COPYRIGHT AND COPY-RELIANT TECHNOLOGIES

Matthew Sag*

Abstract

This article studies the rise of copy-reliant technologies – technologies such as Internet search engines and plagiarism detection software that, although they do not read, understand or enjoy copyrighted works, necessarily copy them in large quantities. This article provides a unifying theoretical framework for the legal analysis of topics that tend to be viewed discretely. Search engines, plagiarism detection software, reverse engineering and Google's nascent library cataloging effort, are each part of a broader phenomenon brought about by digitization, that of copy-reliant technologies. These technologies raise two novel, yet central, questions of copyright law. First, whether a non-expressive use that nonetheless requires copying the entirety of a copyright work should be found to infringe the exclusive rights of the copyright owner. Second, whether the transaction costs associated with copy-reliant technologies justify switching copyright's default rule that no copying may take place without permission to one in which copyright owners must affirmatively opt-out of specific uses of their works.

Keywords: Internet, Copyright, Non-expressive use, Opt-out, Fair use, Transaction costs.

Word-count: 34,987.

* Visiting Assistant Professor, University of Virginia School of Law & Assistant Professor, DePaul University College of Law. Please address comments to msag@depaul.edu. Thanks to Tonja Jacobi, Adam Mossoff, Dotan Oliar, Glen Robinson, and Christopher Sprigman and to participants at the American Intellectual Property Law Association Annual Meeting, and the Intellectual Property Scholars Conference for their comments and suggestions. Special thanks to the University of Virginia School of Law Library staff for their invaluable assistance.

CONTENTS

INTRODUCTION

Although we have been living in the Internet age for more than a decade now, its implications for copyright law and the fair use doctrine are only just beginning to manifest.[1] By expanding the breadth, diversity and sheer number of copyrighted works in existence, the Internet has fundamentally changed the nature of copyright markets. This change is most significant in the context of what I term "copy-reliant technologies" — technologies that copy expressive works for non-expressive ends. Copy-reliant technologies, such as Internet search engines and plagiarism detection software, do not read, understand or enjoy copyrighted works, but they necessarily copy them in order to process them as grist for the mill, raw materials that feed various algorithms and indices.

The copyright implications of Internet search engines, plagiarism detection software, reverse engineering of software and the emerging Google Book Project controversy have been separately considered by other scholars.[2] This article is the first to provide a unifying theoretical framework for the analysis of the issues raised, largely because it is the first to recognize them as sub-parts of a broader phenomenon.

Copy-reliant technologies tend to interact with copyrighted works by copying them routinely, automatically and indiscriminately. These technologies are vital to the operation of the Internet, but they are vulnerable to claims of copyright infringement at key stages of their operation. Copy-reliant technologies typically display three significant traits: the copying of expressive works for non-expressive uses; a high volume of transactions; and the use of technologically enabled opt-out mechanisms to reduce transaction costs.

---

[1] I use the term Internet age here to refer to the period from 1994 to the present – the period in which the Internet was popularized and commercialized. Technically, the first packet-switching node of what would later be called the ARPANET went live on October 29, 1969. The first TCP/IP-wide area network was operational by January 1, 1983, when the United States' National Science Foundation (NSF) constructed a university network backbone that would later become the NSFNet.

[2] On search engines, see Urs Gasser,*Regulating Search Engines: Taking Stock and Looking Ahead*, 9 YALE J. L. & TECH. 124 (2006); James Grimmelmann, *The Structure of Search Engine Law*, 93 IOWA L. REV. 1 (2007). On reverse engineering, see, Pamela Samuelson & Suzanne Scotchmer, *The Law and Economics of Reverse Engineering*, 111 Yale L.J. 1575 (2002).On Plagiarism, see, Samuel J. Horovitz, *Two Wrongs Don't Negate A Copyright: Don't Make Students Turnitin If You Won't Give It Back*, 60 FLA. L. REV. 229 (2008). On Google Book, see e.g. Siva Vaidhyanathan, *The Googlization of Everything and the Future of Copyright*, 40 U.C. DAVIS L. REV. 1207 (2007); Oren Bracha, *Standing Copyright Law on Its Head? The Googlization of Everything and the Many Faces of Property*, 85 TEX. L. REV. 1799 (2007); Emily Anne Proskine, *Google's Technicolor Dreamcoat: A Copyright Analysis of the Google Book Search Library Project*, 21 BERKELEY TECH. L.J. 213 (2006); Elisabeth Hanratty, *Google Library: Beyond Fair Use?*, 2005 DUKE L. & TECH. REV. 10 (2005); Kinan H. Romman, *The Google Book Search Library Project: A Market Analysis Approach To Fair Use*, 43 HOUS. L. REV. 807 (2006); Hannibal Travis, *Google Book Search and Fair Use: Tunes for Authors, or Napster for Books?*, 61 U. MIAMI L. REV. 87 (2006); Steven Hetcher, *The Half-Fairness Of Google's Plan To Make The World's Collection Of Books Searchable*, 13 MICH. TELECOMM. TECH. L. REV. 1 (2006); Michael R. Mattioli, *Opting Out: Procedural Fair Use*, 12 VA. J.L. & TECH. 3 (2007).

The rise of copy-reliant technologies exposes seemingly novel questions. First, should a non-expressive use that nonetheless requires copying the entirety of a copyright work be found to infringe the exclusive rights of the copyright owner? Our historical intuition is that when a work is copied it is copied to communicate at least some part of the work's original expression: books are copied to be read, not to serve as paper weights; compact discs are copied to be played, not to function as drink coasters. This Article concludes that because the copyright owner's exclusive rights are implicitly defined and limited in reference to expressive communication to the public, acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement.

The second important question raised by copy-reliant technologies springs from an empirical observation. The architects of many of the copy-reliant technologies surveyed in this Article have chosen to build in technologically enabled opt-out mechanisms that preserve the autonomy of the copyright owner, but switch the default rule from 'no copying without permission' to one in which copyright owners must affirmatively opt-out of specific uses of their works. Accordingly, the second question whether this switch in copyright's ordinary default rule is justified from either a doctrinal or a utilitarian perspective.

The questions that come into focus in this study of copy-reliant technologies are to a large extent questions about fair use. Technically, the fair use doctrine renders certain otherwise infringing actions relating to copyrighted works non-infringing.[3] More generally, fair use allows the use of copyrighted works without permission; as such it performs a vital function in the modern copyright system by establishing limits on the otherwise expansive rights of copyright owners.[4] Because of the fair use doctrine's pivotal role in adapting copyright law to new technology, it is inevitable that any examination of copyright and new technology becomes a reflection on the nature of fair use. This Article makes a significant contribution to our understanding of the fair use doctrine by explaining its application in the context of non-expressive use and in situations where the alleged infringer has provided copyright owners with the ability to opt-out.

Part I of this Article introduces the phenomenon of copy-reliant technologies by focusing on four significant case studies. The first case-study, *Field v. Google Inc.*, centers on the permissibility of automated archiving in the context of text-based search engines.[5] The second case-study, *Perfect 10 v. Amazon* centers on the creation and display of thumbnail representations of copyrighted photographs used by image-based search engines.[6] The conduct challenged in the third case-study, the Google Book Project, relates both to the generation of metadata and to the display of uncopyrightable fragments of books as part

[3] 17 U.S.C. § 107 ("the fair use of a copyrighted work … is not an infringement of copyright").
[4] As I have argued elsewhere, this function actually allows copyright owners a broader set of exclusive rights than would otherwise be possible. Matthew Sag, *God in the Machine, A New Structural Analysis of Copyright's Fair Use Doctrine*, **[cite]** (2005).
[5] Field v. Google Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006), see *infra* Part I-B-1.
[6] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007), see *infra* Part I-B-2.

of a menu of search results.[7] The final case-study is that of plagiarism detection software which also addresses the use of copyrighted works to generate metadata.[8] These case-studies illustrate common features of copy-reliant technology – the non-expressive use of copyrighted works, the high transaction costs faced by copy-reliant technologies and the use of opt-outs as a method of private ordering to mitigate those transaction costs – which are further developed in Parts II and III.

Part II explores the doctrinal implications of the non-expressive use of copyrighted works. Traditionally, copyright owners have been able to control significant communicative or expressive uses of their works– such as reproduction, display and performance. In contrast, copy-reliant technologies typically use copyrighted works in a way that is non-communicative and non-expressive. A careful review of existing copyright doctrine shows that the rights of copyright owners do not encompass non-expressive uses of their works. I argue that this principle of non-expressive use resolves many questions relating to copy-reliant technologies. It also reconciles many puzzling features of the fair use doctrine more broadly. The doctrinal incorporation of this principle of non-expressive use through the application of the fair use doctrine is then addressed in detail.

Part III studies the doctrinal implications of high transaction costs in relation to copy-reliant technologies and the use of opt-out mechanisms to mitigate those transaction costs. It then integrates that discussion with an analysis of the relationship between transaction costs and the form and content of property rights generally, and the relevance of opt-outs to a fair use analysis.

I. COPY-RELIANT TECHNOLOGIES AND THE INTERNET

This part begins, in Part I-A, with a general discussion of the link between the technological and social changes of the Internet era and the evolution of copyright law. It also explains the centrality of the fair use doctrine in revising that balance as technology and market conditions change. This framework is forms the essential theoretical background for understanding the significance of the copy-reliant technology. Part I-B describes four case-studies of copy-reliant technology which serve to illustrate the concept and its application. These case-studies are the empirical backbone of this Article; they are introduced in this Part and further developed in subsequent Parts. As the case-studies illustrate, copy-reliant technologies tend to raise certain recurring legal issues: the copying of expressive works for non-expressive uses; the potential for high transaction costs associated with copy-reliant technologies; and the role of opt-out mechanisms in addressing these transaction costs problems. These issues are addressed more fully in Parts II and III.

[7] Authors Guild v. Google, No. 05 CV 8136 (S.D.N.Y. Sept. 20, 2005), see *infra* Part I-B-3.
[8] Av et al v. Iparadigms, Llc, US District Court Civil Docket 1:07cv293 (opinion of March 11, 2008 available at http://www.iparadigms.com/iParadigms_03-11-08_Opinion.pdf), see *infra* Part I-B-4. Space constraints preclude addition case-studies such as software reverse engineering.

A. New Technologies, Copyright Markets and Copyright Law

From the printing press, to the photocopier, from the piano-roll to the mp3 player, new technology change has fundamentally altered copyright law.[9] Photography, motion pictures, sound recording and broadcasting have each demanded and (eventually) received accommodation from copyright law.[10] As the technologies of reproduction and communication change, they create new vehicles of creative expression, new communities of interest, and expose latent ambiguities within existing doctrines.[11]

In some respects, the new technologies of copying and distribution that form the Internet represent a continuation of this trend. Napster's peer-to-peer file sharing technology (or more recently, Bittorrent) exemplifies how digital technology and online distribution allows existing works, such as sound recording and motion pictures, to copied and distributed at virtually no cost. Unlocking content from physical delivery has facilitated more than just piracy; it has also enabled legal digital music services which have made more music available at a lower cost than ever before.[12]

Advances in technology have also opened up new possibilities of creative production by reducing the cost of sound and video editing. The video editing software that was used to create *the Phantom Edit*[13] – an edited version of *StarWars I* without the much-reviled character Jar Jar Binks character – used to be reserved for Hollywood studios alone; it is now widely available for less than the cost of a new television.[14] These new possibilities have done more than simply lower costs for existing producers; they have introduced new participants and in some cases dramatically changed the medium. Just as newspapers and television reporting have been changed by the rise of political blogs,[15] advances in digital

[9] See, e.g., PAUL GOLDSTEIN, COPYRIGHT'S HIGHWAY: FROM GUTENBERG TO THE CELESTIAL JUKEBOX (2003) (tracing the development of copyright law in the United States); JESSICA LITMAN, DIGITAL COPYRIGHT (2001) (tracing the history copyright legislation in the United States); Peter Menell, *Envisioning Copyright Law's Digital Future*, 46 N.Y.L. SCH. L. REV. 63 (2002) (arguing that the digital revolution represents a third distinct wave of technological innovation that portends significant changes in copyright protection).

[10] See generally, Litman, Digital Copyright, *supra* note 000.

[11] See, LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE, 22–23 (1999); WILLIAM LANDES & RICHARD POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW __ (2003).

[12] See generally, CHRIS ANDERSON, THE LONG TAIL 139 (2006) (noting the effect on price).

[13] See, Amy Harmon, *'Star Wars' Fan Films Come Tumbling Back to Earth*, N.Y. TIMES, Apr. 28, 2002, § 2 (Arts & Leisure), at 28. (discussing *Star Wars 1.1: The Phantom Edit* and its creation).

[14] See, WIKIPEDIA, *List of video editing software*, http://en.wikipedia.org/wiki/List_of_video_editing_software for a list of video editing software, including several free and open source modules.

[15] Kevin Wallsten, *Agenda Setting and the Blogosphere: An Analysis of the Relationship between Mainstream Media and Political Blogs,* 24(6) REVIEW OF POLICY RESEARCH 567–587 (November 2007) (finding a complex, bidirectional relationship between mainstream media coverage and blog discussion rather than a unidirectional media or blog agenda-setting effect). Stephen A. Banning & Kaye D. Sweetser, *How Much Do They Think It Affects Them and Whom Do They Believe?: Comparing the Third-Person Effect and Credibility of Blogs and Traditional Media*, 55(4) COMMUNICATION QUARTERLY 451–446 (November 2007) (finding no observable differences between the credibility of blogs and that of more traditional media); See also Larry E. Ribstein, *From Bricks To Pajamas: The Law And Economics Of Amateur Journalism* 48 WM AND MARY L. REV. 185 (2006).

technology have created whole new genres of entertainment, such as the mash-ups which blend the vocal tracks from one song over the instrumental or rhythm track of another.[16]

Regarding these changes and their implications for copyright law as a mere continuation of past technological changes risks missing the significant transformation that has occurred.[17] Digital technology and the Internet have significantly expanded the scope, diversity and sheer number of copyrighted works in existence. The cost of reproducing and disseminating digital works has not merely fallen; in many cases it has become entirely trivial. Thus, the Internet has seen not just an increase in copying, but an exponential increase. Similarly, copyright policy in the Internet age requires more than the inclusion of one or two neglected interest groups, copyright law now reaches deep inside the home and must take account of a much broader set of stakeholders than ever before.[18] The proliferation of copyrighted works in the Internet age is not simply a question of scale; the Internet has radically decentralized the production of information and expressive works such that the producers of publicly available copyrighted works are now more numerous and more diverse than at any time in human history.

The magnitude of these changes does not automatically suggest that copyright has no application online, or that the substantial body of copyright law that has developed over the past two centuries should be discarded. On the contrary, many of the principles and distinctions derived from pre-internet cases are equally applicable online. In many cases, the mere fact that copying took place online is of little or no relevance. For example, the legality of the 383,000 self-described parody videos hosted on YouTube[19] will largely depend on the amount of copyrighted material taken by the parodist,[20] and on whether the work is reasonably perceived as a genuine parody or critique of the copyright owner's work.[21] These questions are the same now as they were in 1994 when the Supreme Court last addressed the issue.[22]

However, this apparent continuity should not blind us to significant underlying changes. The advent of discussion boards, blogs, social networking sites, photo sharing sites and other user-generated content has made the fair use doctrine more important to more people than ever before. The fair use doctrine has become increasingly significant to the general public because the digital technology and the Internet have enabled new forums and new ways to interact with copyrighted material which involve copying.[23] For example, whereas posting the contents of newspaper article on an Internet discussion

[16] See, Roberta Cruger, *The Mash-Up Revolution,* Salon.com (http://dir.salon.com/story/ent/music/feature/2003/08/09/mashups_cruger/)
[17] See also, Menell, *supra* note 000 at 64 (discussing the relationship between new technology and new modes of expression).
[18] See, Litman, Digital Copyright, supra note 000.
[19] YouTube.com, search query=parody, performed on February 20, 2008 at 2.56 EST.
[20] Berlin v. E. C. Publications, Inc. 329 F.2d 541 (2d Cir. 1964).
[21] Campbell v. Acuff-Rose Music, 510 U.S. 569 (1994); Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc., 109 F.3d 1394, 1399 (9th Cir. 1997); SunTrust Bank v. Houghton Mifflin Co., 268 F.3d 1257 (11th Cir. 2001).
[22] Campbell v. Acuff-Rose Music, 510 U.S. 569 (1994).
[23] L.A.Times v. Free Republic, 2000 U.S. Dist. LEXIS 5669 (D. Cal. 2000).

board or blog raises the specter of copyright infringement; cutting out the same article and sending it a friend through the U.S. postal service does not. Another significant reason is that sharing user-generated content online makes it much easier for copyright owners to detect any given infringement.[24] The scope of fair use in relation to YouTube videos, blogs and other forms of user-generated content is an important issue, but it is not the focus of this Article. Instead, this Article explores a different set of issues that augers a more fundamental change in the way we think about copyright and fair use. Specifically, this Article addresses the operation of the fair use doctrine in relation to copy-reliant technologies, such as Internet search engines, electronic archives, plagiarism detection software and other applications which rely on copying expressive works for non-expressive ends.

These technological and social changes are significant for copyright because by expanding the breadth, diversity and sheer number of copyrighted works in existence, the Internet has fundamentally changed the nature of copyright markets. To appreciate the significance of these changes for copyright law, it is first necessary to examine the economic function of copyright.

Copyright creates exclusive rights in certain forms of expression in order to give authors an incentive to create those works in the first place. However, these same exclusive rights raise the cost of acquiring works for both consumers and subsequent authors. In the world of tangible objects, these costs become price signals that ensure the efficient allocation of goods to those who value them most; however, given that expressive works can be consumed again and again by different people without diminishing their value, the exclusive rights established by copyright also result in some dead weight loss because those who are unwilling to pay the higher price are forced to go without the work in question.[25]

The author's exclusive rights under copyright law provide a buffer against price competition. This buffer to competition allows the author to charge higher prices than she otherwise would, which in turn has two immediate effects. First, some consumers remain willing to purchase the work at a higher price and consequently pay more. Assuming we value the welfare of both consumers and authors equally, this is simply a wealth transfer and is welfare-neutral. Second, those who are unwilling to pay the higher price are forced to go without the work in question.[26] Market allocation of scarce resources to their highest valued use is usually welfare enhancing, but for nonrivalrous goods, the exclusion of low value users produces a deadweight loss because their consumption is not at the expense of another who values the good more.[27]

---

[24] *See,* Molly Shaffer Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535, 1539 (2005).

