innovator Corp.,* 105 F. Supp 2d 816 (D. Ohio 2000) (Plaintiff claims defendant bought trademarked terms on various search engines so that when a user searched for plaintiff's trademarked terms, the search engine would display competitor's result in front of, and in larger font than, plaintiff's); *Lockheed Martin Corp.* v. *Network Solutions,* 985 F. Supp. 949 (D. Cal. 1997) (Plaintiff complained that defendant, a registrar of Internet domain names, committed direct and contributory infringement, unfair competition, and trademark dilution when it accepted registration of Internet domain names similar or identical to plaintiff's service mark).

[163] Mark Bartholomew, *Making a Mark in the Internet Economy: A Trademark Analysis of Search Engine Advertising,* 58 OKLA. L. REV 179 (2005) (discussing the common role of search queries and trademarks in helping consumers locate goods and services); Misha Gregory Macaw, Google, Inc. v. American Blind & Wallpaper Factory, Inc.: *A Justification for the Use of Trademarks as Keywords to Trigger Paid Advertising Placements in Internet Search Engine Results,* 32 RUTGERS COMPUTER & TECH. L.J. 1 (2005).

[164] Sara Stadler Nelson, *The Wages of Ubiquity in Trademark Law*, 88 IOWA L. REV. 731 (2003).

consumption experiences. Without them, it is easy to imagine being overwhelmed by the array of options available—or simply foregoing choice altogether.

In an era of information overload, attention, not information, is the more pressingly scarce commodity.[165] Like travelers congesting a busy highway, a proliferation of goods can "get in each other's way," providing too many choices for a harried consumer to process. Marks help ease the processing demands, reducing consumers' costs of searching for the product they want.[166]

Trademark law anticipates and mitigates the problem of information overload by giving owners the right to control the marks associated with their products. Mark owners can also sue to enjoin use of marks substantially similar to their own, and

---

[165] RICHARD LANHAM, THE ECONOMICS OF ATTENTION 7 (2006) ("What then is the new scarcity that economics seeks to describe? It can only be the human attention needed to make sense of information."); Herbert A. Simon, *Designing Organizations for an Information-Rich World*, in COMPUTERS, COMMUNICATIONS, AND THE PUBLIC INTEREST 37, 40-41 (Martin Greenberger ed., 1971) ("[A] wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.") (cited in Niva Elkin-Koren, *Let the Crawlers Crawl: On Virtual Gatekeepers and the Right to Exclude Indexing*, 26 U. DAYTON L. REV. 179, 184 (2001).

[166] Imagine, for instance, a world without trademark protection, where each producer's marks were copied freely by competitors. Consumers could not easily recognize, by brand, the source of the products they were consuming. Stephen L. Carter, *The Trouble with Trademark*, 99 Yale L.J. 759, 762 (1990) ("If a firm of good reputation tried to mark its goods with a symbol to let consumers know them—that is, to lower the costs of search for consumers desiring to purchase the firm's goods—other firms could imitate the symbol and trade on the first firm's reputation."). They may well develop secondary authentication techniques, or other ways of assuring that something marked as "Coca Cola" really came from the company that supplied the last beverage they enjoyed with that label on it (such as always traveling to one "authorized distributor" who could certify the source of the beverage). But all these alternative strategies would likely increase the cost of finding what one is looking to buy.

some product packaging (and even product design) that is too similar to the "trade dress" of their products. Such broad protections serve to reduce the noise with which counterfeiters and copiers can occlude the signal of mark owners. In other words, the use of a mark as a source designator is rivalrous; to the extent unauthorized producers appropriate the mark, they reduce its potential to signal the provenance of the good it marks.[167]

At this point, the analogy between systematic protection for trademarks and for categorization, should be clearer. Both serve a similar function by generating metadata structures that help us organize experience and more easily find what we need and grasp the import of signals that promote our search. Both also respond to the problem of congestion.[168] In a world where

---

[167] See Barnes, supra n. 167, at 8 ("Simultaneous use of a trademark by consumers referring to a particular source of coffee is purely non-rivalrous and simultaneous use by competing coffee suppliers in the same geographic market is purely rivalrous. Trademark law permits unrestricted referential use. Infringement actions are directed at conflicting proprietary uses of a mark, source-indicating uses by competitors.").

[168] Congestion is problem readily apprehended in physical space, with important applications to scarce attentional resources. Todd Sandler & John T. Tschirhart, *The Economic Theory of Clubs*, 18 J. ECON. LITERATURE 1481, 1488 (1980) (focusing on the physical aspect of crowding as congestion). Indeed, scholars of copyright have begun examining "congestion externalities," though they tend to focus on whether untrammeled use of a particular work reduces its value. *Compare* William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. Chi. L. Rev. 471, 474-75 (2003) ("just as an absence of property rights in tangible property would lead to inefficiencies, so an absence of copyright protection for intangible works may lead to inefficiencies because of congestion externalities and because of impaired incentives to invest in maintaining and exploiting these works.") with Dennis S. Karjala, *Congestion Externalities and Extended Copyright Protection*, 94 Geo. L. J. 1065, 1067 (2006) (arguing that the Landes and Posner analysis on congestion externalities is "seriously flawed" because it overgeneralizes and ignores the fact that "a downward shift in the demand curve for a product does not necessarily, or even often, represent a loss in 'value' to society as opposed to a change in

a quarter gigabyte of data is produced per person, per year, what appears of far more value (than sheer information) are the filters and maps we need to make sense of this deluge.[169] As Richard Lanham says, the "new scarcity" is not that of information in general, but of "human attention needed to make sense of information."[170] Both trademarks and categorizers help ease the burden of choosing between an ever-increasing number of goods and services.

A skeptic might reply: if trademark law gives a mark's owner control over the use of the mark, why shouldn't copyright law give a copyright owner's the right to control the metadata associated with its works?[171] The flaw in reasoning here is to elide the distinction between *competing* and *referential* uses in trademark.[172] Trademark law largely prohibits the former, but provides robust protections for the latter. Similarly, categorizers seek not to use parts of

---

overall social preferences"). This article focuses on whether works in general can create "congestion externalities" by overtaxing users' ability to find what they need.

[169] The quarter-gigabyte figure is based on RICHARD LANHAM, THE ECONOMICS OF ATTENTION 258 (2006) ("Just as digital storage is the only possible means of preserving the exabytes of information (an exabyte = a billion gigabytes) being generated annually, so the competitive expression of the electronic screen is the only way to express the larger struggle for attention.").

[170] RICHARD LANHAM, THE ECONOMICS OF ATTENTION 7 (2006).

[171] *Merchant & Evans, Inc.* v. *Roosevelt Building Products Co., Inc.*, 774 F. Supp. 1467, 1474 (E.D. Pa. 1991). ("Exclusivity promotes investments in goodwill because consumers interested in quality can easily locate goods manufactured by quality producers and distinguish those of competitors. Likewise, such exclusivity reduces the search costs of consumers, thereby raising their welfare.").

[172] See Barnes, supra n. 159, at 7 ("Consumers would be confused by the simultaneous use of the STARBUCKS mark by two unrelated coffee shops on the same block. But society also benefits from unlimited referential use, to search for or refer to products from Starbucks U.S. Brands. The [referential] information contained in a trademark, like the information disclosed in a patent or revealed in a copyrighted expression, can be consumed non-rivalrously.").

Field Code Changed

copyrighted works to *substitute* for the works themselves, but to *refer* to them, like the nominative uses protected by trademark law.[173] These nominative uses are given favorable treatment at least in part because they reduce consumers' search costs—they permit those writing about or commenting on products to quickly identify the thing they are referring to. Similarly, the type of indexing and commentary provided by commentary improves the "marketplace of ideas" by increasing consumers' level of information about the expressive works they are considering buying.

Trademark law is fundamentally a species of unfair competition policy, a mode of regulating commercial relations among competing offerors of similar products. In its role as an arbiter of categorization processes, copyright law should not shy away from such a function.[174] Just as John Wiley has observed the lessons patent law can teach copyright law, some astute commentators have already described the ways in which trademark theory can inform copyright doctrine.[175] As Gregory Lastowka observes, the Visual Artists' Rights Act already protects rights of attribution (a form of source-designation) for "original works by visual artists who produce single works, limited edition prints, or sculptural casting in editions of less than 200." Lastowka proposes extending that type of protection

---

[173] The nominative use defense permits a party to use a trademark to refer to the trademark holder. For example, the New Kids on the Block once sued U.S.A. Today to prevent the newspaper from publishing a poll on "which New Kid do you like best." The Ninth Circuit held that the newspaper was only using the mark (New Kids on the Block) to refer to the band, not as an indication that the band had endorsed the poll or newspaper, or was the ultimate source of the poll. New Kids on the Block v. News America Publishing, Inc., 971 F.2d 302 (9th Cir. 1992).

[174] See Joseph Liu, *Regulatory Copyright*, N.C. L. REV. (2004) (discussing the increasingly fine-grained role copyright law has in regulating a number of business practices).

[175] Wiley, *Copyright at the School of Patent*, 58 U. CHI. L. REV. 119 (1991); *see also* Grokster v. MGM --- U.S. --- (2005) (discussing the lessons of patent for copyright contributory infringement doctrine).

to copyrighted works generally.[176] Laura Heymann asks us to "see trademark-like activity in . . . the act of authorship," as she carefully parses the types of information conveyed by metadata like names and brands.[177] Each scholar's observations provides fresh ways of understanding copyright-based regulation of categorization. Overlapping trade secret, patent, copyright, and contract laws already govern the protection of search engines' "secret sauce" of the algorithms they use to generate results. Given the centrality of search cost theory to so much of trademark law and theory, a combination of principles drawn from both copyright and trademark might best guide policymaking on the search processes that categorizers promote.

IV. OVERCOMING OVERLOAD

"Second-order" approaches[178] to overload externalities do not attempt to reduce the amount of copyrightable expression, but would rather create incentives for the production of metadata (via categorization services, evaluators, and indexers) which make the welter of extant information easier to navigate.[179] Unfortunately, would-be catalogers, archivists,

---

[176] Lastowka, *The Trademark Function of Authorship*, 85 B.U. L. REV. 1171, 1211 (2005).

[177] *The Birth of the Authornym: Authorship, Pseudonymity, and Trademark Law*, 80 NOTRE DAME L. REV. 1377 (2005).

[178] As noted above, in the introduction and note 156, a "first order" solution to the problem of information overload would involve a Pigouvian tax on expression. Given the normative undesirability of such a tax, the "second order" solutions proposed here are designed not to deter the creation of expression, but to create incentives for the type of expression (categorization) that makes all other expression more useful and navigable.

Field Code Changed

[179] Those fond of the problem of infinite regress might skeptically comment here on the potential problem of too many metadata providers. Nevertheless, for reasons that become more apparent in Part IV.A., I believe that metadata is less susceptible to the problem of information overload than the expression it organizes and categorizes. Consider, for example, rival movie review sites (such as Rotten Tomatoes (rottentomatoes.com) and the *New York Times* film review archive.). It is much easier to take in Rotten

arrangers and guides are often menaced by a thicket of potential copyright claimants. Even if most content owners appreciate the categorizers' service, holdouts tend to demand licensing fees either for the initial, archival copy used to make such compilations digitally searchable, or for results that include portions of the copy as samples. The courts are divided on the merits of such claims, and the cases often hinge on judges' ability to recognize the ways in which unauthorized arrangement and organization of copyrighted works may be essential to a thriving market in information.[180] For instance, courts have affirmed fair use in the case of internet archives of photos, but have resisted it in the case of an interactive site utilizing movie clips.[181] The resulting legal uncertainty has chilled many valuable categorization projects.

This Part suggests two legal routes for courts to begin giving proper weight to the value of categorization services in copyright law. First, categorization and indexing should join the list of especially privileged uses recognized in case law.[182] Like reviewers and educators, categorizers reduce negative externalities associated with information overload. Law should not discourage this productive activity.

Tomatoes' snap statistical summary of movie reviews (indicating that 66% of 127 total reviews of *Legally Blonde* are positive) or even the Times' snarky dismissal of the film than to watch the film itself. See Stephen Holden, *Legally Blonde*, *available at* http://movies2.nytimes.com/gst/movies/movie.html?v_id=246684 (claiming the film "turns gooey when it should be sharp") and Rotten Tomatoes, *Critics Tomatometer for Legally Blonde: 66%*, *available at* http://www.rottentomatoes.com/m/legally_blonde/).

[180] *See* Michael Madison, *A Pattern-Oriented Approach to Fair Use*, *supra* note 73, at 1560 (discussing methods of identifying such patterns, or "social practices").

[181] *See* discussion in Part IV.B. below.

[182] For an overview of how courts have treated various uses under the first factor, *see* WILLIAM F. PATRY, THE FAIR USE PRIVILEGE, *supra* note 58, at 419-504. Categorizing does not yet explicitly appear in this list, though it might be seen as a natural extension of archiving, preservation, abstracting, and research. *Id.*, at xvi.

Second, courts should begin treating aggressive litigation against categorizers as a form of copyright misuse—an illicit effort to leverage control over copyrighted works into secondary markets in which monopoly market power is inappropriate.[183] Only robust fair use and misuse defenses can adequately protect the Google Print project, and categorizers generally, from harassing litigation.

A. The Value of Categorizers

Though many commentators appear to assume that copyright law's chief purpose is to create *more* expression, its *organization* is becoming increasingly important in an era of information overload. Independent categorizers promise to play a vital role in taming information overload externalities, if copyright law permits them to archive, index, and sample copyrighted works.