[25] For a more detailed discussion of the economics of copyright, see Matthew Sag, *Beyond Abstraction: The Law and Economics of Copyright Scope and Doctrinal Efficiency*, 81 TUL. L. REV. 187 (2006).

[26] This assumes, realistically, the absence of perfect price discrimination. See, Kathleen Carroll & Dennis Coates, *Teaching Price Discrimination: Some Clarification*, 66 S. ECON. J. 466, 471-78 (1999) (noting that the assumption that price discrimination is efficient is often implausible).

[27] See e.g., STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW __ (2004) (describing the effect of exclusion on resource allocation).

That copyright requires a balance between "the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand"[28] has long been understood. What is sometimes less clearly grasped is that where this balance should be struck depends not just on the relative needs of authors and consumers, but also on how effectively we expect those parties to cooperate and compromise.[29] In many situations, authors can license their creations with relative ease and the theoretical loss of exclusion is minimal.[30] In other situations however, copyright markets do not function so smoothly. Sometimes copyright owners "wield their economic control with the deftness of a surgeon's scalpel",[31] other times it is more like a cudgel. For example, Stephen Joyce, who controls the literary estate of his grandfather, James Joyce, has been accused of attempting to control access to unpublished material in order to influence historical and literary conceptions of his famous grandfather.[32] The Joyce estate's threats of copyright litigation forced one Joyce biographer to file for a declaratory judgment that her academic book and proposed electronic supplement did not infringe copyright.[33] Biographers of Howard Hughes have faced similar difficulties.[34]

Copyright law addresses potential market malfunctions in a number of ways. Doctrines such as the idea-expression distinction protect the expressive elements of the author's work while guaranteeing subsequent authors the necessary breathing space to make their own contributions by adding to, re-using, or re-interpreting, the facts and ideas embodied in the original work.[35] Statutory exemptions and compulsory licenses – such as the special reproduction rights of libraries and archives,[36] and the compulsory license for making and distributing phonorecords[37] – also provide some breathing space, however their scope tends to be limited. The primary way in which copyright law adjusts to potential market malfunctions is through the evolution of the mercurial doctrine of fair use.

---

[28] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984).
[29] Sag, *supra* note 000 (discussion the relationship between copyright scope and the effectiveness of private ordering).
[30] See Goldstein, *supra* note 000 at 5 (discussing product differentiation through versioning in the book publishing and motion picture industries)
[31] *Id.*
[32] Shloss v. Sweeney, 2007 U.S. Dist. LEXIS 41847 (D. Cal. 2007); R. Anthony Reese, *Public but Private: Copyright's New Unpublished Public Domain*, 85 TEX. L. REV. 585, 618 (2007). See also, D.T. Max, *The Injustice Collector*, NEW YORKER, June 19, 2006, at 34-43 (an account of Stephen Joyce's various threats of copyright litigation).
[33] Shloss v. Sweeney, 2007 U.S. Dist. LEXIS 41847 (D. Cal. 2007).
[34] Rosemont Enters., Inc. v. Random House, Inc., 366 F.2d 303 (2d Cir. 1966).
[35] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984). See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 240 (2d Cir. 1983) (describing the idea-expression distinction as "an effort to enable courts to adjust the tension between these competing effects of copyright protection.")
[36] 17 U.S.C. § 108(a).
[37] 17 U.S.C. § 115.

Fair use is a flexible standard which limits the scope of copyright protection and renders certain actions relating to copyrighted works non-infringing.[38] Activities that courts have regarded as fair use that may have otherwise been infringing include: quoting a significant portion of a work for the purpose of criticism, illustration, comment or clarification; parodying a work; and copying part of a work in the course of classroom activities.[39] Judges and legal scholars frequently attest to the importance of the fair use doctrine,[40] however, the exact nature of fair use remains elusive and resists straight forward definition.[41]

Fair use allows the use of copyrighted works without permission; as such it performs a vital function in the modern copyright system by establishing limits on the otherwise expansive rights of copyright owners. Fair use is necessary, in part, because licensing and other private ordering mechanisms do not provide a solution for cases involving high transaction costs, strategic holdouts, and inadvertent copying.[42] The fair use doctrine is particularly important in situations where the costs of obtaining permission outweigh the benefits of the use. The doctrine also plays a mediating role in situations where the copyright owner withholds permission for reasons that we as a society find unacceptable. For example, a copyright owner usually cannot deny permission to copy in order to stifle parody, criticism or social debate.[43]

B. Four Case-studies of Copy-reliant Technology

Much of the discussion that follows concentrates on various forms of search technology as compelling illustration of the issues that apply to copy-reliant technologies more generally: the copying of expressive works for non-expressive uses; the potential for high transaction costs associated with copy-reliant technologies; and the role of opt-out mechanisms in addressing these transaction costs problems.

Search technology is clearly a significant public policy issue.[44] The Internet has become an integral part of modern existence. For many, it is the dominant medium of communication, research, entertainment, social interaction and political participation.[45]

---

[38] 17 U.S.C. § 107.
[39] *See, Report of the Register of Copyrights on the General Revision of the U.S. Copyright Law* (1961).
[40] *See e.g.* Ty, Inc. v. Publ'Ns Int'l, 292 F.3d 512, 518 (7th Cir. 2002) (Posner) (the fair use doctrine plays an essential role in copyright law).
[41] *See* Dellar v. Samuel Goldwyn, Inc., 104 F.2d 661 (2d Cir. 1939) (describing fair use as one of the most troublesome doctrines in the whole law of copyright).
[42] See, Sag, *supra* note 000 at 250 (criticizing doctrinal recommendations which aim to optimize copyright scope in the abstract but do not account for the effect uncertainty or strategic behavior); see also Gideon Parchomovsky and Kevin A. Goldman, *Fair Use Harbors*, 93 VA. L. REV. 1483 (2007).
[43] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994); SunTrust Bank v. Houghton Mifflin Co., 268 F.3d 1257 (11th Cir. 2001).
[44] Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000).
[45] See, PEW RESEARCH CENTER, *Social Networking and Online Videos Take Off, Internet's Broader Role In Campaign 2008*, available at http://www.pewinternet.org/pdfs/Pew_MediaSources_jan08.pdf (summarizing survey data).

Search technology drives the Internet.[46] Without reliable search technology the world's 1.2 billion Internet users[47] would have very little hope of finding what they were looking for among the hundreds of billions of individual web-pages comprising the World Wide Web.[48] Search engines allow users to sift through massive amounts of data to find the specific information that is of particular interest to them. Without search engines, most content on the Internet would simply never be found and thus, in most cases, probably never created or posted in the first place. [49]

Internet search engines typify copy-reliant technology in that they require the routine and indiscriminate copying of expressive works for non-expressive purposes. Search engines copy expressive works in order to apply certain mathematical functions to their contents, they do not comprehend or enjoy copyrighted works in the way that humans do – they simply process them as raw materials that feed various algorithms and indices. The raw data underpinning modern search engines is gathered initially by automated software agents that continuously "crawl" across the Internet copying web-pages that are latter analyzed and cataloged. As part of this process, search engines both index and copy each web-page they find, and store the HTML code from those pages in a temporary repository called a cache.[50]

Search engine users are directed to particular websites based on the relationship of their search term to the index of pages maintained by the search engine provider.[51] Typically, search engines display search results in a menu which features both the title of the relevant webpage and a short "snippet" or extract from the targeted web-page. The snippet is followed by both a hyperlink to the actual web-page and another link to the cached version of the page stored on the provider's servers. Thus, search engines must copy web-pages to generate the data that allows them to process search requests. They

---

[46] Although the network of interlinked web-pages and resources that comprises the World Wide Web is commonly referred to as "the Internet", the Internet is also comprised of many other key elements, such as electronic mail, online chat services, and various file transfer networks. Technically, the Internet is a physical network, comprised of millions of household, local, academic, business, and government networks, all linked together by copper wires, fiber-optic cables and wireless connections. Whereas the Internet is defined by its physical features, the Web (or World Wide Web) is defined by its content: the Web is a collection of interconnected documents and other resources, linked by hyperlinks and URLs. Following common usage, references herein to the Internet encompass both the physical layer and the content layer. See, Robert E. Kahn &Vinton G. Cerf, *What Is The Internet (And What Makes It Work),* December, 1999, (Corporation For National Research Initiatives). Available at http://www.cnri.reston.va.us/what_is_internet.html.

[47] Internet World Stats, *World Internet Usage And Population Statistics*, available at http://www.Internetworldstats.com/stats.htm.

[48] See *infra* note 232 and accompanying text.

[49] Introna & Nissenbaum, ***supra* note 000**. See generally, JOHN BATTELLE, THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE (2005) (an account of the history and significance of Internet search).

[50] The three most popular search engines are currently Google, Yahoo!, and MSN. Each of these displays "cached" links with their search results. See, Enid Burns, *U.S. Search Engine Rankings, December 2007*, Search Engine Watch, Feb 5, 2008. (estimating Google's market share at 58.4 percent, Yahoo's at 22.9 percent and Microsoft's at 9.8 percent) Available, http://searchenginewatch.com/showPage.html?page=3628341.

[51] See, U.S. Pat. No. 6,285,999, "Method for Node Ranking in a Linked Database".

also must copy web-pages in order to display fragments of them as search results. The centrality of copying to these routine functions leaves search engine vulnerable to claims of copyright infringement. Whether these claims are spurious or well-founded is the central topic of this Article.

The four case-studies which follow provide a brief illustration of the vulnerability of copy-reliant technologies to claims of copyright infringement. These cases are briefly described in the sections that follow and then explored in more detail throughout the remainder of this Article.

1. Archiving Copyrighted Works – Field v. Google Inc.

In 2006, Blake Field, a Las Vegas personal injury attorney sued the Internet search giant Google for copyright infringement.[52] The basis of Field's claim was that Google had infringed his rights by allowing Internet users to access copies of his copyrighted works stored by Google's search engine cache.[53]

Search engines allow users to retrieve items from the cache for two main reasons. First, cached links enable Internet users to detect changes that have been made to a particular web-page over time. The differences such comparisons reveal can have important political, educational, and legal ramifications. Second, the availability of cached links enables users to understand why a seemingly irrelevant web-page was indicated as responsive to their original query. Although these functions relate to the copyrighted expression contained in the original website, they do not replicate the expressive function of the original. By definition, the use of a cached version of a web-page to detect changes is a use that could not be served by the original copyrighted work alone.[54] Likewise, referring to the cache to better understand the relationship between a particular page and a particular search term is also a use that could not be served by the original copyrighted work alone. [55]

Although it was not expressed in this terminology, the essence of the trial court's finding in *Field* was that the non-expressive use of the works in the cache did not interfere with the rights accorded to Field as an author.[56] This conclusion may have been underscored by the court's conclusion that Field had created the works in question by solely for the

[52] Field v. Google Inc., 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).
[53] *Id.*
[54] *Id.* at 1119.
[55] The Internet Archive is also subject to a similar set of copyright issues. The Internet Archive has amassed a collection of over 85 billion screenshots of web-pages which are stored on a computer database in California. These web-pages are available at no cost to the public via the "Wayback Machine". Similar to an Internet search engine, the Internet Archive uses a web crawler to routinely take screenshots of websites as they exist on various days. The Wayback Machine does not direct a user to a live website; instead, the user is presented with a static archived version of the website retrieved from the IA's database. The Wayback Machine is an invaluable tool for researchers, historians, and increasingly litigators, because it provides a record of the contents of a website that is independent of the website author. See e.g. Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey, 497 F. Supp. 2d 627 (E.D. Pa. 2007) (viewing and printing archived web-pages retrieved from the Wayback Machine was fair use).
[56] *Id.*

purposes of his lawsuit.[57] The court's conclusion was also based in part on its assessment that Field could have easily opted out of inclusion in the Google search engine if he had chosen to do so.[58] The significance of both of these rationales will be addressed below.

2. *Displaying Search Results – Perfect 10 v. Amazon*

Google's search engine technology was also at the core of another recent case, *Perfect 10 v. Amazon*.[59] Perfect 10 produces and sells copyrighted images of nude models. It does so primarily through a subscription website which sits behind a password protected paywall.[60] The material on Perfect 10's own website is neither indexed nor cached by the Google search engine; however, the search engine has no mechanism to exclude images republished by third parties without P10's authorization.[61]

To comprehend Perfect 10's objection to Google's image search engine, it is first necessary to understand how image based search technology differs from conventional text-based search technology. Instead of recognizing images themselves, image search software identifies text associated with objects identified as images. If the text associated with an image file is responsive to a user's search query, the search engine will display a small lower resolution "thumbnail" of the image in the search results. If an Internet user selects that thumbnail, the user's browser will be directed to retrieve the full-scale image from its original location.[62] One of Perfect 10's several theories of liability was that by producing these thumbnail-representations, Google was improperly copying Perfect 10's work without its authorization.[63] The district court initially agreed with Perfect 10 on this theory of infringement. However, the Court of Appeals for the Ninth Circuit held that the use of thumbnail-representations in an image based search engine did not constitute copyright infringement.[64] The essence of the court's reasoning – that Google's use of thumbnails "served a different function" unrelated to "artistic expression" – is consistent with the non-expressive use paradigm advanced in this Article.[65]

3. *Generating Meta-Data – The Google Book Project*

The third case-study relates to a different kind of search engine, one that is still under construction and whose very existence will be determined by legal arguments such as those addressed in this Article. Google's self appointed mission "to organize the world's information and make it universally accessible and useful" is not limited to that which is already in digital form.[66] Likened to the Library of Alexandria,[67] the Google Book Search

[57] *Id.* at 1114.
[58] *Id.* at 1119.
[59] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007).
[60] *Id.*
[61] *Id.* at 711.
[62] *Id.*
[63] *Id.*
[64] *Id.*
[65] *Id.* at 721 (citation and quotation omitted). *See also,* Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003).
[66] *Google Corporate Information: Company Overview*, at http://www.google.com/corporate/.

Library Project ("Google Book") aims to make the contents of over nine million books – the entire catalog of some of the world's most prestigious and extensive libraries[68] – searchable by anyone with an Internet connection.[69] To create this search engine, Google is currently in the process of digitizing vast collections of books, one page at a time.

Google Book is designed to allow users to search inside the text of captured books and to generate a list of books relevant to the user's search terms.[70] Google does not allow users to access the entire contents of any book, nor even an entire page of any book, unless the book is known to be in the public domain or the copyright owner has expressly agreed.[71] In the default scenario, a user who clicks on a book title is presented with bibliographic data about the target book and a small extract or "snippet" of the relevant page containing her search terms.[72] Users are also presented with additional information about the books targeted by their search term, including links to online bookstores and links to nearby libraries where the book can be obtained.[73]

Google Book's potential benefits to researchers are easily demonstrated. It takes just three clicks to go from the initial Google Book search screen to the call number of a specific and useful book in the University of Virginia Law library. For example, one might search for a basic statistical textbook discussing the limits of accepting the null hypothesis by entering the search term "accepting the null hypothesis."[74] Entering the search term generates a menu of books containing the term. Selecting any one book leads to a second set of information about the book including, snippets illustrating the relevance of the search term to the contents of the book, bibliographic information, links to reviews, links to references from web-pages, links to references from other books, and details of other editions.

---

[67] Brewster Kahle, *Speech to the Library of Congress in the Digital Future Series* (Dec. 13, 2004), available at http://www.archive.org/details/cspan_brewster_kahle.

[68] Bob Thompson, *Search Me?; Google Wants to Digitize Every Book. Publishers Say Read the Fine Print First.* WASHINGTON POST, August 13, 2006. The University of Michigan Library, Harvard University Library, Stanford University Library, [the University of California Library] and the New York City Public Library, Oxford University Library– the later two libraries will only allow copying of works known to be in the public domain. See, Author's Guild v. Google ___, Answer, para 32.

[69] [Collect citations on law review literature discussing Google Book]

[70] Author's Guild v. Google ___, Answer, para 19; McGraw-Hill v. Google ___, Answer, para 4.

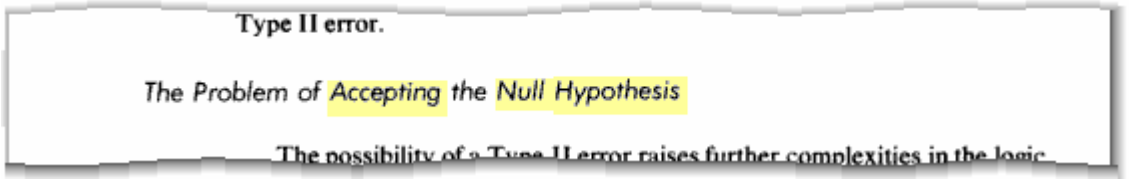
[71] *Id.* Google has several agreements with publishers to do just that. See, McGraw-Hill v. Google ___, Complaint, para 30, 31; McGraw-Hill v. Google ___, Answer, para 30. Amazon's Search Inside feature, also offers similar functionality for a much smaller collection of works for which they have been able to obtain permission from the relevant publishers. See Amazon.com's Response And Objections To Subpoena Served By Google, Inc (05-CV-8136 and 05-CV–8881).

[72] Google, *What you'll see when you search on Google Book Search*, at http://print.google.com/googleprint/screenshots.html.

[73] *Id.*

[74] Search conducted by the author on February 21, 2008 at 1.57 EST using the Google Book search engine at http://books.google.com.

Figure 1: Example of a Google Book "snippet"



The same screen also contains a menu of location options allowing the user to buy the book from online retailers, such as Amazon.com and Barnes&Noble.com, or to find the book in a lending library. A second click generates a list of libraries ranked in order geographic proximity. A third click actually retrieves the call number from, in this case, the University of Virginia Law Library.[75] In this fashion, Google Book allows users to sort vast volumes of information according to relevance and accessibility. Google Book will also provide information about books that are out-of-print or otherwise inaccessible to most of the public.[76] It might be hyperbolic to suggest that "all the books in the world [will] become a single liquid fabric of interconnected words and ideas",[77] but perhaps great advances in human knowledge deserve a little hyperbole.