The scope and creativity of categorizers is astonishing. Feel depressed and only want to read good news? Try out "Mood News," which arrays headlines in order of "good, bad, and neutral."[184] Want to read fresh new political theory and perspectives on the academy? Try meta-blog "Political Theory Daily Review," which posts dozens of recommendations each day.[185] Need to figure out which of your Beanie Babies is a collectable, and which can be safely consigned to Goodwill? A publisher has been generating several guides just for that purpose, as well as picture books permitting a broader overview of the "Beanie Baby" landscape.[186]

---

[183] 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT, *supra* note 61.

[184] Mood News, *BBC News Headlines*, at http://www.latedecember.com/sites/moodnews/index.html (last visited Feb. 20, 2006). If Agence France Press succeeds in its claim against Google, arguing that headlines are copyrightable, such a service would probably be shut down due to liability concerns.

[185] Alfredo Perez, *Political Theory Daily Review*, at http://politicaltheory.info/, (last visited Feb. 2, 2006).

[186] *See* DENISE I. O'NEAL, FOR LOVE OF BEANIE BABIES 5 (1998) (discussing the thriving beanie baby "secondary market").

Though each of these particular examples may seem hopelessly specialized or trivial, the phenomenon of categorizing as a whole is not. Consider, for example, the range of music categorizing sites. MySpace Music,[187] The N,[188] Yahoo! Music Engine,[189] NPR,[190] and Amazon.com[191] all permit users to comment on and rank musical works. In the film arena, Rotten Tomatoes arranges snippets from a cornucopia of movie reviews on pages devoted to nearly all recent films.[192] Blockbuster publicizes a list of "1001 Movies You Must See"[193] and the American Film Institute's Top 100.[194] Netflix,[195] IMDB, and YMDB all offer users the chance to rate films, comment on them, and comment on each other ratings

---

[187] MySpace.com, *MySpaceMusic*, *available at* http://topartists.myspace.com/index.cfm?fuseaction=music.topBands, (last visited Feb. 2, 2006).

[188] The N, *Music Related Rankings*, *available at* http://www.the-n.com/games/rank/index.php?topic=183&v=69170&theme=music, (last visited Feb. 2, 2006).

[189] Yahoo!, *Yahoo! Music Engine 1.1 Help*, *available at* http://help.yahoo.com/l/us/yahoo/music/yme/personalization/personalization-45785.html, (last visited Feb. 2, 2006).

[190] The Best Music of 2005 Countdown, http://www.npr.org/templates/story/story.php?storyId=5054194&sourceCode=gaw, (last visited Feb. 6, 2006); Amazon.com: Rating Items, http://www.amazon.com/exec/obidos/tg/browse/-/13316081/104-9725054-4452710#rate, (last visited Feb. 6, 2006)

[191] Amazon.com: Recommendations Explained, http://www.amazon.com/exec/obidos/tg/browse/-/13316081/ref=br_bx_c_1_6/104-9725054-4452710, (last visited Feb. 6, 2006).

[192] Rotten Tomatoes, *Review Selection*, *available at*: http://www.rottentomatoes.com/pages/faq#gathering

[193] *1001 Movies You Must See*, *available at* http://www.blockbuster.com/homepages/displayPage.action?channel1=1001&nav=true&subChannel=sub&cctr=DVDCollections, (last visited Feb. 6, 2006).

[194] Tomatometer Rating System, *available at* http://www.rottentomatoes.com/pages/faq#tomatometer, (last visited Feb. 6, 2006).

[195] Netflix Recommendations, http://www.netflix.com/Recs, (last visited Feb. 6, 2006).

and comments.[196] Both Amazon.com and Powells.com offer similar services for books.

Given the breadth of such current categorizing services, it is reasonable to wonder whether any legal intervention is necessary to help this field at all. Isn't it just as Robert Merges predicted back in the 1990's—that private parties are working out deals to best promote and expose their content?[197] Admittedly, some large corporations have successfully brokered deals with content owners to set up robust categorizing sites that feature bits of the content presented. For example, Amazon's power as a retailer allowed it to leverage "look inside the book" from publishers.[198] But when we look at the fate of some smaller players, the limitations of a laissez-faire approach become clear.

The Chilling Effects Clearinghouse has documented manycases of content owners bullying categorizers and commenters with cease and desist letters.[199] The high cost of litigation deters many categorizers from even trying to assert their fair use claims. As Lawrence Lessig has said, given the uncertain state of the law, fair use is often little more than "the right to hire a lawyer."[200] Thus content owners may use

---

[196] YMDB: Your Movie Database, *Ratings*, *available at* http://www.ymdb.com/faq/index_ukuk.html, (last visited Feb. 6, 2006).

[197] Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293, 1296 (1996) (praising collecting societies like ASCAP and BMI as models of private ordering).

[198] *See* Gary Wolf, *The Great Library of Amazonia*, *supra* note 91, at 86.

[199] MARJORIE HEINS AND TRICIA BECKLES, WILL FAIR USE SURVIVE? FREE EXPRESSION IN AN AGE OF COPYRIGHT CONTROL, (2005) 38-45 (discussing several instances where copyrightholders' legal threats effectively vetoed apparent fair uses).

[200] See Larry Lessig, Talkback: Manes, Lessig Blog, Mar. 20, 2004, available at http://www.lessig.org/blog/archives/001794.shtml (noting that, given the indeterminacy of extant fair use doctrine, "'Fair use' in America is the right to hire a lawyer."); *see also* David Nimmer, *Fairest of them All*, 66 LAW & CONTEMP. PROBS. 263, 281 (Winter/Spring 2003) ("Basically, had Congress legislated a

expansive rights over derivative works to assure that only "approved" partners have "full-service" rating and comment sites. For example, to assure comity between all players, Amazon screens reviews and does not permit exceptionally cutting or nasty criticism.[201] Ty, Inc., owner of the Beanie Babies trademark and copyrights in these "sculptural works," has systematically tried to suppress criticism of its products and business practices.[202] And as this article has already documented, Google's legal troubles are legion, in part because it refuses to "play by the rules" set by content owners.[203]

Ironically, Google itself may well be hurt in the long run if it manages to succeed in its fair use defense against publishers, the Author's Guild, and Agence France Press. To the extent that these cases establish a precedent of license-free sampling, they permit lower-cost entry for competitors in the search market—as well as for categorizers generally. In a world in which categorizers need licenses for all the content they sample, only the wealthiest and most established entities will be able to get the permissions necessary to run a categorizing site.[204] Fair use for snippets of books, thumbnails of images, and samples of audiovisual and musical works levels the playing field.

---

dartboard rather than the particular four fair use factors embodied in the Copyright Act, it appears that the upshot would be the same.").

[201] For documentation of this phenomenon in the case of one book, *see Battle of the Bailey Reviewers*, *available at* http://www.tsroadmap.com/info/bailey-reviews.html (last visited Feb. 28, 2006).

[202] *See* discussion in Part IV.C., below.

[203] See *Complaint*, *supra* note 49.

[204] For a fuller exploration of the possibility of entrenched concentration here, see Frank Pasquale, *Rankings, Reductionism, and Responsibility*, 54 CLEV. ST. L. REV. 115, 129-130 ("[W]e cannot expect . . . a competitive market to arise organically . . . [I]ndividuals . . . seek out the most comprehensive and authoritative source of information, and the very fact that this occurs gives the leading source enormous leverage to assure that information sources will want to appear (and be highly ranked) on its search results. . . . The "rich get richer" [dynamic can] mak[e] the search and rankings field a very difficult one to enter.).

Only a diverse and independent field of categorizing sites can fully realize their promise of better mapping the information environment. Categorizers help society overcome the fragmentation and colonization of the lifeworld, terms coined by German social theorist Jurgen Habermas to designate the negative consequences of increasing specialization and inaccessibility of knowledge.[205] Fine art and music will tend to become ever more disconnected from daily life if a robust community of critics and commentators is unable to relate them to those outside the often insular community of tastemakers. Popular music improves as niches of thoughtful and independent commenters evaluate and share the artists they enjoy.

In politics, the growing trend toward "narrowcasting" and partisan media erodes the common public sphere of knowledge upon which democratic dialogue depends.[206] As narrowcasting replaces broadcasting, news aggregators like Google may well be the only news source that adequately reflect the full range of opinions on what constitutes news. Finally, categorizers "level" the information playing field, letting outsiders understand the full range of expression available. As proprietary information grows in importance, citizens deserve at least a right to know *what is available*, even if its price makes the expression in question unaffordable.

---

[205] JURGEN HABERMAS, II THEORY OF COMMUNICATIVE ACTION 364 (T. McCarthy, trans., 1989) (describing "the disintegration of life-relations when these are separated, through legalized social intervention, from the consensual mechanisms that coordinate action and are transferred over to [delinguistified steering] media such as power and money.").

[206] See, e.g., Pam Karlan and Eben Moglen, *The Soul of A New Political Machine*, 34 Loyola of Los Angeles L. Rev. 1089 1106 (2001) ("Broadcasting declines, narrowcasting is in."); Michael S. Kang, *From Broadcasting to Narrowcasting: The Emerging Challenge for Campaign Finance Law*, 73 GEO. WASH. L. REV. 1070 (2005); CASS SUNSTEIN, REPUBLIC.COM (describing potential negative effects of narrowcasting); *but see* Dan Hunter, *Phillipic.com*, 90 Cal. L. Rev. 611, 663-65 (2005). (reviewing Sunstein and suggesting that he overstates the negative filtering effects of narrowcasting).

B. The Current Circuit Split on Categorizers

Despite their great promise, categorizers have suffered uneven treatment from courts. The circuits are split on the issue of web archives and categorizations, finding certain types of collections clear fair uses and others infringing.

In *Kelly v. Arriba Soft*, Arriba's search engine, now located at www.ditto.com, permitted Internet users to find images by searching its archives.[207] Kelly, a nature photographer, sued Arriba Soft for including his images in its archive.[208] Arriba's website provided two services: 1) lists of "thumbnail" visions of the images (reduced in size and thus quality) and 2) framing of the full-size image (which appeared on Arriba's website exactly as it had on its source page). The Ninth Circuit ruled that the first use was fair, largely on the basis of its "effect on the market" analysis.[209]

The panel recognized that the plaintiff's images "are related to several potential markets," including attracting internet users to Kelly's own website (which sold digital and print versions of the images and other materials), and being sold or licensed to other websites or to a "stock database."[210] Observing that Arriba's thumbnail images actually directed users to Kelly's site, the panel found no evidence that it reduced the value of his images as a type of advertising for his site.[211]

---

[207] Kelly v. Arriba Soft, 336 F.3d 811 (9th Cir. 2003).

[208] For example, if one goes to Arriba Soft's website (ditto.com) and types in "dog," the site provides at least 12 "thumbnail" images of dogs, permitting the user to click on the source of each image and thereby be directed to the website on which the dog image appears.

[209] It remanded the latter issue with instructions to the district court. *Id.*

[210] Id., at 818.

[211] The panel's finding on the first factor (the purpose and character of the use) informed its effect on the market analysis: "Arriba was neither using Kelly's images to directly promote its website nor trying to profit by selling Kelly's images. Instead, Kelly's images were among thousands of images in Arriba's search engine database. Because the use of Kelly's images was not highly exploitative, the commercial nature of the use only slightly weighs against a finding of fair use." *Id.*

The court also found that the "low resolution" thumbnails in no way competed with the full size images in markets for images.[212] However, since the full-size images Arriba made available did divert internet users from Kelly's website, and effectively substituted for the images Kelly would have sold, the panel was agnostic on the fairness of this use and ordered the district court to consider more closely the economic effects of this type of reproduction.[213] The *Kelly* panel's opinion offers a model of "fourth factor" analysis that recognizes the complexity of economic effects of unauthorized use.[214]

However, fourth factor analysis not only "giveth" to categorizers, but also "taketh away." In *Video Pipeline v. Buena Vista Entertainment*, a company specializing in the business of movie preview compilation and organization sold clips of movies, without permission from the movie copyright holders, to retailers for use on their websites.[215] Users could not

---

[212] *Id.* Google is expected to rest a good deal of its fair use defense on an analogy of snippets to thumbnails. Thumbnails are to pictures what snippets of text are to books (sentences are to books). Books cannot be modified (useful yet not as valuable to user) by shrinking the size of text (as thumbnails are just pictures reduced in size), but instead are reduced in form (length) by only allowing the user/searcher to see a small applicable portion.

[213] The panel addressed the diverse markets for the photos involved. *Id.*, at 819.

[214] However, the vitality of that precedent was recently called into question when a district court judge found Google liable for providing almost exactly the same image-search service at issue in *Kelly*. The only significant difference in the two cases was that Google's antagonist, a purveyor of erotic images, could demonstrate that it had licensed small-scale reproductions of its images to a cell-phone company. *See* Xenia Kobylarz, *Perfect 10 Racks Up Preliminary Injunction Against Google*, LAW.COM, Feb. 22, 2006, *available at* http://www.law.com/jsp/article.jsp?id=1140516320952; Siva Vaidhyanathan, *Thumbnails (and Google) in Danger*, SIVACRACY.NET, Feb. 25, 2006 (noting that Google may be "inviting its own death" if this case "forc[es] the courts to overturn *Kelly [v. Arriba Soft*.]").