Not everyone is so enamored with Google Book. Two significant lawsuits have been filed in relation to Google Book. The first is by the American Association of University Presses.[78] The second is a class action representing published authors and The Authors Guild.[79] Both suits seek declaratory and injunctive relief and money damages. The copyright challenge to Google Book is primarily focused on the way Google is building its search engine, rather than the output of the search engine *per se*.[80] The information contained in the search results of any one Google Book search is not by itself likely not infringe the copyright of any author for two reasons. First, most of the information Google Book generates falls into the category of uncopyrightable facts about books.[81] Second, even the snippets of material that Google directly copies from the print version a book will not amount to copyright infringement because the amounts taken are too fragmented and insignificant to meet the test of substantial similarity.[82]

---

[75] The book located in this example was *Science and Behavior: An Introduction to Methods of Research*, which contains a useful discussion of the problem of accepting the null hypothesis at page 149. JOHN M NEALE & ROBERT M. LIEBERT, SCIENCE AND BEHAVIOR, AN INTRODUCTION TO METHODS OF RESEARCH (2d) 149 (1980).

[76] See, Edward Wyatt, *Google Adds Library Texts to Search Database*, N.Y. TIMES, Nov. 3, 2005, at C11.

[77] Kevin Kelly, *Scan This Book!*, N.Y. TIMES, May 14, 2006, §6 (Magazine), at 42.

[78] The Association of American University Publishers on behalf of the McGraw-Hill Companies, Pearson Education, Penguin Group (USA), Simon & Schuster and John Wiley & Sons, also filed suit against Google on October 19, 2005. McGraw-Hill companies, Inc., et al. v. Google, Inc., Civila Action No. 05-CV-8881-JES (S.D.N.Y. Oct. 19, 2005).

[79] The Authors Guild filed a lawsuit in relation to Google's scanning and digitizing of library books on September 20, 2005. Author's Guild et al. v. Google, Inc., Civil Action No. 05-CV-8136-JES (S.D.N.Y. Sept. 20, 2005).

[80] Admittedly, the Author's Guild's Class Action Compliant is not so precise. See, Author's Guild et al. v. Google, Inc., First Amended Class Action Compliant, paras 3, 4.

[81] **See *infra* Part II-B.**

[82] **See *infra* Part II-B.**

However, the manner in which Google is building its formidable database presents more serious copyright issues. In the same way that Internet search engines routinely, automatically and indiscriminately copy html pages as part of the indexing process, the Google Book project requires the routine, automatic and indiscriminate copying of printed library books. Like the other search engine case-studies above, Google does not copy these literary works to disseminate a substantive amount of their expressive content to the public; but rather as grist for the search engine mill. Google Book is like the first two case studies in another important way: just like with its other search engines, Google has provided a method by which authors who do not want to have their works included in Google Book the ability to opt-out. The implications of both of these features are discussed in more detail in the remainder of this Article.

4. TurnItIn.com – Plagiarism Detection Software

A broad range of educational institutions have turned to technological solutions to combat the threat of plagiarism.[83] Harvard University,[84] the International Baccalaureate program[85] and thousands of high schools across the United States[86] use plagiarism detection software to detect and deter cheating by their students. Plagiarism detection services, such as Turnitin.com, detect improper and unaccredited copying in student papers by comparing new papers to an archive of material available on the Internet and to proprietary databases of previously submitted papers.[87]

This technology has obvious benefits for educators and for students. However, like other copy-reliant technologies, anti-plagiarism software also has its share of critics.[88] In 2006, students at McLean High School in Virginia objected when the school mandated the compulsory use of anti-plagiarism software.[89] The students took umbrage to both the implied accusation of cheating and to the fact that a commercial software company would

[83] Darby Dickerson, *Facilitated Plagiarism: The Saga Of Term-Paper Mills And The Failure Of Legislation And Litigation To Control Them*, 52 VILL. L. REV. 21 (citing various studies of academic integrity that show including a 1999 survey finding 50 percent of students admitted to Internet plagiarism).
[84] See, *Nation In Brief,* WASH. POST, November 3, 2006.
[85] See, S. Mitra Kalita, *Schools Turn to Software to Help Stop Plagiarism,* WASH. POST, April 15, 2004, p. T4.
[86] See, Andy Dehnart, *The Web's Plagism Police*, SALON.COM, June 14, 1999, http://www.salon.com/tech/feature/1999/06/14/plagiarism (reviewing several different services); Brock Read, "Anti-Cheating Crusader Vexes Some Professors", 54 CHRON. HIGHER ED. Issue 25, page A1, February 29, 2008. Available at http://chronicle.com/free/v54/i25/25a00101.htm; Maria Glod, *McLean Students Sue Anti-Cheating Service; Plaintiffs Say Company's Database of Term Papers, Essays Violates Copyright Laws,* WASH. POST, March 29, 2007, p B5.
[87] See, Turnitin website, *Proprietary Matching Technology*, http://turnitin.com/static/plagiarism.html.
[88] See e.g., Conference on College Composition and Communication, *CCCC-IP Caucus Recommendations Regarding Academic Integrity and the Use of Plagiarism Detection Services*, available http://ccccip.org/files/CCCC-IP-PDS-Statement-final.pdf (arguing somewhat incoherently that anti-plagiarism software undermine students' authority over the uses of their own writing and fosters an artificial view of originality and the role of imitation and borrowing in writing). See also Maria Glod, *Score One for McLean High Students; Administration Amends Anti-Cheating Policy After Protests,* WASH. POST, October 4, 2006, p. B1 (discussing the Conference on College Composition and Communication resolution).
[89] *Id.*

be able to make use of their works by adding them to a reference database.[90] Two McLean High School students followed up their protest with a copyright infringement lawsuit against iParadigms, the company that provides the Turnitin.com service.[91] The students sought a total of $900,000 in damages based on alleged copyright infringement of six term papers. At least one of the papers contained an express instruction that it was not to be archived.[92] Some other plagiarism detection services avoid similar disputes by allowing students to opt-out of inclusion in their reference databases. Nonetheless, like iParadigms, these services are still vulnerable to claims of copyright infringement in relation to the web-based material they incorporate into their services.[93]

Plagiarism detection services rely on access to entire copies of student term papers and any works from which they might be copied; yet the services do not necessarily cause any of the copyrighted content they process to be displayed to or read by human end-users.[94] As such, anti-plagiarism software also presents the paradox of non-expressive copying: copyrighted works are copied in their entirety in order to compute a result, but only the result itself contains none of the copyrighted expression of the original works.

II. THE DOCTRINAL IMPLICATIONS OF NON-EXPRESSIVE USE

A. The Principle of Non-Expressive Use

Copyrighted works are typically used to enable the expression contained in those works to enjoyed, appreciated or at least comprehended by some human actor. The enjoyment of watching a film, listening to music or reading a book is derived from the creative expression contained within those objects. We chose some films, songs and books over others because of the quality of their expression. It is convenient to think of these uses as "expressive" in that they relate to, and are motivated by, the expression embedded within a copyrighted work.[95] This observation, that expressive works are usually copied in contemplation of expressive uses, extends to partial copies as well. Because meaning is derived from context,[96] sampling a segment of music might change what that music expresses, but the end product is expressive in the general sense nonetheless.[97]

---

[90] Glod, *McLean Students Sue Anti-Cheating Service, supra* note 000.
[91] *Id.* A.V. v. iParadigms, 544 F. Supp. 2d 473 (2008).
[92] *Id.*
[93] The students' copyright claim was also weakened by the fact that the students themselves load their papers into the database after apparently accepting a click-wrap agreement. A.V. v. iParadigms, 544 F. Supp. 2d 473 (2008).
[94] See *infra* note 000 and accompanying text.
[95] To the extent that this definition of the "expressive use" of a copyright work departs from a conventional understanding, the reader should understand that it is employed herein as a term of art.
[96] *See*, Stanley Fish, *Normal Circumstances, Literal Language, Direct Speech Acts, the Ordinary, the Everyday, the Obvious, What Goes Without Saying, and Other Special Cases*, in IS THERE A TEXT IN THIS CLASS? 268-92 (1980) (arguing that words only have meaning because of their context); Richard Bandler and JOHN GRINDER, REFRAMING 2 (1982) ("In general communication theory there is a basic axiom that a signal only has meaning in terms of the frame or context in which it appears."); Wallace, *Only in the Context of a Sentence Do Words Have Any Meaning*, in CONTEMPORARY PERSPECTIVES IN THE PHILOSOPHY OF LANGUAGE 305 (P. French, T. Uehling, Jr., & H. Wettstein eds. 1979). See also, FDA v. Brown & Williamson Tobacco Corp., 529 U.S. 120, 132 (2000) ("The meaning–or ambiguity–of certain words or phrases may only become evident when placed in context."); Smith v. United States, 508 U.S.

The legal status of actual copying for non-expressive uses was not a burning issue before digital technology: there simply was no commercially relevant total literal copying that was not directed towards some expressive end. Digital technology and the increasing value of metadata (which is itself driven by digital technology) have combined to make the legality of non-expressive copying arguably the most significant issue in copyright law today.

In a world of analog works, non-expressive uses of copyrighted works are fairly uncontroversial. The metadata contained in library catalogs, topic indices or even plot synopses are unquestionably valuable. Nonetheless, because such uses do not typically involving copying the work in question, copyright owners have no legal right to object. By the same token, prior to digital technology, any instance of actual copying of the copyright owner's work could be assumed to be directed at some expressive end. Accordingly, the exclusion of facts and ideas from copyright subject matter was rarely important in cases of total copying – in an analog world it was almost inconceivable that someone could make a non-expressive use of a copyright work that involved physically copying the entire work. However, given the significant role of non-expressive copying in Internet search engines and other copy-reliant technologies, the legality of non-expressive copying is an issue that must now be addressed.

The purpose of this section is to demonstrate three related propositions: one descriptive, one normative, and one prescriptive. The descriptive proposition is that all of the copyright owner's exclusive rights are implicitly defined and limited in reference to expressive communication to the public. The normative proposition follows from the descriptive: acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. Nonetheless, in light of potential ambiguities in the application of the non-expressive use principle, a categorical rule that non-expressive copying is non-infringing may not be advisable. Accordingly, the prescriptive proposition advanced is that the non-expressive nature of the defendant's copying can and should be incorporated into a fair use analysis. As discussed in more detail below, these propositions are consistent with the goals of copyright generally and existing copyright doctrine.

223, 229 (1993) ("Language, of course, cannot be interpreted apart from context. The meaning of a word that appears ambiguous if viewed in isolation may become clear when the word is analyzed in light of the terms that surround it."); Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 Tex. L. Rev. 507 (2008) (criticizing the cognitive theory of trademark dilution).

[97] See, Andrew Ross, *Princes Among Thieves: Sampling in the 80s*, ARTFORUM, March 2003 at 249 (discussing the social meaning of sampling in American hip-hop music of the 1980s). See also Thomas G. Schumacher, *"This Is a Sampling Sport": Digital Sampling, Rap Music and the Law in Cultural Production*, 17(2) MEDIA, CULTURE AND SOCIETY 253, 268 (1995) (arguing that by facilitating the mixing of different voices in a musical text, sampling technology implicitly challenges "the concept of the singular artist as the only embodied voice in the text"); David Hesmondhalgh, *Digital Sampling and Social Inequality*, 15 SOCIAL & LEGAL STUDIES 53 (2006) (summarizing the literature and addressing the social and legal issues of music sampling as cultural "borrowing").

The Copyright Clause in the U.S. Constitution is expressly directed to the promotion of "the Progress of Science and useful Arts".[98] Copyright exists to encourage the creativity of authors and to promote the creation and dissemination of information.[99] As the Supreme Court has noted on a number of occasions, the promotion of science and the useful arts requires a balance between "the interests of authors and inventors in the control and exploitation of their writings and discoveries on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand."[100] Where that balance is struck dictates what the public can copy and what authors can control. Just as importantly, it also mediates relationships between different generations of authors: initial authors and those who build upon their works.[101] Thus, while copyright aims to give authors an incentive to create and share their works, it also strives to provide subsequent authors with sufficient "breathing space" to make their own additive contributions.[102] The copyright system is predicated both on the existence of certain rights to protect authors from unfair competition, and on significant gaps in those rights that give other authors freedom to breath.

Viewing copyright in terms of the communication of the expressive elements of the author's work is consistent with both economic and rights-based understandings of copyright. For the economist, copyright creates certain exclusive rights to give authors an incentive to invest in the creation of works that would otherwise be freely copied. Copyright protection thus allows authors to internalize more of the benefits of their creations and thus makes them more likely to want to create in the first place.[103] The natural rights argument for copyright is primarily an extension of the Lockean Framework of labor as the basis of property ownership to intangibles.[104] However, this justification for property does little by itself to establish either its form or its limitations.[105] "Personhood" provides an alternative non-utilitarian view of copyright, the premise being that "property provides a unique or especially suitable mechanism for self-actualization, for personal expression, and for dignity and recognition as an individual

[98] U.S. CONST. ART. I, § 8, cl. 8.
[99] Eldred v. Ashcroft, 537 U.S. 186 (2003).
[100] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984).
[101] *See generally,* Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989 (1997) (discussing sequential innovation in copyright and patent law).
[102] MGM Studios Inc. v. Grokster, Ltd., 545 U.S. 913, 933 (2005); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 479 (1984) ("The fair use doctrine must strike a balance between the dual risks created by the copyright system: on the one hand, that depriving authors of their monopoly will reduce their incentive to create, and, on the other, that granting authors a complete monopoly will reduce the creative ability of others.")
[103]
[104] See JOHN LOCKE, TWO TREATISES OF GOVERNMENT bk. II 27 (Peter Laslett ed., 1988) (As Locke famously argued, "[e]very Man has a Property in his own Person. This no Body has any Right to but himself. The Labour of his Body, and the Work of his Hands, we may say, are properly his. Whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joyned to it something that is his own, and thereby makes it his Property.")
[105] For two quite different views of Locke's implications for intellectual property, see Wendy J. Gordon, *A Property Right in Self-Expression: Equality and Individualism in the Natural Law of Intellectual Property*, 102 YALE L.J. 1533 (1993) (arguing that natural rights theory is necessarily concerned with the rights of the public as well as with the rights of those whose labors create intellectual products). Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287 (1988).

person." [106] In either case, the guiding principle of copyright is that one should not generally be entitled to offer the author's copyrighted expression to the public as a substitute for the work of the author.

Copyright consists of a bundle of discrete exclusive rights, such as the reproduction right, the derivative right, and the public performance and display rights.[107] These rights are defined, articulated and limited by a number of initially judge-made doctrines, such as the idea-expression distinction, the threshold of substantial similarity and the fair use doctrine.[108] As this section explores in more detail below, these doctrines each limit copyright protection to the expressive aspects of original works of authorship in a way that confirms the centrality of communication to the public at the heart of copyright.

Copyright's focus on expressive substitution is also evident in the exclusion of non-expressive elements from copyright subject matter itself. Nonetheless, the centrality of expressive substitution does not rest on the idea-expression distinction alone. A number of other significant copyright doctrines also demonstrate that communication to the public is the touchstone of copyright infringement and that no copyright liability should be found without such expressive communication. In particular, the communication of original expression to the public defines the metes and bounds of the publisher's collective right in section 201(c) of the Copyright Act; it defines the threshold of substantial similarity which is the test of copyright infringement; furthermore it explains why courts exclude unpublished drafts from copyright liability altogether.

1. The Exclusion of Non-Expressive Elements from Copyright Subject Matter

Copyright in an expressive work does not confer any exclusive rights in the facts, ideas, concepts, or discoveries contained in that work, regardless of the form in which they are described, explained or illustrated in such a work.[109] This principle, often simply abbreviated to the "idea-expression distinction," is longstanding at common law and was expressly incorporated into the Copyright Act in the 1976 revision.[110]

At least since *Baker v. Selden* in 1879, courts have recognized that "there is a clear distinction between the book, as such, and the art which it is intended to illustrate."[111] The distinction holds even in those unusual cases where the true value of the work lies in the methods, systems, and ideas it discloses, rather than in the way those concepts are

---

[106] Hughes, *Id.* at 330. See generally, Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957 (1982) ("to achieve proper self-development – to be a *person* – an individual needs some control over resources in the external environment.")

[107] 17 U.S.C. §§ 106(1) – (6).

[108] The idea-expression distinction and the fair use doctrine are also reflected in the sections 102(b) and 107 of the Copyright Act of 1976, however, they remain essentially common law features of the copyright system.

[109] Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 547 (1985) (holding that "no author may copyright facts or ideas"); 17 U.S.C. § 102(b).

[110] 17 U.S.C. § 102(b) provides: "In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such a work."

[111] Baker v. Selden, 101 U.S. 99, 102 (1879).

expressed.[112] For example, in the *Selden* case itself, the plaintiff had developed a novel and useful method of bookkeeping, the practice of which created value regardless of how the method was communicated.[113] Nonetheless, the plaintiff's copyright in his instructional material was limited to the expression of his useful methods and did not encompass those methods themselves.[114] Of course, in most cases, protecting the unique expression of an idea is sufficient to ensure that the author will be able to appropriate a return on her investment.

Copyright law also clearly distinguishes between facts and their expression, providing no protection for the former and only limited protection for the latter.[115] In *Feist v. Rural Telephone*, the Supreme Court ruled that copying listings form a telephone directory did not infringe the copyright in that directory because the information itself was not copyrightable. As the Court explained facts — whether they be telephone numbers and addresses or the details of historical occurrences — are not "original" to the author.[116] The author's copyright is limited to her expression of those facts, not the facts themselves.[117] The *Feist* Court further held that the selection and arrangement of that information was also not copyrightable because the organization of the telephone directory was "so mechanical or routine as to require no creativity whatsoever."[118]

Through the idea-expression distinction, copyright law protects the expressive elements of the author's work while guaranteeing subsequent authors the necessary breathing space to make their own contributions by adding to, re-using, or re-interpreting the facts and ideas embodied in the original work. Subsequent authors may not compete with the copyright owner by offering her original expression to the public as a substitute for the copyright owner's work, but they are free to compete with their own expression of the same facts, concepts and ideas. Accordingly, the idea-expression distinction is a central element of the balance between the interests of authors in the exploitation of their writings and society's competing interest in the free flow of ideas, information, and commerce.[119]

---

[112] *Id.* Note also that the copyright protection available for maps is somewhat thin as a result. See, 1-2 Nimmer on Copyright § 2.08 [A] and the cases cited therein.