[215] Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc., 192 F. Supp. 2d 321, 342-43 (D.N.J. 2002), aff'd, 342 F.3d 191 (3rd Cir. 2003).

download the clips, but each time a user viewed a clip on a retailer's website, the retailer paid a fee to the movie preview company.[216] The copyright holders of the movies claimed that the use of the clips constituted copyright infringement.[217]

The district court sensitively addressed the "effect on the market" factor accounting for both potential negative as well as positive effects resulting from the unauthorized distribution of the clips.[218] In addition, the trial judge found that the movie clips were not substitutes for the copyrighted films[219] and recognized that the contested site would increase exposure to the work.[220] Visitors to retailers' websites, "who might otherwise be unaware of, or unattracted to" the films, would have a chance to view clips.[221] These determinations left the district court unconvinced by the plaintiffs' assertions that the Video Pipeline service reduced the value of their copyrighted works.

Nevertheless, the Third Circuit found in favor of Buena Vista (a Disney subsidiary) by restricting the scope of the fourth factor inquiry: "Because the issues pertaining to the potential harm to the market for Disney's derivative trailers are more straightforward, we focus our analysis on this area and do not

---

[216] *Id.*

[217] *Id.*

[218] There exists "the possibility that potential customers will be discouraged from purchasing or renting certain videos due to the depiction of the movie as provided by Video Pipeline's clip previews . . . [and,] [m]oreover, the evidence that Video Pipeline's video previews are low in quality . . . also suggests that the market for purchasing or renting the copyrighted motion pictures may be detrimentally affected." *Id.* at 340. The district court also concluded that "Video Pipeline's service of providing online previews to retailers' customers may also affect the marketability of the copyrighted motion pictures due to the retailers' competition with . . . [the copyright holder] in online sales." *Id.*, at 341.

[219] *Id.*

[220] *Id.* ("While Video Pipeline's previews may attract customers to its retailers' websites and lead to increased purchasing, as they submit, such purchases would most likely detract from the sales of home videos on [the plaintiff's] official website.").

[221] *Id.*

review the District Court's" consideration of the site's effect on the value of the underlying films.[222] The appellate court found (rather unsurprisingly) that Video Pipeline's unauthorized use of the trailers denied the plaintiffs the right to charge for that content.[223] The appellate panel did not even consider whether potential positive effects on sales or rentals of the underlying movies might swamp these negative effects.[224]

C. Directions for the Future

Arriba Soft and *Video Pipeline* create a circuit split on the proper analysis of categorizers in a fair use "effect on the market" analysis. In previous work, I have focused on resolving this split by refining the fourth factor of the fair use test, proposing ways of making the requisite judicial analysis more economically sophisticated and more respectful of the legal methodology adopted in the landmark *Sony* decision.[225] I now argue that categorization projects are so necessary to counteract the negative effects of information overload that they deserve positive recognition in the *first* fair use factor, which focuses on the "purpose or character of the use."[226] Traditional analysis of whether the use is commercial and transformative has extremely limited utility in the categorization context. Courts can short-circuit these endlessly manipulable formal distinctions by recognizing categorization as a *per se* pro-defendant finding in the first fair use factor. That would not mean an automatic fair use finding—there are, of course, three other factors to examine. But it would at least provide some

[222] *Video Pipeline*, 342 F.3d 191, at 202.

[223] *Id.*

[224] Id.

[225] Pasquale, *Toward an Ecology of Intellectual Property*, *supra* note 17; *Breaking the Vicious Circularity*, *supra* note 67. I still believe that detailed inquiry into the actual effects of a use on the value of a copyrighted work is essential to applying the statute. However, given that the costs of copyright litigation are one of the main impediments to fair use, richer fourth factor inquiry may ultimately prove not to be much of a help to defendants, especially if they cannot afford the experts commonly necessary in such litigation.

[226] 17 U.S.C. § 107(1) (2004).

measure of judicial recognition of the value of categorizers and indexers.

Next, in order to level the litigation playing field, I suggest that aggressive efforts by content holders to shut down categorizing sites should constitute a form of copyright misuse. Developed from the doctrine of patent misuse in the 1990's, the misuse defense may reasonably balance Congress's recent expansion of copyright (and paracopyright) protections. As an equitable defense, misuse doctrine protects innovators in fields related to, but ultimately not directly covered by, the legal rights of a copyrightholder. Many content owners have used aggressive litigation tactics not only to control the use of their copyrighted material, but also to leverage that control into veto power over any categorizers who sample their work.[227] Such aggressive tactics are exactly the type of inefficient and unfair competitive tactics that misuse doctrine was designed to combat.

1. Categorization as Privileged Fair Use

The moral and economic arguments for this position have already been laid out in Section A above: information overload is a real problem and search engines do much to alleviate it. Yet these arguments must find a basis in extant doctrinal analysis if they are to convince courts. We can find such roots in a rather unlikely place—a 2002 dispute between the extraordinarily litigious stuffed animal manufacturer, Ty, Inc., and a publisher of guides to Ty's "beanie baby" products.[228] In this case, Ty, the owner of copyrights in

[227] Such efforts also serve to dampen new, unauthorized commentary on works in order to (comparatively) raise the profile of extant distribution and promotion networks. See Mark S. Nadel, *How Current Copyright Law Discourages Creative Output: The Overlooked Impact of Marketing*, 19 BERKELEY TECH. L.J. 785 (2004) (arguing that "protection against unauthorized copying provides dramatically disproportionate benefits to the most popular creations: it enables the publishers seeking to create blockbusters to finance enormous promotional campaigns, which drown out valuable, artistic creations that lack competitive marketing efforts.").

[228] Ty, Inc. v. Publ'ns Int'l, 292 F.3d 512 (7th Cir. 2002).

various "Beanie Babies" (stuffed animals copyrighted as "sculptural works") sued the publisher of books featuring images of Beanie Babies (including a collector's guide and a "picture book" entitled *For Love of Beanie Babies*). Writing for the Seventh Circuit panel, Judge Posner characterized *For Love of Beanie Babies,* a children's book whose central appeal was amusing arrangements of particular "species" of Beanie Babies into scenes, as "*essentially* just a collection of photographs of Beanie Babies, and photographs of Beanie Babies are derivative works from the copyrighted Beanie Babies themselves."[229] The *categorizing* work merited distinctly more favorable treatment:

> PIL's *Beanie Babies Collector's Guide* . . . is a small paperback book with small print, clearly oriented toward adult purchasers--indeed, as the title indicates, toward collectors. Each page contains, besides a photograph of a Beanie Baby, the release date, the retired date, the estimated value of the Beanie Baby, and other information relevant to a collector, such as that "Spooky is the only Beanie ever to have carried his designer's name," or that "Prance should be a member of the Beanie line for some time, so don't panic and pay high secondary-market prices for her just because she's fairly new."[230]

Judge Posner notes that Ty licensed the right to publish photos of Beanie Babies only to authors of collectors' guides who promise not to criticize Ty in their guides.[231] This state of affairs indicates the importance of independent categorizers; if all collector's guides are licensed, consumers won't be able to trust whether they're getting accurate information about the market or are simply being fed talking points helpful to the interests of dominant producers. Melding first and fourth factor fair use analysis, Judge Posner observes:

---

[229] *Id.*.

[230] *Id.*, 519-520.

[231] *Id.* ("Some of the text is quite critical, for example accusing Ty of frequent trademark infringements.").

> Generalizing from this example in economic terminology that has become orthodox in fair-use case law, we may say that copying that is complementary to the copyrighted work (in the sense that nails are complements of hammers) is fair use, but copying that is a substitute for the copyrighted work (in the sense that nails are substitutes for pegs or screws), or for derivative works from the copyrighted work, is not fair use. . . . The hammer manufacturer wants there to be an abundant supply of cheap nails, and likewise publishers want their books reviewed and wouldn't want reviews inhibited and degraded by a rule requiring the reviewer to obtain a copyright license from the publisher if he wanted to quote from the book.[232]

The existence of reviewing sites uncontrolled by the owners of the material reviewed may be essential to the assurance of trustworthy sources of information about such works.[233] Although *Ty* focuses on reviews, categorization and indexing may count as just as socially useful a purpose, and may be the only effective way of keeping track of materials to be reviewed (or reviews themselves). In an era of information overload, there are many reasons to immunize the efforts of those who give us a sense of "what's out there" from holdouts who would make the task prohibitively expensive.

Fair use findings for snippets are also important to Google's archiving project. In a series of cases involving software, courts have protected users' rights to make an *intermediate copy* of a work in order to reverse engineer its

---

[232] Ty, Inc., at 518 (emphasis added) (citing On Davis v. The Gap, Inc., 246 F.3d 152, 175-76 (2d Cir. 2001); Suntrust Bank v. Houghton Mifflin Co., 268 F.3d 1257, 1277 (11th Cir. 2001) (concurring opinion); Wendy J. Gordon, *Fair Use as Market Failure*, *supra* note 80, at 1643.

[233] *See* 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT, *supra* note 61, at 13-293..

noncopyrightable elements—and to circumvent technological measures designed to prevent such intermediate copying.[234] In other words, a software coder is entitled to make a copy of a work in order to discover how it works, and then to replicate those elements of it that are not copyrightable. The doctrine appears tailor-made for the Google Library project, which intends not to provide full copies of copyrighted works to searchers, but only small snippets of text deemed relevant to their queries. To the extent the snippets are protected, the larger archiving project may be eligible for the intermediate copying defense.[235] Given the accidental (and inevitable) destruction of so much analog data over time, this digital archiving project is of immense cultural importance.[236]

Per se favorable first factor treatment for categorizers who merely provide metadata and samples (and not copies of

---

[234] *See* 17 U.S.C. § 1201(f) (2004) (addressing reverse engineering); Paul Ganley, *Google Book Search: Fair Use, Fair Dealing and the Case for Intermediary Copying*, *supra* note 70 (advocating British adoption of an American-style intermediate use doctrine in order to immunize actions like Google's).

[235] See Jonathan Band, *The Google Library Project: Both Sides of the Story*, 1(2) PLAGIARY 1, -17 (8 February 2006) ("The owners respond that the intermediate copying cases are distinguishable because they address a problem specific to software: translation of the programs is the only means of accessing ideas unprotected by copyright that are contained within the program. This problem, of course, does not exist with books. Furthermore, in the intermediate copying cases, the software developer discarded the translation once it developed its new, non-infringing program. Google, conversely, will retain the scanned copy in its search index. While acknowledging these factual differences, Google's supporters stress the underlying principle of the intermediate copying cases: that copying may be excused if it is necessary for a socially useful non-infringing end use.").

[236] *See* Mary Sue Coleman, *Google, the Khmer Rouge, and the Public Good*, Address to the Association of American Publishers, Feb. 6, 2006 (copy on file with author) (discussing how disasters like Hurricane Katrina or fascist regimes like the Khmer Rouge can wipe out all unique analog copies of works, and how the Google digitization project is essential to preservation efforts).

Formatted: Font: Not Bold

works themselves) would do much to immunize the Google Library project, as well as efforts to extend its coverage to music, films, and other forms of expression. Currently, it is very easy for a court to give inordinate power to holdouts unwilling merely to indicate their refusal to be in the database to Google.[237] Per se favorable first-factor treatment would not end the fair use analysis; there are still the three other factors, and they can likely address the meritorious objections of copyrightholders.[238] Yet a first factor analysis favorable to categorizers and organizers would do much to dispel the fear, uncertainty, and doubt that aggressive content owners have used to chill legitimate fair uses over the past few decades.[239]

Admittedly, once such a privileging subfactor as categorization is established, there will be more or less legitimate uses of it. It may well be hard to define what precisely constitutes a categorizer in the abstract. However such difficulties have not proven insuperable for other "subfactors" in the first factor of fair use analysis, including transformativeness and commerciality. Courts can distinguish between highly transformative and less transformative uses, and similarly can distinguish between paradigm cases of categorization and less promising ones.

For example, one reasonable development of the doctrine would hold that a privilege for categorizers should be proportional to the scope of works categorized. A project

---

[237] As Lawrence Lessig has observed, such veto power would essentially keep about three quarters of copyrighted works out of the database, because it is impossible to find the owners of these orphan works. Joan Indiana Rigdon, *Google, Books, and Fair Use*, WASHINGTON LAWYER, Mar. 2006, at 26 (quoting Lessig).

[238] These meritorious objections may include a) security considerations (copyright owners worry about their industry being "Napsterized" if someone breaches the security of Google's or the partnering libraries' digital copies of the works) and b) an overbroad "search engine" exception (which would permit, say, a fanfic site to digitally copy in all versions of the work ostensibly in order to let users find their favorite quotes, but really in order to let them download and print works at will).

[239] For examples of such overclaiming, see Brennan Center report and Loren on contractual overclaiming and Posner and Mcleod.

premised on indexing and categorizing, say, all Japanese films, would be a more privileged categorizer than, say, one devoted only to indexing the films of Akira Kurosawa. In the latter case, the relatively more easily found copyrightholders should have more of a claim to get some licensing revenues from the project than the enormously dispersed owners of all Japanese films. Moreover, the more comprehensive categorization is more of a public service than the narrower one. As copyright law develops, more such paradigm cases of categorization may emerge.[240]

2. Misuse Defense

Intellectual property rights are, at their core, monopolies.[241] Even though legitimately attained, the rights are subject to abuse. A copyrightholder's efforts to leverage control over content into control over a field uncovered by its derivative works rights can result in a finding of copyright misuse (completely independently of any antitrust liability).[242]

---

[240] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S at,578, citing Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990), See also NIMMER ON COPYRIGHT, *supra* note 67 § 13.05(A)(1)(b). For the weight given to commercial use, see *id.* § 13.05(A)(1)(c), discussing ., 510 U.S. at 585, *Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 and *American Geophysical Union v. Texaco, Inc.*, 60 F.3d 913, 922 (2d Cir. 1994).