[113] Seldon's patent application may well have been patentable under today's standards. *See*, State Street Bank & Trust Co. v. Signature Financial Group, Inc., 149 F.3d 1368, 1373 (Fed. Cir. 1998)). *But see*, Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 548 U.S. 124 (2006) (Breyer dissenting).

[114] *Id.*

[115] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340. (holding that facts are not copyrightable and that the copyright in a factual compilation is thin).

[116] *Id.* at 348 ("Copyright protection may extend only to those components of a work that are original to the author").

[117] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 556 (1985) ("No author may copyright his ideas or the facts he narrates.")

[118] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 380 (holding that the selection, coordination, and arrangement of Rural's white pages did not satisfy the minimum constitutional standards for copyright protection). See also, Matthew Bender v. West Publishing Co., 158 F.3d 674 (2d Cir. 1988); Assessment Techs. of WI, LLC v. WIREdata, Inc., 350 F.3d 640 (7th Cir. Wis. 2003).

[119] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984). See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 240 (2d Cir. 1983) (describing the idea-expression distinction as "an effort to enable courts to adjust the tension between these competing effects of copyright protection.")

The exclusion of facts and ideas from the ambit of copyright protection, applies with equal force to non-expressive copying in the digital age. In spite of the fact that metadata is increasingly valuable in the information age, it is no more copyrightable than it was 100 years ago. The undisputed value of individual facts, such as the title of book or its location in a library, does not change the copyright status of those facts. As a general rule, metadata is not subject to copyright protection: one can extract and reproduce facts, names and dates from a news paper article, or ideas and processes from an instructional text, without infringing the author's copyright.[120] Whether congress should, or even could, alter the traditional contours of copyright by extending copyright protection to facts and ideas is a worthy topic of debate – the fact remains that as of now, it has not.[121]

The idea-expression distinction is a central element of copyright's balance between the interests of authors in the exploitation of their writings and society's competing interest in the free flow of ideas, information, and commerce. The idea-expression distinction limits the rights of the copyright owner to the expressive elements of the author's work: in the analog context this is achieved by simply declining to hold that the copying of facts and ideas alone amounts to infringement. Preserving the functional force of idea-expression distinction in the digital context requires a slightly different application: copying for purely non-expressive purposes, such as the automated extraction of data, should not be regarded as infringing.

2. The Collective Work Right

The collective work right also demonstrates that communication to the public is the touchstone of copyright infringement. The Copyright Act gives authors the exclusive right to reproduce their works in copies; however, the Act also confers a special privilege on the owners of collective works, such as magazines and newspapers, which allows them to reproduce and distribute individual contributions as part of the collective work and revisions thereof.[122] The collective work right creates an apparent conflict with the general reproduction right by allowing magazines and newspapers to reproduce the works of individual authors without their consent in certain circumstances. That conflict came to a head in the 2001 case of *New York Times v. Tasini.*[123] In that case, six freelance authors sued a group of publishers, including the New York Times, for allowing articles written by the authors to be placed in electronic databases without the authors' consent. The publishers relied on their 'privilege', contained in section 201(c) of the Copyright Act, to reproduce and distribute the freelance author's contributions as part of a revision to a collective work.[124]

The Supreme Court's resolution of the conflict between the general reproduction right and the collective work right confirms the centrality of public perception and expressive

[120] Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991); Baker v. Selden, 101 U.S. 99 (1880).
[121] See generally, Jonathan Band & Makoto Kono, *The Database Protection Debate in the 106th Congress*, 62 OHIO ST. L.J. 869 (2001).
[122] 17 U.S.C. § 201 (c) Contributions to collective works.
[123] N.Y. Times Co. v. Tasini, 533 U.S. 483 (2001).
[124] *Id.*

communication to the public in determining the rights of the copyright owner. In *Tasini*, the Supreme Court rejected the New York Times' broad construction of its collective right, holding that because the articles in question were "presented to, and retrievable by, the user in isolation, clear of the context the original print publication" they did not qualify as part of a revision to the original collective work.[125]

The defendants in *Tasini* had argued that their conversion of printed back-issues to an electronic form amounted to revision of the collective work, and was thus sheltered under § 201(c) of the Act. From the New York Times' perspective, electronic storage was no different to the conversion of newsprint to microfilm – in either case the entire issue was archived exactly as initially printed to facilitate later retrieval of specific articles. However, the Court held that the form of storage simply did not matter, given that users did not perceive the articles as stored but only as retrieved by the New York Times' database.[126] Unlike microfilm files, the database presented the individual articles to the user devoid of their initial context. The Court's view was that only user perception mattered and that the fact that the articles were stored in their initial sequence was irrelevant to both readers and authors alike. The Court thus held that "[i]n determining whether the Articles have been reproduced and distributed as part of a revision of the collective works in issue, we focus on the Articles as presented to, and perceptible by, the user of the Databases."[127]

Although *Tasini* is not a non-expressive use case, it nonetheless supports the proposition that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. By defining the scope of the publishers' collective works privilege in terms of that which is communicated to the public and dismissing the relevance of unseen uses within the defendants' databases, the Court reinforces expressive communication to the public is the touchstone of copyright infringement.

3. Substantial Similarity

The centrality of the expressive communication to the public is inherent in the tests applied by the courts to determine the threshold of infringement: i.e. the tests that determine when some copying becomes too much copying. As discussed in more detail below, the application of the test of substantial similarity further demonstrates that copying which does not interfere with the exclusivity of the copyright owner's communication of her work to the public does not infringe the exclusive rights of the author.

---

[125] *Id.* at 488.
[126] *Id.* at 504 ("The crucial fact is that the Databases, like the hypothetical library, store and retrieve articles separately within a vast domain of diverse texts. Such a storage and retrieval system effectively overrides the Authors' exclusive right to control the individual reproduction and distribution of each Article.") (citations omitted)
[127] *Id.* at 499.

The copyright owner's exclusive right to "reproduce the work in copies" extends to both exact and inexact reproductions.[128] In both cases, however, the Copyright Act leaves the threshold of reproduction – the question of how much of the copyrighted work must be copied – undefined. In cases of "non-literal infringement", i.e. where the accused work is not an exact copy of the copyright owner's work, courts assess whether the allegedly infringing work possesses a "substantial similarity" to the copyrighted work.[129]

The threshold of substantial similarity is defined in reference to the perspective of the ordinary observer.[130] Infringement is defined in reference to the perspective of the consuming public, because the copyright owner's "legally protected interest is not, as such, his reputation . . . but his interest in the potential financial returns from his [work] which derive from the lay public's approbation of his efforts."[131] As such, the tests of substantial similarity provide further evidence that copyright primarily protects the author against expressive substitution.

Courts also apply the threshold of substantial similarity in cases of fragmented actual copying, such as in music sampling or collage.[132] In *Newton v. Diamond*, for example, the plaintiff alleged that the Beastie Boys had infringed his copyright in a musical composition by including a six second sample of a licensed sound recording in their own musical creation, "Pass the Mic".[133] The Beastie Boys had obtained a license with respect to the sound recording, but had not thought it necessary to seek a license from the composer given the parsimonious nature of the composition.[134]

Where the defendant copies a portion of the plaintiff's work exactly or nearly exactly, without appropriating the work's overall essence or structure, the courts apply a test of "fragmented literal similarity" to determine if the substantial similarity threshold has been met.[135] In cases of fragmented literal similarity, courts determine whether the

[128] 17 U.S.C. § 106(1); Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir, 1930).
[129] See, Laureyssens v. Idea Group, Inc., 964 F.2d 131, 140 (2d Cir. 1992); Tufenkian Import/Export Ventures, Inc. v. Einstein Moomjy, Inc., 338 F.3d 127, 131 (2d Cir. 2003).
[130] See, Shine v. Childs, 382 F. Supp. 2d 602, 614 (S.D.N.Y. 2005) (summarizing authorities).
[131] Arnstein v. Porter, 154 F.2d 464, 473 (2d Cir. 1946) (footnotes omitted); See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 240 (2d Cir. 1983).
[132] See, Newton v. Diamond, 388 F.3d 1189, 1195 (9th Cir. 2004) (holding that "the substantiality requirement applies throughout the law of copyright"). As David Nimmer notes, the Sixth Circuit's *Bridgeport* decision suggests otherwise, however, that decision is almost certainly in error on this point. See 4-13 Nimmer on Copyright § 13.03.
[133] Newton v. Diamond, 388 F.3d 1189 (9th Cir. 2004).
[134] The sample corresponds to three notes on the original composition, C - D flat - C, over a held C note. The score to "Choir" also indicates that the entire song should be played in a *largo/senza-misura* tempo, meaning "slowly or without-measure." *Id.* 1191. Note that sound recordings and their underlying compositions are separate and distinct copyrighted works. 17 U.S.C. § 102(a)(2), (7).
[135] As the Second Circuit explained in *Twin Peaks,* "the concept of similarity embraces not only global similarities in structure and sequence, but localized similarity in language. In both cases, the trier of fact must determine whether the similarities are sufficient to qualify as substantial." Twin Peaks Prods. v. Publ'ns Int'l, Ltd.,996 F.2d 1366, 1372 (2d Cir. 1993). See also, Palmer v. Braun, 287 F.3d 1325, 1330 (11th Cir. 2002) ("[T]he work may copy only a small part of the copyrighted work but do so word-for-word. If this fragmented copy is important to the copyrighted work, and of sufficient quantity, then it may support a finding of substantial similarity.")

copying amounts to infringement "by considering the qualitative and quantitative significance of the copied portion in relation to the plaintiff's work as a whole."[136] Applying this test to the Beastie Boys appropriation of a fragment of Newton's original musical composition, "C - D flat - C, over a held C note", the court found that "no reasonable juror could find the sampled portion of the composition to be a quantitatively or qualitatively significant portion of the composition as a whole."[137]

This focus on the qualitative and quantitative significance of the copied portion in the plaintiff's work is consistent with the prohibition against expressive substitution. Even where some of the copyright owner's original expression has been copied directly, such copying does not rise to the level of infringement unless the expression was significant, in either quantity or quality, in the author's original work. Just as copyright law does not prevent the copying of facts and ideas; it also permits copying of trivial expressive elements from an existing work, because to do so does not unfairly compete with the copyright owner.[138] In other words, trivial copying of expressive elements is not copyright infringement because it does not interfere with the copyright owner's exclusive right to communicate her work to the public.

The law relating to fragmented literal similarity not only shows that the copyright owner's exclusive rights are implicitly defined and limited with respect to communication of some expression to the public. It also demonstrates that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement.

4. Allegations of Intermediate Copying in Hollywood

The dismissal of the seemingly routine allegations of copyright infringement that accompany the release of major motion pictures also illustrates that no copyright liability should be found without expressive communication to the public. In the screen-play infringement cases discussed below, and many others, courts have refused to entertain requests for discovery with respect to drafts of a non-infringing final work because only the final product released to the public is capable of infringing the copyright owner's rights.

Meritless claims of copyright infringement are a recognized cost of doing business in Hollywood.[139] Some of these claims are merely opportunistic, others appear to be

[136] Newton v. Diamond, 388 F.3d 1189, 1195 (9th Cir. 2004) (citing Worth v. Selchow & Righter Co., 827 F.2d 569, 570 n. 1 (9th Cir. 1987); Jarvis v. A&M Records, 827 F. Supp. 282, 289-90 (D.N.J. 1993); 4 Nimmer § 13.03[A][2], at 13-47 to 48).
[137] Newton v. Diamond, 388 F.3d 1189, 1195 (9th Cir. 2004).
[138] Id. at 1193 ("The principle that trivial copying does not constitute actionable infringement has long been a part of copyright law."); *Id*. at 1195 ("the dispositive question is whether the copying goes to trivial or substantial elements.")
[139] Matthew Belloni, *THR Esquire*, HOLLYWOOD REPORTER, June 26, 2007 ("Like expensive CGI and flashy premieres, defending copyright lawsuits by writers who think their screenplays have been ripped off is just another cost of doing business for studios."); Verne Gay, *Flash!: The Latest Entertainment News and More...,* NEWSDAY, Jan. 30, 1998, at A12 (In thanking the studios in defending against a claim that his screenplay for the movie *Twister* had stolen from another screenplay, Michael Crichton said, "I hope it will

motivated by the plaintiff's genuine belief that all his or her own ideas are unique and that there are no coincidences. *Madrid v. Chronicle Books* is representative of the phenomenon. In that case, the author of a one-page poem about a land of monsters who are afraid of human children alleged that the Pixar film *Monsters, Inc.*, infringed her copyright.[140] The court, in contrast, held that the inverted plot of monsters afraid of children was generic.[141] Some of these cases involve similar themes,[142] others involve similar descriptive titles applied to the same general subject,[143] and several others involve a similarity discernable only to the plaintiffs themselves.[144]

Confronted with motions for summary judgment, plaintiffs often urge the courts to allow them to scrutinize every single draft of the defendant's screen play in the hope that some earlier version of the work will disclose a greater resemblance to their own copyrighted work than the finished film did.[145] These requests are invariably denied.[146] The reasons why they are refused provide an important insight into the structure of copyright law.

---

usher in a new era where studios fight these frivolous charges and don't treat it as a cost of doing business.")

[140] Madrid v. Chronicle Books, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002).

[141] *Id.* Two antecedents spring immediately to mind. First, E.T. hiding in the cupboard from Elliot, *E.T.: The Extra-Terrestrial*, Amblin Entertainment (1982). Second, Max's dominion over the fearful monsters in *Where The Wild Things Are*. See, MAURICE SENDAK, WHERE THE WILD THINGS ARE (1963).

[142] In *Litchfield v. Spielberg*, in which the author of a musical play about two aliens stranded at the north pole accused the producers of the motion picture, *E.T., – The Extra Terrestrial*. Litchfield v. Spielberg, 736 F.2d 1352 (9th Cir. 1984) (finding no substantial similarity between the sequences of events, mood, dialogue and characters of the two works). In *Madrid v. Chronicle Books* the author of a one-page poem about a land of monsters who are afraid of human children alleged that the Pixar film Monsters, Inc., infringed her copyright, Madrid v. Chronicle Books, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002). See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 235 (2d Cir. 1983) (finding that the protagonist in the television series, *The Greatest American Hero*, was not sufficiently similar to the D.C. Comic's *Superman* character to warrant consideration of the plaintiff's copyright infringement claim by a jury.) *Id.* at 243 ("In the genre of superheroes, Hinkley follows Superman as, in the genre of detectives, Inspector Clouseau follows Sherlock Holmes.")

[143] In *Davis v. United Artists, Inc.,* the author of the 1972 Vietnam novel entitled "Coming Home*"* claimed copyright infringement in relation to the 1978 Vietnam film, also titled "Coming Home". Davis v. United Artists, Inc., 547 F. Supp. 722 (S.D.N.Y. 1982) (finding no similarity between the two works). In *Walker v. Time Life Films, Inc.,* the author of the autobiographical policeman's tale, *Fort Apache* alleged that the Time Life film, *Fort Apache, The Bronx* amounted to copyright infringement. Walker v. Time Life Films, Inc., 615 F. Supp. 430, 435 (S.D.N.Y. 1985) (holding that no reasonable observer could find substantial similarity and that any similarity that may exist is either trivial, abstract or non-protectible as a matter of law).

[144] For example, in *Stromback v. New Line Cinema*, the author of the disturbing screen-play outline about a callous reporter who brings down a corrupt governor entitled, *The Keeper,* accused the writers of the film, *Little Nicky,* of copyright infringement. *Little Nicky* is a comedy about the Devil and three sons, one of whom has a speech impediment and is played by Adam Sandler. Stromback v. New Line Cinema, 384 F.3d 283 (6th Cir. 2004) (finding no similarity between the works other than at the most superficial level). Equally incomprehensible is the claim in *Flaherty v. Filardi,* in which the producers of *Bringing Down the House*, an odd couple film about a lonely tax attorney who meets a woman on the Internet who unknown to him happens to be in prison, were alleged to have infringed the copyright the screen play screenplay *Amoral Dilemma*, the rather grim story of a disaffected young Manhattan insurance attorney who knowingly corresponds with a death row prisoner. Flaherty v. Filardi, 2007 U.S. Dist. LEXIS 69202, 8-9 (S.D.N.Y. 2007).

[145] Walker v. Time Life Films, Inc., 615 F. Supp. 430, 434 (S.D.N.Y. 1985) (request to discover drafts denied). Stromback v. New Line Cinema, 384 F.3d 283 (6th Cir. 2004). Flaherty v. Filardi, 2007 U.S. Dist.

Courts refuse to entertain discovery with respect to drafts of a non-infringing final work precisely because infringement requires at least some potential interference with the copyright owner's expectation of exclusivity. As noted in *Davis v. United Artists,* "the ultimate test of infringement must be the film as produced and broadcast, we do not consider the preliminary scripts."[147] Courts do not refuse to examine interim drafts merely because of judicial economy; as the Second Circuit noted in *Warner Bros., Inc. v. American Broadcasting Cos.,* "a defendant may legitimately avoid infringement by intentionally making sufficient changes in a work which would otherwise be regarded as substantially similar to that of the plaintiff's."[148]

The refusal of courts to entertain copyright infringement allegations in relation to unpublished drafts and preliminary scripts demonstrates the practical importance of the principle that expressive substitution lays at the heart of copyright infringement. Because the copyright owner's rights are generally limited to the communication of their original expression to the public, a film maker is perfectly entitled to start with Jane Austen's *Emma* and rework the plot over and over again until she comes out with *Clueless*.[149] Intermediate scripts that never see the light of day do not communicate the author's original expression to the public and thus cannot constitute copyright infringement.

***

This section has confirmed the centrality of expressive substitution to a variety of doctrines and applications: the idea-expression distinction, substantial similarity and the collective work right, and finally, the refusal of courts to entertain infringement actions solely based on unpublished screen-play drafts. Communication to the public is the touchstone of copyright infringement and that no copyright liability should be found without such expressive communication. Once the centrality of centrality of expressive substitution to copyright infringement is properly understood, the implications for copy-reliant technologies become clear: the non-expressive use of a copyrighted work should not result in copyright infringement.

LEXIS 69202, 8-9 (D.N.Y. 2007) (copyright claim to interim drafts of a published non-infringing final work dismissed as a matter of law).

[146] See, e.g., Walker v. Time Life Films, Inc., 615 F. Supp. 430, 434 (S.D.N.Y. 1985) (noting that courts routinely reject requests to consider earlier drafts of screenplays).

[147] Davis v. United Artists, Inc., 547 F. Supp. 722, 724 n.9 (S.D.N.Y. 1982) (citing Fuld v. National Broadcasting Co., Inc., 390 F. Supp. 877, 882 n.4 (S.D.N.Y. 1975). See also, Walker v. Time Life Films, Inc., 615 F. Supp. 430, 435 (S.D.N.Y. 1985) ("The Court considers the works as they were presented to the public."); Madrid v. Chronicle Books, Pixar, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002)("Since a court considers the works as they were presented to the public, discovery in this case … would be pointless."); Stromback v. New Line Cinema, 384 F.3d 283, 299 (6th Cir. 2004). ("In deciding infringement claims, courts have held that only the version of the alleged infringing work presented to the public should be considered.")

[148] Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 241 (2d Cir. 1983)(citing 3 Nimmer § 13.03[B] at 13-38.1 to -38.2, Eden Toys, Inc. v. Marshall Field & Co., 675 F.2d at 501; Durham Industries, Inc. v. Tomy Corp., 630 F.2d at 913 & n.11.)

[149] CLUELESS (Paramount 1995).

B. Doctrinal Incorporation of Non-expressive Use

As discussed above, the principle of non-expressive use – that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement – flows naturally from an analysis of existing copyright doctrines. Just as authors possess no copyright in the facts and ideas contained within their works, the rights of authors to control the copying of their works should not include copying that is non-expressive in nature. As the doctrines surveyed above clearly demonstrate, authors possess a set of limited and largely economic rights to control the expressive uses of their works. Extending those rights to encompass non-expressive uses would constitute a significant departure from existing copyright principles.

As set forth in more detail below, the copying at issue in the *Field*, *Perfect 10*, *Google Book* and *Turnitin.com* case-studies is manifestly different to the usual copyright scenario. In all four cases the entire works were copied, but the purpose of that copying was not to convey the work's expressive qualities to the public, but rather to enable banks of microprocessors to index the content of those works and to generate meta-data about the works. Explicit recognition of this principle of non-expressive use would significantly clarify the legal status of copy-reliant technologies such as Internet search engines, plagiarism detection software and the Google Book project.[150] However, acknowledging the principle of non-expressive use raises the subsidiary question of how the principle should be implemented.

Consistent with the traditional contours of copyright, it is certainly open to a court to rule that the use of a copyrighted work that does not communicate its expressive content to the public is *per se* non-infringing. However, there are also reasons why courts might hesitate to adopt such a categorical rule. First, as this section explores, potential ambiguities in the concept of non-expressive use suggest that adopting a categorical rule that non-expressive uses are non-infringing may simply shift the focus of argument from substantive questions to questions of category definition. Second, as the next section explains in greater detail, courts may wish to hesitate before adopting a categorical rule in relation to non-expressive use because the same considerations that would determine whether a use was non-expressive are already a significant part of fair use analysis.

Although the principle of non-expressive use articulated in this Article is important, it is not free from ambiguity. The extraction of factual information – such as names dates and places – is a non-expressive use, in that it does not relate to the expression of these facts, but to the facts themselves.[151] Similarly, generating factual information about a work should also be categorized as a non-expressive use of the underlying work. For example, publishing the fact that the novel *Moby Dick* was written by Herman Melville in 1851 and contains the word "whale" 783 times would not infringe any copyright in the book

[150] See *infra* notes ___ to ___ and accompanying text.
[151] *See* 17 U.S.C. § 102(b) ("In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.")

because this information about the work is independent of the expressive value of the work.[152]

In relation to copy-reliant technologies, plagiarism detection software illustrates one of the clearest applications of the non-expressive use principle. Plagiarism detection services rely on access to entire copies of student term papers and any works from which they might be copied; yet the services do not necessarily cause any of the copyrighted content they process to be displayed to or read by human end-users. Plagiarism detection software works by comparing strings of text in new works to strings of text in existing works. [153] If a match is found, the software indicates as much. By itself, the report that a new work is similar to another work already in the database in no way reproduces or communicates the expressive qualities of either work. In practice, plagiarism detection providers also issue reports identifying the text allegedly copied and the source document, however, the basic matching function can be performed with no communication of expression at all. Thus, in its basic function at least, anti-plagiarism software is paradigmatic non-expressive use.

Nonetheless, many other significant copy-reliant technologies present some degree of ambiguity as to whether they should be regarded as expressive or non-expressive. In both *Field* and *Perfect 10*, the courts effectively found that the primary purpose of the copying at issue was non-expressive. In *Field*, the court held that although allowing html pages to be retrieved from the search engine cache also allowed them to be read, the primary use of the cache was non-expressive and thus non-infringing.[154] The court found that to the extent that Google itself copied or distributed Field's copyrighted works by allowing access to them through cached links, Google had engaged in a fair use of those copyrighted works. The court's fair use analysis relies heavily on the differences between Google's use of the works and any expressive or artistic value that Field's work might have otherwise had. Although the court does not employ the terminology applied in this Article, it is clear that what made Google's use different in the relevant sense was the fact that it was non-expressive. The court gave a number of reasons why Google's use of the works in the form of cached links did not serve the same function as the original works. Primarily, the court noted that cached links enable Internet users to detect changes that have been made to a particular web-page over time – changes which may have important

[152] *Moby Dick* is in the public domain in the United States and is available at Project Guttenberg at http://www.gutenberg.org/etext/2701.

[153] The similarities between two works can be assessed by simply looking for common strings of words. See, Amy Argetsinger, *Technology Snares Cheaters at U-Va.; Physics Professor's Computer Search Triggers Investigation of 122 Students*, WASH. POST, May 9, 2001, pA1. However, there are also various algorithms that can be applied to a document to create a digital fingerprint which captures other characteristics of the work. These digital fingerprints allow a document to be characterized by its structure, vocabulary and content; they are essentially abstractions of the original documents and allow for faster comparisons which will not be as easily deceived by minor text alterations. *See e.g.,* Khair Eddin M. Sabri & Jubair J. Al-Ja'afer, *The JK System to Detect Plagiarism,* 6(2) JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY, 66 (2006). The Turnitin software uses statistical techniques originally designed to analyze brain waves to compare the fingerprints of student papers to more than a billion documents that have been fingerprinted in a similar fashion. *See,* THE ECONOMIST, *Plagiarise. Let No One Else's Work Evade Your Eyes*, March 16, 2002 (U.S. Edition).

[154] Field v. Google Inc., 412 F. Supp. 2d 1106 (D. Nev. 2006)

political, educational, or legal ramifications.[155] As the court notes: "by definition, this information location function cannot be served by the original web-page alone. To conduct such a comparison, a user would need to access both Google's archival copy of a Web-page and the current form of the web-page on the Internet."[156] In addition, the court also noted that the availability of cached links enables users to understand why a seemingly irrelevant web-page was indicated as responsive to their original query.[157]

Although these functions relate to the copyrighted expression contained in the original website, they do not replicate the expressive function of the original. Axiomatically, the use of a cached version of a web-page to detect changes is a use that could not be served by the original copyrighted work alone.[158] Likewise, referring to the cache to better understand the relationship between a particular page and a particular search term is also a use that could not be served by the original copyrighted work alone.

Perfect 10 presents a similar ambiguity and a similar resolution. In *Perfect 10*, the thumbnail-representations were clearly visible to the public; however the Court of Appeals found that the thumbnails did not fulfill a demand for the originals as expressive works.[159] The court held that the plaintiff had raised a *prima facie* case of infringement because Google's thumbnail-representations were literally copied from Perfect 10's works and were displayed by the search engine.[160] However, the court found that Google's creation of thumbnail-representations did not infringe Perfect 10's rights, in large measure because the copying the plaintiff complained of was directed to a different use, a non-expressive use.[161] The essence of the Court of Appeals' decision with respect to thumbnails was that the use of the pictures by Google as a pointing device must be distinguished from the use of the pictures to fulfill the demand for the pictures as expressive works of themselves. In the court's words, Google's replication of the original works as thumbnails "served a different function" unrelated to "artistic expression."[162] As the court explained: "[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information."[163] In other words, although the thumbnail-representations were technically a copy of Perfect 10's original works, they were not used to fulfill the public's demand for small grainy photos of unclad women, but rather as pointing devices to instruct users where they might find the photos they are looking for.

In both *Field* and *Perfect 10* there was at least the technical, if unlikely, possibility that the search engine copying could function as an expressive substitute for the copyright

[155] Field v. Google Inc., 412 F. Supp. 2d 1106 (D. Nev. 2006).
[156] *Id.* at 1119.
[157] *Id.*
[158] *Id.*
[159] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007)
[160] *Id.* at 719.
[161] *Id.* at 725. (reversing the district court's ruling that the use of thumbnails was not fair use).
[162] *Id.* at 721 (citation and quotation omitted). *See also,* Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003).
[163] *Id.* at 721.

owners' original works. Nonetheless, in both cases the courts found that the copying at issue did not fulfill a demand for the originals as expressive works.[164] In *Field*, the court found that the mere technical possibility that someone might recall an object from the cache to enjoy its expressive qualities was insufficient to characterize caching in general as an expressive use of copyright works given that its predominant uses – verifying the integrity of search results, checking the authenticity of documents and confirming the date that they were actually posted – were unrelated to the expressive function of the original works.[165] In *Perfect 10*, the court acknowledged the possibility that some users might see the thumbnail representations as substitutes for the originals; however, despite this possibility, the court dismissed the plaintiff's claim of expressive substitution as speculative and unlikely.[166] In the court's opinion, because the thumbnail-representations were used by the image search engine to show users which websites contained images relevant to their search terms, they were not substitutes for the originals.[167]

Google is in the process of scanning the text of millions of books in order to create the metadata that drives the Google Book search engine. The object of all this indiscriminant copying is the production of metadata – thus to understand the Google Book controversy, it is first necessary to appreciate the value of metadata in the information age. Information is only useful to the extent that it is relevant, discernable and available. There are thousands of volumes of information in even the smallest libraries; however, these dusty tombs are mere ornaments unless a user has some means to locate a particular book, or better yet a particular page, that may be of interest. This is the point where "metadata" becomes valuable. Metadata refers simply to information about information, or data about data.[168] The traditional (and now obsolete) library "card catalog" is an archetypal metadata repository – the card catalog contains information on the author, title and subject matter, but it does not contain the volumes themselves.[169] As the volume of available information increases, so to does the value of metadata used to organize, search, rank and retrieve that information

The copyright issues relating to Google Book must be analyzed in two distinct parts: first, the intermediate copying which produces metadata; and second, the copying and displaying of fragments of books to display along with search results. It is the connection between these two parts that raises a potential ambiguity about whether Google's copying in aggregate should be viewed as non-expressive. The construction of the Google Book

[164] Field v. Google Inc., 412 F. Supp. 2d 1106 (D. Nev. 2006); Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007).
[165] Field, *Id.*
[166] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007)
[167] *Id.*
[168] Metadata is defined as "structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities." American Library Association, Task Force on Metadata *Summary Report*, June 1999 (http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html); Ganesan Shankaranarayanan & Adir Evan, *The Metadata Enigma*, 49 COMMS. ACM 88 (2006).
[169] IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records, *Final Report*. — München: K.G. Saur, 1998. — (UBCIM publications ; new series, vol. 19). — ISBN 3-598-11382-X; Tillett, Barbara. FRBR: A Conceptual Model for the Bibliographic Universe. Library of Congress Cataloging Distribution Service, 2004.

database involves the actual copying of millions of expressive works for an intermediate purpose that is itself entirely non-expressive. In this regard it is exactly analogous to plagiarism detection software.[170] However, while the process of data-generation itself is not an expressive use, the search engine linked to that data does provide expressive snippets of copyrighted books to end-users in response to their search requests.

Does this mean that the intermediate copying performed by Google should be deemed to be expressive in nature? Probably not. First, most of the information produced by the Google Book search engine is metadata about the books which is not amenable to copyright protection because of the idea-expression distinction. Second, even to the limited extent that the search engine displays expressive snippets of books to end-users, those snippets – even if deemed to be expressive – are too fragmented and insubstantial to amount to infringing copies of the books themselves. The better view is that a later expressive use that is non-infringing does not detract from the non-expressiveness of an intermediate use. However, if the snippets displayed by Google were more extensive, the answer might be different. As in the *Field* and *Perfect 10* cases, this context sensitivity indicates difficulties with a categorical approach to the issue of non-expressive use.

In sum, Internet search engines are strong candidates for non-expressive use, but the extent to which that label actually fits may depend on a detailed assessment of specific facts. For example, the claim of non-expressive use in relation to an image search engine that reproduced full scale images as opposed to thumbnails would be doubtful. The categorization of intermediate non-expressive uses that are intertwined with expressive uses is also a source of some ambiguity. Given these and other potential ambiguities, a categorical approach to the issue of non-expressive use may prove to be unwieldy. As the next section sets forth in more detail, courts can avoid the unwieldiness of a categorical approach to non-expressive use by incorporating the same substantive inquiries within the framework of a fair use analysis.

C. Fair Use and Non-Expressive Use

Recognition of the principle of non-expressive use does not require a radical reinterpretation of copyright law. It merely requires an analysis of the existing elements of the fair use doctrine in light of the principle that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement.

The Copyright Act requires courts to consider four factors in making a fair use determination. These factors are (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the

[170] Google's process for generating the metadata behind its book search engine is also clearly analogous to the intermediate copying approved by numerous federal courts in reverse engineering cases. See, Sag, *supra* note 000 at 425-428.

potential market for or value of the copyrighted work.[171] In reality, the Section 107 factors are neither complete,[172] nor are they individually or cumulatively determinative.[173] Even the notion that there are four factors is misleading: beneath the statutory factors lies an amalgamation of interconnected meta-factors, sub-factors and presumptions. The implications of non-expressive use in relation to fair use are explored below.

1. The "Purpose and Character" of Non-Expressive Uses

The defense of non-expressive use is perhaps most clearly relevant under the first fair use factor, "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes."[174] Indeed, recognizing that all of the copyright owner's exclusive rights are implicitly defined and limited in reference to expressive communication to the public, makes sense of both expressive and non-expressive fair uses.

According to the Supreme Court's most recent fair use decision, *Campbell v. Acuff-Rose*, the first factor turns primarily on:

> whether the new use merely supersedes 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'. … Although such transformative use is not absolutely necessary for a finding of fair use, … the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works.[175]

Traditionally, the concept of transformative use has been applied to new expressive uses that "provide social benefit, by shedding light on an earlier work, and, in the process, creating a new one."[176] Transformative use is most obvious when the work itself is literally transformed; however, in many cases courts have held that the mere recontextualization of a copyrighted work from one expressive context to another is

[171] 17 U.S.C. § 107.
[172] Bond v. Blum, 317 F.3d 385, 394 (4th Cir. 2003) ("These factors are not meant to be exclusive, but rather illustrative, representing only general guidance about the sorts of copying that courts and Congress most commonly have found to be fair uses.") (citations omitted); Universal City Studios, Inc. v. Sony Corp. of America, 480 F. Supp. 429, 448 (C.D. Cal. 1979) ("The factors are illustrative, not definitive.")
[173] Sag, God in the Machine, *supra* note 000. See also, Madison *supra* note 000 at 1564 (describing the statute as facially empty); David Nimmer, *"Fairest of Them All" And Other Fairytales of Fair Use*, 66 LAW & CONTEMP. PROBS. 263 (2003) (arguing that "the four factors fail to drive the analysis, but rather serve as convenient pegs on which to hang antecedent conclusions.")
[174] 17 U.S.C. § 107(1).
[175] *See* Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994) (citations omitted). See also, Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1111 (1990).
[176] *Id.* at 579.

sufficient to sustain a finding of fair use – the work itself need not be altered.[177] Understanding the rationale for transformative use is the key to grasping the link between transformative use and non-expressive use. The privileged status of transformative uses under the fair use doctrine allows for the creation of new works from old, but this is not a sufficient explanation because the same effect could be achieved by other doctrinal levers, such as a narrower understanding of the author's exclusive right to make derivative works.[178] Beyond a simple enthusiasm for new works, the reasons that courts accord special status to transformative uses because they do not substitute for the author's original expression – i.e. they do not "merely supersedes the objects of the original creation."[179]

Cognizant of the Supreme Court's focus on transformative uses, some courts have simply equated non-expressive uses with transformative uses. In *Perfect 10* the court held that Google's use of thumbnails in its Internet search engine "may be more transformative than a parody because a search engine provides an entirely new use for the original work, while a parody typically has the same entertainment purpose as the original work."[180] While this might seem to be stretching the concept of transformativeness, there can be no doubt that uses which do not relate to the expressive appeal of a work may find favor under the first fair use factor – whether they qualify as transformative in the expressive sense or not. In *Perfect 10*, the defendant's claim to fair use was strengthened by the fact that it put the images "in a different context so that they are transformed into a new creation;"[181] and also by the fact that it in this new context, the images served a purpose that was non-expressive and thus entirely different from the original photos in their original context. By construction, the non-expressive use of copyrighted works does not substitute for the author's original expression. As such, non-expressive uses should be regarded as equivalent to highly transformative uses – their "purpose and character" is such that they do not merely supersede the objects of the original creation. [182] Non-expressive uses of should be presumed to be fair uses because, by their very nature, they do not give rise to expressive substitution.