[241] Frischmann and Moylan, Chapter on Copyright Misuse (forthcoming 2006); 4 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT, *supra* note 61 ("The Copyright Act accords to each copyright owner a limited form of monopoly."). The doctrine of misuse arose in part out of concern about sham litigation by copyright and patent holders designed to intimidate rivals into not exercising rights that were legally theirs. Id., at § 13.09 [A] [1] (discussing *Colum. Pict. Indus., Inc. vs. Prof. Real Estate Investors, Inc*., 866 F.2d 278 (9th Cir. 1989). As Heins and Beckles have demonstrated, the chilling effects of "weak IP claims" are legion. Heins and Beckles, *Will Fair Use Survive*, supra note 70, at 33.

[242] *Id.*, at 2; Lasercomb v. Reynolds, 911 F.2d 970 (4th Cir. 1990) (finding misuse where software copyright owner's licensing agreement sought to suppress any attempt by the licensee to independently implement the software's idea); Alcatel USA, Inc. v.

After such a finding, a content owner's copyrights are invalid until the misuse is "purged."[243] Although misuse findings have not been common, they may prove a more effective "shield" for categorization projects than fair use. While fair use doctrine is by nature extraordinarily malleable and indeterminate, misuse presents a relative straightforward assessment of whether copyrightholders (either alone or in concert) have attempted to "strong-arm" control over given works into control over a whole other industry or field.

Several commentators have praised the development of the misuse doctrine as a balance to copyrightholders' overreaching.[244] Of all the defenses in copyright law, misuse appears ideally suited to categorizers. To be reliable, categorizers should be independent of the owners of the content they are reviewing and commenting on. They should not be subject to sanctions or reprisals from large content owners angry at the categorizer's treatment of their properties—be it a low ranking, a bad review, or a brusque

---

DGI Techs., Inc., 166 F.3d 772 (5th Cir. 1999) (finding misuse where software copyright owner tried to prevent defendant from testing its cards in conjunction with DSC's software (and thereby illicitly attempted to secure a monopoly over uncopyrighted microprocessor cards)); Practice Mgmt. Info. Corp. v. Am. Med. Assoc'n, 121 F.3d 516, 520 (9th Cir. 1997) (finding misuse in licensing agreement which plainly required the Health Care Financing Administration to use the AMA's copyrighted coding system and no other); Scott A. Sher, Case Note, In re Napster Inc. Copyright Litigation: *Defining the Contours of the Copyright Misuse Doctrine,* 18 SANTA CLARA COMP. & HIGH TECH. L.J. 325, 329 (2002) (discussing district court order indicating the potential viability of Napster's misuse defense before it was bought by one of the companies suing it).

[243] Nimmer, *supra* note 61, at 2 (citing Lasercomb, 911 F.2d at 979 n.22; *see also* Practice Management Information Corp. v. American Medical Ass'n, 121 F.3d 516, 520 n.9 (9th Cir.).

[244] *See*, e.g., Dan L. Burk, *Anticircumvention Misuse*, 50 U.C.L.A. L. REV. 1095 (2003); Lydia Pallas Loren, *Slaying The Leather-Winged Demons In The Night: Reforming Copyright Owner Contracting With Clickwrap Misuse,* 30 OHIO N.U. L. REV. 495, 497 (2004).

dismissal.[245] Unfortunately, a recent categorizer case to address the misuse defense directly, *Video Pipeline*, appeared to confuse it with a more general First Amendment argument against excessive control over copyrighted work.[246] The *Video Pipeline* court conceded that "anti-competitive licensing agreements may conflict with the purpose behind a copyright's protection by depriving the public of the would-be competitor's creativity."[247] However, the court refused to find misuse, holding that the defendant was free to criticize Disney films on websites lacking Disney trailers.[248] If the sole value of categorizing sites were commentary, perhaps this crabbed view

[245] Circuit courts addressing potential misuses of copyright have also focused on the illegitimacy of using copyright to expand a copyrightholder's power over information that is not covered the copyright. See, e.g., Assessment Techs. v. Wiredata, 350 F.3d 640, (7th Cir., 2003) ("To try by contract or otherwise to prevent the municipalities from revealing their own data, especially when, as we have seen, the complete data are unavailable anywhere else, might constitute copyright misuse. The doctrine of misuse "prevents copyright holders from leveraging their limited monopoly to allow them control of areas outside the monopoly.") (quoting A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004, 1026-27 (9th Cir. 2001)).

Formatted: Font: 11 pt, Not Bold

[246] See Frischmann and Moylan, supra n. *, for a comprehensive critique of Video Pipeline.

[247] *Id.*, at 204.

[248] *Id.*, at 206 ("The licensing agreements in this case do seek to restrict expression by licensing the Disney trailers for use on the internet only so long as the web sites on which the trailers will appear do not derogate Disney, the entertainment industry, *etc.* But we nonetheless cannot conclude on this record that the agreements are likely to interfere with creative expression to such a degree that they affect in any significant way the policy interest in increasing the public store of creative activity. The licensing agreements do not, for instance, interfere with the licensee's opportunity to express such criticism on other web sites or elsewhere.")

of the misuse doctrine would be valid.[249] But as Part III demonstrated, even categorizers that offer the barest comment on copyrighted content create value by *sorting* content. The misuse and fair use defenses do not simply serve to facilitate more expression about extant expression. They are also designed to protect the independent categorization and organization of copyrighted work.

From a strictly economic perspective, bargains between categorizers and content owners may appear welfare-maximizing.[250] For example, Google's recent deal with the Associated Press clears enables it to develop new content and provides a new revenue source for an embattled old media stalwart.[251] Google has also started licensing content from

---

[249] Followed literally, *Video Pipeline* suggests that some courts will respect copyrightholders' demands to set up *their own* method of categorizing and providing samples of works and exclude others from entering this market. I believe such demands may be a kind of copyright misuse (an illicit effort to leverage control over copyrighted works into control over other markets which are not properly considered derivative works). *See, e.g.*, Matt Richtel, *Aggressive Strategy Brought on Inquiry of Recording Industry*, *N.Y. Times*, Oct. 22, 2001 (discussing antitrust inquiry catalyzed by RIAA's effort to dominate the online music retailing market); Matthew Fagin, Frank Pasquale, and Kimberlee Weatherall, *Beyond Napster: Using Antitrust Law to Advance and Enhance Online Music Distribution*, 8 B.U. J. SCI. & TECH. L. 451, 513 (2002) (discussing potentially anti-competitive practices in the recording industry).