2. Non-Expressive Use and Commercial Fair Use

As part of their consideration of the first factor – "the purpose and character of the use" – courts are instructed to consider "whether such use is of a commercial nature or is for nonprofit educational purposes."[183] Although the application of the fair use doctrine to

[177] See, e.g., Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 796-98, 800-06 (concluding that photos parodying Barbie by depicting "nude Barbie dolls juxtaposed with vintage kitchen appliances" was a fair use); Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 609-610 (2d Cir. 2006) (use of promotional posters in a rock-biography was "a purpose separate and distinct from the original artistic and promotional purpose for which the images were created").
[178] 17 U.S.C. 106(2).
[179] Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994)
[180] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 721 (9th Cir. 2007). See also *Kelly v. Arriba Soft Corp*., 336 F.3d 811, 818-19 (9th Cir. 2003) (holding that "even making an exact copy of a work may be transformative so long as the copy serves a different function than the original work.").
[181] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 721 (9th Cir. 2007).
[182] Campbell v. Acuff-Rose Music, 510 U.S. 569, 583 (1994).
[183] 17 U.S.C. § 107(1).

commercial entities has been uncertain for some time, due deference to the Supreme Court's most recent pronouncement on the issue and the economic logic of copyright both suggest that commerciality has no *per se* relevance. The status of commercial fair use has proved to be confusing, in part because it is so closely linked with the question of market substitution under the fourth factor.[184] Conceiving of copyright as a set of exclusive rights in relation to the communication of original expression to the public sheds considerable light on the status of commercial uses under the fair use doctrine. The fact that most copy-reliant technologies are developed and maintained by commercial entities does not weaken their claim to fair use. As explained in more detail below, if a use is non-expressive, its status as commercial or non-commercial is irrelevant because non-expressive uses do not substitute for the author's original expression.

In both *Sony* and *Harper & Row,* the Supreme Court indicated that commercial uses are disfavored under the fair use doctrine. Writing for the majority in *Sony*, Justice Stevens suggested that: "If the Betamax were used to make copies for a commercial or profit-making purpose, such use would presumptively be unfair. The contrary presumption is appropriate here, however, because the District Court's findings plainly establish that time-shifting for private home use must be characterized as a noncommercial, nonprofit activity."[185] Similarly, the majority in *Harper & Row* declared that: "[t]he fact that a publication was commercial as opposed to nonprofit is a separate factor that tends to weigh against a finding of fair use."[186]

However, as the Court later discovered in *Campbell*, a fixed presumption against commercial fair use is difficult to reconcile with the economic logic of copyright. As the Court has reaffirmed most recently in *Eldred*, copyright promotes the creation and publication of free expression "by establishing a marketable right to the use of one's expression."[187] As Neil Netanel observes, the great virtue of copyright is that it "supports a sector of creative and communicative activity that is relatively free from reliance on state subsidy, elite patronage, and cultural hierarchy."[188] The virtues of creative production freed from the shackles of patronage and direct government control apply equally to all forms of private production, regardless whether they rely on the fair use doctrine or not. Thus, the economic and political logic of copyright is inconsistent with placing special burdens on the private sector for no other reason than its pursuit of profit. Noncommercial uses may have other characteristics, such as a greater degree of spillovers, which justify fair use, [189] but there are no inherent differences between the uses of commercial and non-commercial actors. In a modern free-market economy, most

[184] Indeed, the Ninth Circuit's approach to commerciality in *Napster* defines the concept exclusively in terms of market substitution. A&M Records v. Napster, Inc., 239 F.3d 1004, 1015 (9th Cir. 2001) (holding that "commercial use is demonstrated by a showing that repeated and exploitative unauthorized copies of copyrighted works were made to save the expense of purchasing authorized copies.")
[185] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 449 (1984).
[186] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 562 (1985).
[187] Eldred v. Ashcroft, 537 U.S. 186, 219 (2003) (citing Harper & Row 471 U.S., at 558).
[188] Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288 (1996).
[189] See generally, Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257, 261 (2007); Lydia Pallas Loren, Redefining the Market Failure Approach to Fair Use in an Era of Copyright Permission Systems, 5 J. INTELL. PROP. L. 1, 51-53 (1997).

copyright works of interest to the public at large are created by private commercial actors. Newspapers, television broadcasts and Internet search engines are predominantly commercial, and even though schools and universities are often operated by "not for profit" corporations, they are still commercial in the sense that they operate on a fee for service basis. As the Supreme Court recognized in *Campbell,* "[if] commerciality carried presumptive force against a finding of fairness, the presumption would swallow nearly all of the illustrative uses listed in the preamble paragraph of § 107, including news reporting, comment, criticism, teaching, scholarship, and research, since these activities are generally conducted for profit in this country."[190]

The Court in *Campbell* rejected the notion that commerciality by itself had any "hard presumptive significance."[191] Instead, the Court adopted a sliding scale to commercial use, arguing that because "the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works… the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use."[192] This sliding scale approach to commercial uses makes sense in light of the principle of expressive substitution articulated in this Article. The hallmark of transformative works protected by the fair use doctrine is that they do not substitute for the author's original expression. Rather they "add[] something new, with a further purpose or different character, altering the first with new expression, meaning, or message."[193] Arguably, the commerciality of non-expressive uses is irrelevant by definition – non-expressive uses by themselves are incapable of substituting for the author's original expression.

There is ready support for this position in the case law to date. In *Kelly v. Arriba Soft Corp.,* an image search case preceding *Perfect 10*, the Ninth Circuit ruled that the replication of copyrighted images in thumbnails would not substitute for the full-sized images.[194] The court in *Perfect 10* likewise concluded that Google's thumbnail-representations were unlikely to interfere with the market for Perfect 10's original expression.[195] The court expressly rejected the application of any commerciality inference or presumption noting that "this presumption does not arise when a work is transformative because market substitution is at least less certain, and market harm may not be so readily inferred.[196]

[190] Campbell v. Acuff-Rose Music, 510 U.S. 569, 584 (1994) (citing 3 BOSWELL'S LIFE OF JOHNSON 19 (G. Hill ed. 1934, other citations omitted).
[191] *Id.* at 585.
[192] Id. at 579. As Barton Beebe notes, while commentators have assumed that the commerciality presumption was finally discarded in *Campbell*, it remains a tenacious meme in the court of public opinion, and probably some district courts as well. See, Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. Pa. L. Rev. 549, 598 (2008).
[193] Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994); Leval supra note 000 at 1111.
[194] Kelly v. Arriba Soft Corp., 336 F.3d 811, 815 (9th Cir. 2003).
[195] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 724 (9th Cir. 2007).
[196] *Id.* (citing Campbell v. Acuff-Rose Music, 510 U.S. at 591).

3. Non-Expressive Use and "Amount and Substantiality"

Non-expressive use is also significant in terms of the third fair use factor, "the amount and substantiality of the portion used in relation to the copyrighted work as a whole."[197] The third factor eschews mechanical quantification and recognizes that the amount of tolerable copying varies according to both the purpose of the defendant's use and the effect of that use on the copyright owner. The issue at the heart of the third factor is not simply what percentage of the copyright owner's original work has been taken, but what proportion of the work's expressive value has been appropriated. The argument made here is that because a non-expressive use does not substitute for the expressive value of the author's original expression.

Even in the realm of expressive uses, there no linear relationship between the percentage of a work copied and its propensity to fair use. All other things being equal, the more a defendant copies, the more likely she is to interfere with the copyright owner's right to market her works to the public. Thus Napster users who trade complete copies of copyrighted music over the Internet are treated very differently from collage artists who copy only parts of works and add their own significant creative input.[198] But all other things are rarely equal, and courts have repeatedly found that even total copying of expressive works can be fair use in the right circumstances. Courts have held that total copying is permissible in personal use cases, such as those testing the legality of the video cassette recorder and the mp3 player.[199] In cases relating to photography and other visual works, courts have occasionally allowed defendants to reproduce entire images where it was unlikely that any market harm would result and the complete reproduction was considered necessary for the defendant's transformative purpose.[200]

Far from being linear or arithmetic in nature, the reading the tea-leaves of the third factor is contingent on both the purpose, character and effect of the defendant's use. As the Supreme Court recognized in *Campbell*, "the extent of permissible copying varies with the purpose and character of the use".[201] In that case, the Court held that the degree to which rap-musicians 2 Live Crew had copied from Roy Orbison's original song, Pretty

---

[197] 17 U.S.C. § 107 (3). This inquiry can be traced back to Justice Story's original formulation of the fair use doctrine in Folsom v. Marsh 9 F. Cas. 342 (D. Mass. 1841) (No. 4901). In that case, Justice Story was concerned to protect the "chief value of the original work" against the extraction of its "essential parts" through the mere "facile use of scissors" or its intellectual equivalent. *Id.*

[198] A&M Records v. Napster, Inc., 239 F.3d 1004, 1015 (9th Cir. 2001); Blanch v. Koons, 2006 U.S. App. LEXIS 26786 (2d Cir. N.Y., Oct. 26, 2006).

[199] Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 449-450 (1984); Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., 180 F.3d 1072 (9th Cir. 1999); Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc., 964 F.2d 965, 971 (9th Cir. 1992).

[200] See, Kelly v. Arriba Soft Corp., 336 F.3d 811, 821 (9th Cir. 2003) (concluding that images used for a search engine database are necessarily copied in their entirety for the purpose of recognition); Nunez v. Caribbean Int'l News Corp., 235 F.3d 18, 24 (1st Cir. 2000) (concluding that to copy any less than the entire image would have made the picture useless to the story); Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 613 (2d Cir. 2006) (holding that total copying "does not necessarily weigh against fair use because copying the entirety of a work is sometimes necessary to make a fair use of the image"); Mattel Inc. v. Walking Mt. Prods., 353 F.3d 792, 803 n.8 (9th Cir. 2003) (holding that "entire verbatim reproductions are justifiable where the purpose of the work differs from the original").

[201] Campbell v. Acuff-Rose Music, 510 U.S. 569, 587 (1994).

Woman, must be assessed in light of their parodic purpose. Because the art of parody "lies in the tension between a known original and its parodic twin", parody requires copying enough of the original so that the object of derision is made clear to the audience.[202] Just as the extent of permissible copying varies according to purpose, it also varies according to effect. In *Harper & Row*, the defendant copied only a few hundred words from an entire manuscript of the biography of former President Gerald Ford, yet the Supreme Court held that this constituted a substantial taking under the third factor because The Nation had selected its quotes "precisely because they qualitatively embodied Ford's distinctive expression," taken "the most interesting and moving parts of the entire manuscript," and structured its article around these quoted excerpts.[203] The Court's finding in relation to the third factor rests on the fact that The Nation had taken essentially the heart of the book's expressive value.[204]

The third factor does not rely on mechanical quantification of the amount of the original work used; it asks courts to asses how much of the value of the original work is present in the later use.[205] Accordingly, the extent to which a use is non-expressive plays a vital role in the assessment of the third fair use factor. A use that is non-expressive does not substitute for the expressive value of the author's original expression; thus, a non-expressive use should be viewed as qualitatively insignificant under the third statutory factor, even if it requires total literal copying.

Again, existing case law is consistent with this proposition. In *Perfect 10,* the court held that although the thumbnails were copies of the original images, their reduced size and image quality was consistent with their use as pointing devices which did not substitute for the expressive value of the author's original expression.[206] Consistent with its earlier decision in *Kelly*, the court found that the representation of an entire photographic image was reasonable in light of the purpose of an image search engine.[207] As the court explained, while a user can identify relevant text by seeing merely a fraction of it, recognizing images necessitates seeing a representation of the complete image.[208] In *Perfect 10*, as in *Kelly*, the court found that the third fair use factor did not weigh in favor of either party.[209]

Likewise, in *Field,* the court found that making entire web-pages available in the search engine cache served a purpose which could not be effectively accomplished by using only portions of the web-pages. The court found that Google's non-expressive uses of the cached pages – such as verifying the authenticity of live pages and assessing the

[202] *Id.*
[203] Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 565 (1985).
[204] *Id.* at 566 ("In view of the expressive value of the excerpts and their key role in the infringing work, we cannot agree with the Second Circuit that the magazine took a meager, indeed an infinitesimal amount of Ford's original language.")
[205] Sag, God in the Machine, *supra* note 000 at 391.
[206] Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 721-722 (9th Cir. 2007) ("Here, Google uses Perfect 10's images in a new context to serve a different purpose.")
[207] *Id.*
[208] *Id.* (citing Kelly, 336 F.3d at 821.)
[209] *Id.;* Kelly v. Arriba Soft Corp., 336 F.3d 811, 821 (9th Cir. 2003)

relevance of search queries – required caching complete reproductions of the plaintiff's html pages. Accordingly, the district court concluded that because "Google uses no more of the works than is necessary in allowing access to them through "Cached" links, the third fair use factor is neutral, despite the fact that Google allowed access to the entirety of Field's works."[210]

Furthermore, the numerous copyright cases dealing with the practice of reverse engineering computer software also support the proposition that a non-expressive use should be viewed as qualitatively insignificant under the third statutory factor. In *Sony v. Connectix*, for example, the court acknowledged that Connectix had copied an entire section of Sony software multiple times; however it concluded that "in a case of intermediate infringement when the final product does not itself contain infringing material, this factor is of very little weight."[211]

The third factor requires a holistic assessment of the extent to which a work's expressive value has been appropriated measured against the need and justification of the defendant in appropriating it. Accordingly, an untransformative expressive use of a copyrighted work is frowned upon, but transformative expressive uses are granted considerable latitude. Furthermore, non-expressive uses, even those that require total copying in some mechanical sense, should be deemed to be qualitatively insignificant because they do not substitute for the expressive value of the author's original expression.

4. The Market Effect of Non-Expressive Uses

The fourth fair use factor is "the effect of the use upon the potential market for or value of the copyrighted work."[212] The *Harper & Row* Court described the fourth fair use factor as "undoubtedly the single most important element of fair use."[213] Barton Beebe, in contrast, concludes that the fourth factor is "no factor at all."[214] As detailed below, although the fourth factor risks collapsing into circularity because everything is a potential market effect, courts have in fact avoided this outcome by applying certain limiting principles which emphasize that the copyright market is limited to expressive substitution. The logical implication of the exclusion of economic consequences that do not arise from expressive substitution is that non-expressive uses have no *cognizable* market effect under the fourth factor.

To ascertain the market effect of an unauthorized use necessitates first defining the relevant market. If the market is defined purely in terms of that which *might* be licensed if the law says that it *must* be licensed, then the fair use ruling collapses into

[210] Field v. Google Inc., 412 F. Supp. 2d 1106, 1121 (D. Nev. 2006).
[211] Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 606 (9th Cir. 2000).
[212] 17 U.S.C. § 107(1).
[213] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 566 (1985)
[214] Beebe, supra note 000 at 620–621 ("Ultimately, the paradox of the fourth factor is that it is everything in the fair use test and thus nothing. To assert, as a descriptive matter, that it is the most important factor - or, as a normative matter, that it is too important is meaningless, primarily because it is no factor, no independent variable, at all."

circularity.[215] The concept of market effect becomes even more elusive if a trial judge adopts the *Harper & Row* Court's *slippery slope presumption.* In *Harper & Row* the Court announced that "to negate fair use one need only show that if the challenged use should become widespread, it would adversely affect the potential market for the copyrighted work."[216] The aggregation of any harm which is likely to result from widespread use is reasonable, however the Court offers no particular reason why *all* uses should be presumed to become widespread.[217]

Combining the slippery slope of aggregation with a broad concept of derivative works, copyright owners frequently claim that almost any new use of their work – either in whole or in part – is part of an unexplored derivative market.[218] Taken at face value it becomes impossible for a defendant to prove that her particular use, if widely replicated, would not displace some potential future market in some derivative of the copyright owner's work. As the Second Circuit noted in *Texaco* "were a court automatically to conclude in every case that potential licensing revenues were impermissibly impaired simply because the secondary user did not pay a fee for the right to engage in the use, the fourth fair use factor would *always* favor the copyright holder."[219]

Courts avoid this potential circularity by adopting a number of limiting principles in relation to the fourth factor. First, the derivative market is limited by likelihood: "The market for potential derivative uses includes only those that creators of original works would in general develop or license others to develop."[220] Second, fair use cases often turn on the simple question of whether the particular market claimed by the plaintiff is one that is cognizable under copyright. This is not merely a question for the fourth factor; it permeates consideration of all of the factors. The market harms which courts refuse to recognize illustrate yet again the copyright owner's exclusive rights are limited to the communication of their original expression to the public. This principle is reflected in the seemingly unrelated cases involving parody and the reverse engineering of computer software. In both scenarios, courts exclude consideration of market effects that do not arise from expressive substitution.

---

[215] James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 947-51 (2007). But note that although Gibson describes a one-way ratchet effect, the potential circularity of the fourth factor can be set to spin in either direction: if the use is fair, there is no need to license and thus no harm to the market, thus the use is fair; but equally, if the use in unfair, there is axiomatically at least one potential licensee (the defendant) and thus the copyright owner's market has been adversely effected.
[216] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 562 (1985).
[217] This is arguably a distortion of the Senate Report which comments that "Isolated instances of minor infringements, *when multiplied many times*, become in the aggregate a major inroad on copyright that must be prevented." Senate Report, at 65 (emphasis added). Note that in *Campbell* the slippery slope presumption is weakened to a matter for consideration, but still without any analysis of which uses are likely to become widespread and which are not. Campbell v. Acuff-Rose Music, 510 U.S. 569, 590 (1994).
[218] For example, although it had shown no interest in licensing a derivative of "Pretty Woman" in the rap genre before its lawsuit against 2 Live Crew, Acuff-Rose (Roy Orbison's publisher) argued that 2 Live Crew's parody diminished its potential to do so.
[219] American Geophysical Union v. Texaco Inc., 60 F.3d 913, 930, n.17 (2d Cir. N.Y. 1994) (citations omitted). See also, Leval, *supra* note 000, at 1124 ("[b]y definition every fair use involves some loss of royalty revenue because the secondary user has not paid royalties").
[220] Campbell v. Acuff-Rose Music, 510 U.S. 569 (1994).