[250] See, e.g., Douglas Lichtman, *Google Print*, UNIVERSITY OF CHICAGO FACULTY BLOG, Oct. 5, 2005 (arguing for the "develop[ment of] a legal system that allows authors to share in that revenue stream").

[251] *See* Caroline McCarthy, Google Reveals Payment Deal with AP, CNet News, Aug. 3, 2006 ("On the surface, paying the Associated Press seems to conflict with the stance Google has traditionally taken regarding its Google News service. Because Google News is an aggregator, the company has argued, Google is not obliged to reimburse news outlets for linking to their content. But Wednesday's announcement said the AP content will be the foundation for a new product that will merely complement Google News. Thus Google maintains that the deal supports its original stance on fair use.").

studios and sees such "deals as a key to long-term growth."[252] In the short-run, such licensing practices may seem like an ideal compromise between a categorizer and the owners of the content it orgnanizes.

However, traditional economic analysis does not take into account adverse cultural consequences of categorizers dependent on (or otherwise aligned with) the owners of categorized content.[253] If Google and other large search engines are forced by adverse fair use decisions to license content from copyright holders, we can easily imagine deals

---

[252] See Kevin J. Delany, *Google Sees Content Deals as Key to Long-Term Growth*, WALL ST. J., Aug. 14, 2006, B1 ("[Google] announced a deal to distribute video from Viacom Inc.'s MTV Networks on the Web and a separate agreement with News Corp.'s Fox Interactive Media division to provide it with search technology and broker advertising. Google has pledged $900 million in minimum payments to Fox under the tie-up.").

[253] For an analysis of the cultural "blind spots" of economic analysis of intellectual property, see Julie Cohen, *Copyright, Commodification, and Culture: Locating the Public Domain*, in THE FUTURE OF THE PUBLIC DOMAIN: INDENTIFYING THE COMMONS IN INFORMATION LAW 140 (Lucie Guibault and P. Bernt Hugenholtz, eds., 2006) ("Economic models of creativity treat motivation as both exogenous and abstract. . . blithely consign inspiration to the category of 'fixed costs' (or, worse, assumed inputs). . . [and] lack[] appropriate tools to study audience response to creative works [so pervasively that] . . . . what remains most important is what the models leave out."); Frank Pasquale, *Net Neutrality: Law, Money, and Culture*, CONCURRING OPINIONS, Oct. 18, 2006, available at http://www.concurringopinions.com/archives/2006/10/net_neutrality.html#more ("[W]e may want to avoid giving already-dominant entities even more opportunities to leverage existing networks of distribution into an ever more powerful hold over our collective imagination. Churches, schools, museums, indie musicians--all deserve as much of a shot at our computers as iTunes, Disney, or Comcast."). Though my post addresses "net neutrality," there are powerful analogs between "net neutrality" concerns and categorizer independence and neutrality. See, e.g., James DeLong, Search Engine Neutrality, IPCentral Weblog, May 22, 2006, available at http://weblog.ipcentral.info/archives/2006/05/search_engine_n.html; Thomas Hazlett, Google and the Myth of an Open Net, Fin. Times, Oct. 10, 2006, available at http://www.freepress.net/news/18236.

between dominant content providers and dominant categorizers that squeeze out smaller players in both fields.[254] Just as trademark law helps preserve a fair competitive playing field, copyright law should avoid, whenever possible, enabling aggressive leveraging that entrenches the dominance of large content providers or renders categorizers mere subsidiaries of such copyrightholders.[255]

Strong fair use and misuse defenses for categorization are essential to the reliability, objectivity, and diversity of the services that help us navigate the maze of copyrightable expression. Like the "licensed reviewers" whom Judge Posner mocks in Ty, licensed categorizers may be overly inclined to praise their partners' work, while ignoring others' (whatever its relevance or merits). With little fear of a successful misuse defense, large content owners like Disney can give "take it or leave it" ultimata to categorization sites via license terms, forcing them to "say nothing but good" of the content they index and comment on. Finally, if copyright becomes a major barrier to entry in the categorization field, we can expect the diversity of such sites to quickly decline. It is doubtful that any of these outcomes would promote copyright's constitutional purposes.

V. CONCLUSION

[254] *See* Frank Pasquale, *Would Google Go Out of Business Without Fair Use?*, MADISONIAN WEBLOG, available at http://madisonian.net/archives/2006/08/07/would-google-go-out-of-business-without-fair-use/ (August 7, 2006) (criticizing the "licensing solution" to the struggle between categorizers and copyrightholders); *Google's Fight and Flight Response*, MADISONIAN WEBLOG, available at http://madisonian.net/archives/2006/08/16/googles-fight-and-flight-response/ (August 16, 2006) (same).

[255] For a prescient look at the dangers of such a regime, see Niva Elkin-Koren, *Let the Crawlers Crawl: On Virtual Gatekeepers and the Right to Exclude Indexing*, 26 U. DAYTON L. REV. 179, 183 (2001) (discussing problematic consequences of concentration in the culture industry).

Imagine a comprehensive source of data—one that included all relevant material in a single index. With the advance of digitization and interconnection, there are less and less *technical* obstacles to such a "celestial jukebox," "new library of Alexandria," or consolidated collection of all types of expression.[256] As technology governed by Moore's Law advances, storage and search costs continue to decline.

However, *legal* and *business* obstacles appear to arise as quickly as technical barriers come down. Some of these obstacles may be necessary to secure compensation to copyrightholders and other entrepreneurs. But the mere indexing and archiving of readily available works—the core of categorization projects—has little if any negative commercial impact on information creators. Holdouts should not be permitted to stop such projects in the same way that permission culture has crippled innovation in the music and film industries.[257]

Giving content owners the right to control all mentions and samples of their work conduces to the creation of a desert (albeit a peaceful one) of self-serving and unreliable categorizers. A robust information ecosystem depends on spontaneous creativity, serendipitous appropriation, and accountable information sources—precisely the type of positive developments that an untrammeled market in "snippet licenses" appears less and less likely to provide. The growing burden of information overload makes all the more important a revision of fair use doctrine favoring independent categorization, and a robust misuse defense designed to deter its enemies.

---

[256] *See* PAUL GOLDSTEIN, COPYRIGHT'S HIGHWAY: FROM GUTENBERG TO THE CELESTIAL JUKEBOX (1994).

[257] MARJORIE HEINS AND TRICIA BECKLES, WILL FAIR USE SURVIVE? FREE EXPRESSION IN AN AGE OF COPYRIGHT CONTROL 25 (2005) (discussing several instances where copyrightholders' legal threats effectively vetoed apparent fair uses); PAT AUFDERHEIDE & PETER JASZI, UNTOLD STORIES: CREATIVE CONSEQUENCES OF THE RIGHTS CLEARANCE CULTURE FOR DOCUMENTARY FILMMAKERS 36-38 (2004) (discussing the negative consequences of "clearance culture" for documentary makers without extensive corporate backing).