In *Campbell*, the Supreme Court quite plainly differentiated the copyright owner's general economic interests from the limited protection afforded by copyright:

> when a lethal parody, like a scathing theater review, kills demand for the original, it does not produce a harm cognizable under the Copyright Act. Because parody may quite legitimately aim at garroting the original, destroying it commercially as well as artistically, the role of the courts is to distinguish between biting criticism that merely suppresses demand and copyright infringement, which usurps it.[221]

Just as *Campbell* recognizes that criticism is outside of the copyright owner's protectable sphere of interest, the reverse engineering cases recognize that the copyright owner has no protectable interest in preventing the copying of unprotectable expression and ideas buried within its object code. Federal courts have consistently held that making unauthorized copies of a computer program as a necessary step in reverse engineering is fair use.[222] For example, in *Sony v. Connectix*, the Ninth Circuit held that although the defendant's Virtual Game Station console directly competed with Sony in the market for Sony-compatible-gaming-platforms, the Virtual Game Station was a "legitimate competitor" in that market.[223] The court concluded that Sony's desire to control the market for gaming platforms was understandable but that "copyright law ... does not confer such a monopoly."[224]

Both parody and reverse engineering cases illustrate the exclusion of market effects that do not arise from expressive substitution. This rationale is most explicit in the reverse engineering cases. From the beginning of its decision in *Sony v. Connectix*, the court emphasized the importance of the idea-expression distinction: "We are called upon once again to apply the principles of copyright law to computers and their software, to determine what must be protected as expression and what must be made accessible to the public as function."[225] Consistent with its decision in *Sega*,[226] the Ninth Circuit held in *Connectix* that intermediate copying of software could be protected as fair use if the copying was necessary to gain access to the functional elements of the software.[227] The court based its ruling firmly in the importance of maintaining the idea-expression distinction: "We drew this distinction because the Copyright Act protects expression

---

[221] Campbell v. Acuff-Rose Music, 510 U.S. 569, 591-592 (1994) (quoting in part, Benjamin Kaplan, An Unhurried View of Copyright 69 (1967))

[222] *E.g.,* Sony, 203 F.3d at 602, cert. denied, 531 U.S. 871 (2000); Atari Games Corp. v. Nintendo of Am., Inc., 975 F.2d 832 (Fed. Cir. 1992); Sega Enter. Ltd. v. Accolade, Inc., 977 F.2d 1510, 1520 (9th Cir. 1992); see also David A. Rice, *Copyright and Contract: Preemption After Bowers v. Baystate*, 9 ROGER WILLIAMS U. L. REV. 595, 601 n.19 (2004) (further references). Circumventing encryption for the purpose of reverse engineering is also allowed under the safe-harbor provisions of the DMCA, see [cite].

[223] Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 607 (9th Cir. 2000); see also, Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510, 1522-23 (9th Cir. 1993).

[224] Sony, 203 F.3d at 607; see also, Sega, 977 F.2d at 1523-24.

[225] *Id.* at 598.

[226] Sega Enterprises Ltd. v. Accolade, Inc., 977 F.2d 1510 (9th Cir. 1993).

[227] Sony Computer Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 607 (9th Cir. 2000)

only, not ideas or the functional aspects of a software program ... Thus, the fair use doctrine preserves public access to the ideas and functional elements embedded in copyrighted computer software programs."[228] As in the parody cases, although for different reasons, the reverse engineering cases exclude consideration of market effects that do not arise from expressive substitution.

In the case of expressive uses, such as parody, and non-expressive uses, such as reverse engineering, courts have consistently held that the protection that copyright affords is limited to certain cognizable markets. Transformative expressive uses are usually found not to affect the market in any relevant sense because the second author's expression does not substitute for that of the original author. The absence of any cognizable market effect is even more apparent in non-expressive use cases because there is literally no potential substitution effect. Accordingly, acts of copying which do not communicate the author's original expression to the public should also be presumed not to affect the potential market for or value of the copyrighted work.

***

As established earlier in this Part, the exclusive rights of the copyright owner are limited to the communication of original expression to the public. Furthermore, acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. As demonstrated above, these principles are not merely compatible with the fair use doctrine; moreover, they are in many cases necessary to make sense of existing case law. It may be unrealistic to attempt to reduce the entirety of fair use jurisprudence into any one coherent principle. Nonetheless, the general proposition that acts of copying that are unlikely to substitute for the copyright owner's original expression are favored by the doctrine explains the majority of cases. Non-expressive uses should be presumed to be fair uses because, by their very nature, they do not substitute for the author's original expression. Accordingly, non-expressive use should be favored under the first, third and fourth factors – such uses are non-substitutive in 'purpose and character', appropriate a qualitatively insignificant proportion of the value of the copyright owner's original expression, and produce no cognizable market effect under the fourth factor.[229]

This Part has addressed the first question with respect to copy-reliant technologies: whether non-expressive uses that nonetheless require copying the entirety of a copyright work be found to infringe the exclusive rights of the copyright owner. It has demonstrated that because the rights of the copyright owner are generally limited to a monopoly over the communication of the expressive aspects of their works, extending the rights of copyright owners to encompass non-expressive uses by copy-reliant technologies would constitute a significant departure from existing copyright principles.

[228] *Id.* at 603.
[229] As is so often the case, the second statutory factor does not appear to have much bite in the context of non-expressive uses, and so does little to "sort the fair use sheep form the infringing goats." Campbell at 586.

Finally, this part has established the plausibility of addressing the issue of non-expressive use through the fair use doctrine.

The second issue to be addressed with respect to copy-reliant technologies relates to transaction costs. Specifically do the transaction costs associated with copy-reliant technologies justify switching copyright's default rule that no copying may take place without permission to one in which copyright owners must affirmatively opt-out of specific uses of their works? Part III which follows considers the doctrinal implications of high transaction costs in relation to copy-reliant technologies and the use of opt-out mechanisms to resolve those transaction costs.

III. THE DOCTRINAL SIGNIFICANCE OF TRANSACTION COSTS

Copyright disputes involving copy-reliant technology should be completely resolved once a court determines that the use in question is non-expressive. However, there are nonetheless two reasons for delving further into the issues of transaction costs and opt-outs that preoccupy this final Part. First, although the correct application of the non-expressive use principle expounded in this Article will be clear in many cases, it may not be clear-cut in every case. In those cases where ambiguity persists, courts will have to consider the totality of the defendant's fair use claim. In that context, the defendant's opt-out mechanism takes on considerable significance. Second, opt-outs are common feature of copy-reliant technology and their ubiquity deserves some explanation. Part III-A begins with an explanation of why copy-reliant technologies face significant transaction costs problems and the role of opt-out mechanisms in reducing those transaction costs. Part III-B discusses the relationship between transaction costs and the form and content of property rights generally. Part III-C then specifically addresses how and why the use of opt-outs by copy-reliant technologies is relevant to a fair use analysis.

A. Transaction Costs and Copy-Reliant Technologies

The transaction costs faced by copy-reliant technologies are different in kind and in magnitude to those pertaining to analog works. The sheer number of transactions that must typically be accommodated by copy-reliant technologies makes the transaction costs problems they face somewhat unique. The irony of copy-reliant technology is that, while technology has helped reduce the per-unit transaction costs in relation some discrete objects, such as motion pictures and sound recordings, the proliferation of copyrighted content means that total transaction costs for any technology that must cover the whole of the Internet have increased dramatically. While private ordering through collective rights management may be a common solution in various fields of intellectual property, it is unlikely to offer any solution in the context of copy-reliant technologies due to the scale, decentralization and heterogeneity of Internet. Nonetheless, Internet entrepreneurs have found other ways to address transaction costs, primarily through the combination of well understood default rules and technologically enabled opt-out mechanisms. These issues are now addressed in detail.

1. Scale, Diversity and Decentralization

The sheer scale of the Internet is truly daunting. No technology since the printing press has given rise to a proliferation of copyrighted works equivalent to the explosion of Internet content witnessed since the mid-1990's. A simple comparison helps illustrate this point. The U.S. Library of Congress is the world's largest library, with more than 134 million books, photographs, maps, works of music, manuscripts and other printed materials.[230] The volume of material available on the Internet has dwarfed this number in a very short period. There are now an estimated 1.2 billion Internet users world-wide.[231] It is difficult to estimate the number of web-pages available on the world-wide-web at any given time; however the Internet Archive – which is only a partial collection – contains 85 billion searchable pages archived from 1996 to the present.[232] This e number alone exceeds the entire collection of the Library of Congress by a ratio of more than 600 to 1.

The volume of material on the Internet presents a significant transaction costs problem for Internet search technology in particular because the value of any search engine grows exponentially with its coverage. The Google Book project and plagiarism detection software confront a similar network effect. The difference between the requirements of copy-reliant technologies and other more traditional consumers of copyrighted works is illustrated by the contrast between an Internet search engine and a book publisher. To provide a useful product, a book publisher must sift through a large number of submitted manuscripts, select one, and negotiate a license with the author. The publisher is fortunate that once it finds one good manuscript, there is no need to read the remainder. Furthermore, if the publisher's preferred author is intransigent in negotiations, she can proceed to her second best alternative at very little sacrifice. In contrast, an Internet search engine cannot just select one or two websites under each search term and rest on its laurels. First, search engines must be able to cope with unexpected queries. Second, search engines are subject to a significant network effect such that they are only really useful at a threshold of near-complete coverage.

The Internet has not only expanded information production, it has radically decentralized it as well. The Gartner consulting firm estimates that around 100 million writers actively maintained a personal website or blog in 2007.[233] Even as the mainstream press continues to consolidate into fewer and fewer media empires, the Internet has decentralized news

---

[230] *See* The Library of Congress, *About the Library,* http://www.loc.gov/about/facts.html. Only 32 million of these items are books.

[231] Internet World Stats, *World Internet Usage And Population Statistics*, available at http://www.Internetworldstats.com/stats.htm.

[232] The Internet Archive is a non-profit organization that was founded to build an Internet library, with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format. *See,* The Internet Archive, *About the Internet Archive*, at http://www.archive.org/about/about.php.

[233] Antony Savvas, *Gartner's top 10 forecasts for 2007 and beyond,* COMPUTERWEEKLY.COM, December 15, 2006, available at http://www.computerweekly.com/Articles/2006/12/15/220726/gartners-top-10-forecasts-for-2007-and-beyond.htm.

production and increased both its volume and its diversity.[234] Not only are these actors decentralized, they are also diverse.[235] The "new media" and distributed production exemplified by blogs and social networking sites are characterized by (i) a blurring of the lines between producers and consumers, (ii) a significant degree of interaction between participants who are both producers and consumers and (iii) low monetary costs, at least for the majority of participants.[236] As a result, the Internet has complicated the economics of copyright by expanding the range of viable information production strategies. The proliferation of content producers and their heterogeneity is no doubt beneficial, but it presents copy-reliant technologies with a difficult set of transaction costs problems if they were to clear rights before unleashing their automated processes.

The high transaction costs environment is not limited to Internet search engines. Because plagiarism detection software must search the whole Internet for possible sources of plagiarism, it faces a similar problem of scale. The transaction costs issues in Google Book are similar, but distinct. Google faces substantial costs in building out its database of library books. There are approximately 18 million books in the combined collections Google's partner libraries, each one of these needs to be pulled off a shelf, scanned and re-shelved.[237] The average cost to scan each book is estimated at around $10.[238] In addition to these costs, if Google's intermediate copying is not fair use, it will also be confronted with a substantial rights clearance problem. For each book Google will have to (1) determine whether the book is in the public domain, (2) determine the identity of the copyright owner(s), (3) locate the copyright owner(s) and negotiate to obtain their permission.

Google's clearance costs will vary according to the book in question: they will be lowest for very old works (pre-1923), modest for very new works (2001 onwards) and highest for those in between (1923 to 2000). The clearance costs for very old and very new works are quite low. If a work was published in the United States before January 1, 1923 it is safe to assume that is in the public domain.[239] The clearance costs for very new works are also quite low because publishers now insist on obtaining the relevant rights from authors and are in a position to grant Google permission to include these works in its database.[240]

---

[234] See, YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM 223 (2006). See also, Brett M. Frischmann, *Cultural Environmentalism and The Wealth of Networks*, 74 U. CHI. L. REV. 1083 (2007)

[235] *Id.* (describing information production in a networked information economy and its inclusion of a broader range of participants).

[236] *Id.*

[237] Brian Lavoie, Lynn Silipigni Connaway & Lorcan Dempsey, *Anatomy of Aggregate Collections The Example of Google Print for Libraries*, 11(9) D-LIB MAGAZINE, September 2005, available at http://www.dlib.org/dlib/september05/lavoie/09lavoie.html.

[238] See e.g., Carolyn Said, *Revolutionary Chapter; Google's Ambitious Book-Scanning Plan Seen As Key Shift In Paper-Based Culture*, S.F. CHRON. Dec. 20, 2004, F1. ("press reports have pegged it at about $10 per book.") See also, Eleanor Yang Su, *Google Will Post UC Library Books Online For Public*, SAN DIEGO UNION-TRIBUNE, Aug.10, 2006, Pg. A-4 (reporting that the University of California estimates that it would cost it $30 to $40 per book to scan its collection).

[239] Note that a work created but not published prior to 1923 may not be in the public domain.

[240] Google and Amazon.com have each established cooperative agreements with publishers which allow them to display considerable portions of books in a searchable database.

However, for the vast numbers of books published between 1923 and 2001, the applicable clearance costs are likely to be quite high.[241] First, although it may not be obvious from the date of publication alone, many of these works are in fact in the public domain. The duration of copyright under the 1909 Act, was limited to 28 years, plus a renewal period of 28 years.[242] This renewal period was subsequently extended to 47 years, and the again to 67 years,[243] the result being that if a work was published in the United States before 1950 and not renewed, it belongs to the public domain. Likewise, certain works may also be in the public domain because they were published in the United States without the appropriate copyright notice. However, the notice requirement only applies to works first published before March 1, 1989.[244] Additionally, notice defects will not affect a work's copyright status if the defective copy was published without the authorization of the copyright owner or the notice defect only applied to a small number of copies.[245] Furthermore it was possible for the author to correct a notice defect in some circumstances.[246] It is important to note that any work created by a U.S. government employee or officer also belongs to the public domain, provided that the work is created in that person's official capacity.[247] The status of unpublished works and the works of foreigners adds several additional layers of complexity, including the possibility that some works in the public domain may have been restored to copyright protection by the Uruguay Round Agreements Act of 1994.[248]

Second, even if a work remains subject to copyright, the ownership of those rights is can be highly uncertain. The basic principle of copyright law is that copyright vests initially in the author or authors of the work.[249] However, those rights may be assigned in an infinite chain of transactions, bankruptcies, and by the laws of succession. These laws are by no means uniform. The disposition to copyright ownership through a will or intestacy is determined by the domicile of the author,[250] even if that domicile is in a foreign country.[251] Furthermore, unvested renewal rights under the 1909 Copyright Act and rights of termination under the 1976 Copyright Act pass according to complex statutory

[241] *See,* Elizabeth Townsend Gard, *Vera Brittain, Section 104(a) and Section 104A: A Case Study in Sorting Out Duration of Foreign Works Under the 1976 Copyright Act.* Tulane Public Law Research Paper No. 07-09 Available at SSRN: http://ssrn.com/abstract=1015575.
[242] 17 U.S.C. § 24 (1909 Act).
[243] 17 U.S.C. § 304(a).
[244] 17 U.S.C. §§ 401-402 (prior to the Berne Convention Implementation Act). Note that whether a work was in fact "published" turns on a number of factual questions, see Nimmer on Copyright § 4.04 for a general discussion.
[245] A limited publication without copyright notice does not inject a work into the public domain if the work is communicated "to a definitely selected group and for a limited purpose, without the right of diffusion, reproduction, distribution or sale." See e.g. Regents of the Univ. of Minn. v. Applied Innovations, Inc., 685 F. Supp. 698, 710 (D. Minn. 1987), *aff'd*, 876 F.2d 626 (8th Cir. 1989). Publication can be especially ambiguous for copies of letters deposited with a library.
[246] 17 U.S.C. §§ 405 (prior to the Berne Convention Implementation Act).
[247] 17 U.S.C. §105.
[248] The Uruguay Round Agreements Act of 1994 restores copyright protection to certain foreign works which had at some stage fallen into the public domain due to failure to comply with formalities, but which would otherwise now be eligible for protection in the Unites States. 17 U.S.C. § 104A(a)(1)(A).
[249] 201(a). This is subject to the work for hire doctrine and the possibility of joint-authorship.
[250] 17 U.S.C. § 201(d)(1).
[251] Brecht v. Bentley, 185 F. Supp. 890 (S.D.N.Y. 1960).

provisions, which may, or may not, mirror the author's will or the default positions of her domicile.[252] Beyond the name of the initial author, almost none of the facts relevant to determine the current ownership of copyright in a work are disclosed in the work itself or the records of the United States Copyright Office. As the work ages, complexity of these legal and factual issues multiplies exponentially.

Third, Google faces significant costs related to opportunism and strategic behavior. As discussed in more detail below, even authors who favor inclusion in the Google Book database may have an incentive to hold-out for higher payment if their copyright gives them an effective veto over the project.[253]

How significant are these costs likely to be? There are an estimated 18 million books in the combined collections of the libraries participating in the Google Book project. Approximately 10.5 million of these books are unique – i.e. they are only held by one the participating libraries.[254] It is estimated that slightly less than 20% of these works were published before 1923 and thus present no copyright issues.[255] That leaves about 8.4 million books with some potential copyright constraint. If the average clearance cost (the cost of determining the status of the book, finding the relevant copyright owners and negotiating a license) was $200, then the total cost of rights clearance before any royalties have been paid would be $1.68 billion. It is of course easy to imagine that clearance costs could be in the thousands, not merely the hundreds, in which case the total cost of pro-actively clearing rights on every book could exceed $10 billion. This does not include any royalties paid to authors. As these very preliminary estimates show, the problem of high transaction costs is common to most copy-reliant technologies and is not limited to Internet search engines.

2. Technology Reduces Some Transaction Costs While Increasing Others

Advances in technology have reduced transaction costs in many areas by reducing the cost of communication and increasing the effectiveness of search, however, such advances have done little to offset the significant transaction costs problems faced by copy-reliant technologies. Commentators have been predicting the death of fair use on the Internet since the late 1990's. Specifically, the prediction was that digital rights management technology will allow copyright owners to automatically enforce their rights and to prevent uses that were once considered fair.[256] Those who embraced the death of fair use online argued that DRM would allow copyright owners to define the permissions associated with their works make it possible to charge different prices to different users thus reducing the need for fair use.[257] Those who feared the death of fair use made the

[252] 17 U.S.C. § 304(a).
[253] See *infra* part ___
[254] Lavoie, Connaway & Dempsey, *supra* note 237.
[255] *Id.*
[256] *See* Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems,* 15 HARV. J.L. & TECH. 41 (2001).
[257] *See* Paul Goldstein, *Fair Use in a Changing World*, 50 J. COPYRIGHT SOC'Y U.S.A. 133, 137 (2003) ("For the great bulk of uses previously excused because of transaction costs, the [fair use] doctrine will simply become irrelevant."). See also, Edmund W. Kitch, *Can the Internet Shrink Fair Use?*, 78 NEB. L.

same prediction, arguing that that the control facilitated by DRM will enable an end-run around the public policy values embedded in copyright law.[258]

Those predictions have proven to be extravagant. DRM permission systems have had a muted impact on dvds and digital music and no impact whatsoever on the majority of transactions relevant to copy-reliant technologies. The effect of DRM in the context of dvds and digital music has been muted because permissions systems are fragile and hard to maintain for several reasons, two of which are addressed here. First, once the encryption on any one copy of a work is broken, that copy can be used to propagate an infinite number of un-encrypted copies. Second, users tend to gravitate toward unrestricted formats precisely because they offer fewer restrictions.[259] The network effects of this preference for unrestricted formats means that content providers are only in a position to impose permission systems if they are able to control both the content format and the playback technology. Content providers have been partially successful in developing permissions systems with respect to dvds, but similar attempts with respect to audio CD's have been somewhat disastrous.[260] To the extent that copyright owners have had any success with DRM, these successes look more like "sufficient" control within the

---

REV. 880, 881 (1999)(same); Tom W. Bell, *Fair Use vs. Fared Use: The Impact of Automated Rights Management on Copyright's Fair Use Doctrine*, 76 N.C. L. REV. 557 (1998) (same); Robert P. Merges, *The End of Friction? Property Rights and Contract in the "Newtonian" World of On-Line Commerce,* 12 BERKELEY TECH. L.J. 115 (1997) (same); I. Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217 (same); Jane C. Ginsburg, *Authors and Users in Copyright*, 45 J. COPYRIGHT SOC'Y U.S.A. 1, 15 (1997) ("The primary justification for exempting private copying as fair use has been transaction costs, but these are much attenuated in the digital world").

[258] *See* Burk & Cohen, *supra* note 256. *See also,* Glynn S. Lunney, Jr., *The Death of Copyright: Digital Technology, Private Copying, and the Digital Millennium Copyright Act,* 87 VA. L. REV. 813, 814 (2001).

[259] Both Apple and Amazon have adopted this strategy, see Apple, *Apple Unveils Higher Quality DRM-Free Music on the iTunes Store* at http://www.apple.com/pr/library/2007/04/02itunes.html; Engadget, *Amazon announces DRM-free MP3 music store*, at http://www.engadget.com/2007/05/16/amazon-announces-drm-free-mp3-music-store. Note also that consumer's may distrust because of a potential lack of backwards compatibility. For example, music bought from Microsoft's MSN music which uses "plays4sure" DRM cannot play on the Microsoft Zune digital media player. See, posting to cdfreaks.com, *Microsoft postpones MSN Music DRM server shut-off*, June, 19, 2008, available at http://www.cdfreaks.com/news/Microsoft-postpones-MSN-Music-DRM-server-shut-off.html?news=14762.

[260] See, Megan M. LaBelle, *The "Rootkit Debacle": The Latest Chapter in the Story of the Recording Industry and the War on Music Piracy*, 84 DENV. U.L. REV. 79; J. Alex Halderman & Edward Felton, *Lessons from the Sony CD DRM Episode* (Princeton University, May 16, 2006), available at http://itpolicy.princeton.edu/pub/sonydrm-ext.pdf. See generally R. Polk Wagner, *Information Wants to be Free: Intellectual Property and the Mythologies of Control*, 103 COLUM. L. REV. 995, 1015-16 (2003); Peter S. Menell, *Envisioning Copyright Law's Digital Future,* 46 N.Y.L. SCH. L. REV. 63, 193 (2002-2003). The loose-permission system developed by Apple in conjunction with iTunes demonstrates the problems of maintaining permissions systems. Apple is able to impose certain conditions on its music downloads because it controls both the file format and the playback technology. However, because the iTunes platform is tied to the iPod line of digital music players, Apple has been under sustained pressure to unlock its proprietary technology to allow music purchased on competing digital music stores to be transferred to the iPod. For an overview, see Kevin J. Harran, *Challenges in the Global IT Market: Technology, Creative Content, and Intellectual Property Rights*, 49 ARIZ. L. REV. 29 (2007). In response to this pressure Apple has advocated simply removing DRM from music downloads altogether, and at least one record company has actually agreed to take this step. See, Deana Sobel, *A Bite out of Apple? iTunes, Interoperability, and France's Dadvsi Law*, 22 BERKELEY TECH. L.J. 267 (2007).

traditional contours of copyright law, not the Orwellian vision of a digital lock-down predicted over a decade ago.[261]

Dvds and digital music notwithstanding, advances in the technological protection of content have been overwhelmed by advances in the technologies of reproduction and distribution. The proliferation of easy to copy content on the Internet has actually increased the economic significance of transaction costs. The dominant transaction costs problem on the Internet relates to negotiating basic permissions for billions of pages, not sophisticated bargaining over relatively few high volume items such as popular movies, books and music. Thus while DRM technology may have the potential to reduce transaction costs with respect to any one individual pre-existing work, the magnitude of transactions that copy-reliant technologies must process has increased exponentially. So, ironically, while Internet search engines have reduced transaction costs in relation to many copyrighted markets, they themselves are subject to increasing transaction costs by virtue of their own success.

The pre-millennial consensus that technology would reduce the significance of transaction costs in relation to copyright failed to take into account the difference between the costs attending any one transaction and the total volume of transaction costs faced by copy-reliant technologies. The former have been reduced by the adoption of new technologies, the latter have not.

3 Collective Rights Management and Copy-Reliant Technologies

Collective management is the exercise of copyright and related rights by organizations acting in the interest and on behalf of the owners of rights.[262] In many intellectual property contexts, transaction costs problems are addressed through collective rights management, in the form of collection societies such as the American Society of Composers, Authors, and Publishers (ASCAP), or patent pools and joint ventures, such as the 3G Patent Platform Partnership.[263] The success of collective rights management in some fields demonstrates that high transactions costs may be overcome by market-based solutions where the individual management of rights is impossible or impractical.[264] However, this particular type of private ordering solution is unlikely to be effective in relation to copy-reliant technologies because of the scale of transactions required and the decentralization and diversity of the relevant rights holders.

---

[261] See, Jane Ginsburg, T*he Pros And Cons Of Strengthening Intellectual Property Protection: Technological Protection Measures and Section 1201 Of The U.S. Copyright Act,* (version of February 1, 2007) at 24 ("To date, "digital lock-up" persists in spectral guise, a grim, yet untranspired, anticipation.").
[262] See, World Intellectual Property Organization, *About WIPO, Collective Management of Copyright and Related Right*, http://www.wipo.int/about-ip/en/about_collective_mngt.html#P46_4989.
[263] 3G Patent Platform Partnership is a standard setting organization designed to cap total fees paid to patentees that own rights in the 3G mobile phone standard.
[264] Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293, 1293-94 (1996).

The problem with collective rights management is that it has been most successful in the context of homogeneous transactions among repeat-players with similar preferences. [265] Collective rights management is unlikely to reduce the transaction costs faced by copy-reliant technologies. First, copy-reliant technologies typically rely on close to complete coverage – a search engine or plagiarism detection program that only covers half the Internet is of very limited use. This means that each html page is a complementary good, whereas collective rights organizations like ASCAP typically license a range of potential substitutes. Second, collective rights organizations like ASCAP only work because a significant percentage of relevant copyright owners affirmatively opt-in to that system. Given the billions of works at issue and the hundreds of millions of rights owners that would be required to proactively clear rights for an Internet search engine, similar levels of participation seem unlikely. This is particularly so given decentralized nature of the relevant actors and the diversity of their motivations. In fact, even where the perquisites for effect collective rights management appear to exist, experience suggests that rights holders have been slow to take advantage of the potential savings collective action offers.[266]

4. Private Ordering Through Opt-Outs

It has been established so far that transaction costs present a formidable potential obstacle to copy-reliant technologies, one which is unlikely to be overcome by either DRM technology or collective rights management. How then do copy-reliant technologies continue to function? The primary mechanism for reducing transaction costs in relation to copy-reliant technologies has been the combination of well understood default rules and technologically enabled opt-out mechanisms.

To grasp how these opt-out mechanisms work requires some (but not too much) understanding of the basic structure of the Internet. The Internet is an open system which allows any end-point to communicate with any other end-point through a set of standard protocols.[267] The architecture of the Internet thus embeds a default rule of unrestricted access. This default requires anyone who does not wish their material to be available to affirmatively opt-out. For example, website owners remain free to restrict access by blocking specific IP addresses, or by requiring a user account and/or password. They can also control how search engines interact with their copyrighted material by employing a technological device known as the Robots Exclusion Protocol.[268] The default is, however, an open system.

The Internet norm of open access stands in marked contrast to the usual assumptions made with respect to copyrighted works. This norm has remained stable for some time

[265] *Id.* at 1319 ("Only repeated transactions among right holders will give rise to the private institutions discussed . . . One-shot or sporadic interactions do not justify investments in exchange institutions.")
[266] *See* WILLIAM W. FISHER & WILLIAM MCGEVERAN, THE DIGITAL LEARNING CHALLENGE: OBSTACLES TO EDUCATIONAL USES OF COPYRIGHTED MATERIAL IN THE DIGITAL AGE 80 (2006) (noting that there is a greater reluctance by rights holders regarding licensed digital uses of content as opposed to analog uses such as photocopies).
[267] JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? 23 (2006).
[268] See *infra* notes 271 to 275 and accompanying text.

for three reasons. First, the initial design of the Internet and its basic protocol for the exchange of information embedded an open architecture.[269] The open, minimalist and neutral design of TCP/IP has enabled an unparalleled diversity of social and technological innovations.[270] Open systems and end-to-end architecture were fundamental early Internet technical standards, thus the default of open access continues, in part, simply as a result of path dependence. Second, the norm of openness also continues to flourish because it reflects the preferences of the majority of Internet users. Most people want their websites to be seen and their emails to be received. Third, those with minority preferences generally have no quarrel with the default of open access; they simply opt-out of the default as it suits them.

This third point requires some elaboration. Those with minority preferences can easily opt-out of the default rules that govern the Internet in a number of ways, the most significant of which is probably the Robots Exclusion Protocol.[271] The Robots Exclusion Protocol is particularly significant in the context of copy-reliant technology. Every major Internet search engine relies on the Robots Exclusion Protocol to prevent their automated agents from indexing certain content and to remove previously indexed material from their databases as required. Although it has been widely adopted, the Robots Exclusion Protocol is not controlled by any standards setting organization and thus remains a de-facto standard. The success of the Robots Exclusion Protocol is attributable to two factors: its low cost and high degree of customization. The monetary cost of using the Robots Exclusion Protocol is zero and the information costs are not significantly higher. Adding a robots.txt file a web site is fairly trivial and there a number of widely available free tools for automatically generating a robots.txt file.[272] To disallow all robots from a website simply requires two lines of code:

```
User-Agent: *
Disallow: /
```

[269] See Vint Cerf & Robert Kahn, *A Protocol for Packet Network Intercommunication*, IEEE TRANSACTIONS ON COMMUNICATIONS, Vol. Com-22, No. 5, May 1974, pp. 637-648 (1974). (http://www.cs.princeton.edu/courses/archive/fall06/cos561/papers/cerf74.pdf) (the original specification of the "Transmission Control Protocol"); Christos J. P. MOSCHOVITIS, HISTORY OF THE INTERNET: A CHRONOLOGY, 1843 TO THE PRESENT 80 (1999); Goldsmith & Wu, *supra* note 000 at 23.

[270] See, Goldsmith & Wu, *supra* note 000 at 22-24.

[271] For a general discussion of the Robots Exclusion Protocol, see *David Gourley & Brian Totty*, HTTP: THE DEFINITIVE GUIDE, 225 –241 (2002). See also *The Web Robots Pages, About/robots.txt*, available at http://www.robotstxt.org/robotstxt.html. The original Robots Exclusion Protocol was set out in 1994 by Martijn Koster. See Martijn Koster, *A Standard for Robot Exclusion* (1994), available at http://www.robotstxt.org/orig.html. See also, Martijn Koster, *A Method for Web Robots Control* (1996), available at http://www.robotstxt.org/norobots-rfc.txt (a draft internet specification of the Robots Exclusion Protocol); WC3 Recommendation, HTML 4.01 Specification, Appendix B.4.1, available at http://www.w3.org/TR/html4/appendix/notes.html#h-B.4.1.1.

[272] The Web Robots Pages provides a tutorial on creating robots.txt files with a text editor, see, http://www.robotstxt.org. Google offers a free robots.txt generator as part of their webmaster tools at http://www.google.com/support/webmasters/bin/answer.py?answer=83098&topic=13648. Microsoft's "How to Write a Robots.txt File" Knowledgebase article: http://support.microsoft.com/kb/217103. See also, http://robots.googletoad.com/ and http://www.seochat.com/seo-tools/robots-generator/.

Adding these instructions to the robots.txt file at the root level of a website,[273] will block all compliant search engine robots and other information harvesting software agents.[274]

Apart from its low cost and relative simplicity, the real attraction of the Robots Exclusion Protocol is its extraordinary flexibility. To block a particular directory rather than the entire site requires simply changing the second line to include the name of the directory.[275] The Google search engine, for example, is designed to allow site owners to prevent individual pages, sections of a web-site or an entire web-site from being indexed.[276] In the event that content has already been indexed and the web site owner changes its preferences, the Google search engine will remove this content from the cache once the robots exclusion standard is activated.[277] Google's implementation of the Robots Exclusion Protocol is also highly customizable: among other things, site owners can also remove the snippets and/or images that appear below their page's title in Google search results.[278]

The important thing to realize about opt-out mechanisms such as the Robots Exclusion Protocol is that they do not displace private ordering – they are the means of private ordering. When transaction costs are otherwise high, opt-out mechanisms can play a critical role in preserving a default rule of open access while still allowing individuals to have their preferences respected. In the context of search engine technology, opt-out mechanisms such as the Robots Exclusion Protocol have reduced seemingly insurmountable transaction costs and made them trivial.[279] *Field*, *Perfect 10* and *Google Book* are interesting in part because the plaintiffs in those cases chose to object to the default rule instead of simply opting out.

B. Transaction Costs and Property Rights

Critics of various copy-reliant technologies are quick to invoke the rhetoric of property in service of their claims. Former Authors Guild president, Nick Taylor is illustrative, he argues that Google is "in effect, stealing people's property and providing others with access to it for its own gain."[280] Despite its headline appeal, the rallying cry of property rights tells us very little about the scope and form of those rights when novel policy questions present themselves. When a new office building casts a shadow over a hotel

[273] i.e. http://www.example.com/robots.txt
[274] See generally, THOMAS A. POWELL, WEB DESIGN: THE COMPLETE REFERENCE 247–249 (2000); TARA CALISHAIN & RAEL DORNFEST, GOOGLE HACKS: 100 INDUSTRIAL-STRENGTH TIPS & TOOLS 309 (2003)
[275] i.e. Disallow: /nameofdirectory/
[276] See, Calishain & Dornfest supra note 000 at 315; Google, *How can I prevent content from being indexed or remove content from Google's index?* Available at http://www.google.com/support/webmasters/bin/answer.py?answer=35301&topic=8459.
[277] *Id.*
[278] *Id.*
[279] *See* John S. Sieman, *Using the Implied License To Inject Common Sense into Digital Copyright*, 85 N.C.L. REV. 885 (2007) (The transaction costs in getting permission before viewing every website would be so high that people would be likely to stop visiting websites. An opt-in Internet would be virtually unusable.)
[280] Nick Taylor, *Letter from the President - Q4 2005*, Authors Guild Website, available http://www.authorsguild.org/?p=18221.

swimming pool or obstructs the air current so as to impede the operation of a windmill, it is pointless for the effected parties to simply proclaim their property rights are sacrosanct.[281] The owners of the office building and the windmill each have property rights: the real question is what exactly is the content of those rights? Whose rights prevail when conflicting claims are asserted? Both the doctrinal and welfare economics answers to this question must be resolved with reference to transaction costs.

Transaction costs are central to an economic understanding of property rights because they dictate both the scope and the form of private rights. Private property generally reduces transaction costs by lowering the costs of coordination among disparate individuals. Secure property rights are generally thought to be essential for the increased specialization that sustains economic development because they provide the institutional framework needed for long term and complex relationships.[282] The allocation and definition of property rights determines both which individuals have the authority to decide how specific resource is used and to whom the costs and benefits of that use will flow.[283] As every student of the *Coase Theorem* knows, in a world without transaction costs, the specific allocation of these costs and benefits is unimportant because all the relevant parties will bargain to an efficient outcome regardless of their initial entitlement.[284] However, as every student of the *Coase Theorem* also knows, in the real world reallocation and enforcement are costly and many transaction costs persist. Indeed, because the specification of rights is itself a costly endeavor, it is axiomatic that rights will never be fully specified.[285]

How then should property rights be allocated given that transaction costs abound? One view is that, given transaction costs stand in the way of efficient reallocation, the primary objective of the law should be to reduce transaction costs by defining simple and clear property rights which enable private exchange. [286] On this view, law-makers should not devote considerable resources to optimizing initial allocation; they should just ensure that all the rights worth specifying are allocated.[287] A second view is that, because substantial transaction costs persist even after private rights have been allocated, law-makers must attempt to allocate property rights to their best initial use so as to minimize the harm caused by inevitable failures to reach private agreements.[288]

Although these contending implications of the *Coase Theorem* are frequently offered as a binary choice, in fact, neither prescription should be accepted as dogma, because the applicability of either depends on the exact nature of the transaction costs at issue. In fact, just as the decision of a firm to either make or buy is determined by relative transaction

[281] These examples come from Ronald H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 8 (1960).
[282] DOUGLASS C. NORTH, INSTITUTIONS, INSTITUTIONAL CHANGE AND ECONOMIC PERFORMANCE 33–35 (1990). Property rights are but one element of the institutional matrix that sustains the rule of law.
[283] NICHOLAS MERCURO AND STEVEN G MEDEMA, ECONOMICS AND THE LAW, 249 2d (2006).
[284] Coase, *supra* note 281
[285] North, *supra* note 282.
[286] ROBERT COOTER & TOMAS ULEN, LAW & ECONOMICS (4d) 97 (2004).
[287] *Id.* at 97
[288] *Id.*

costs, so is the structure of property rights.[289] In this context it should be noted that the form that property rights take can play a significant a role in reducing *or exacerbating* transaction costs. There is in consequence a vast legal literature devoted to understanding various features of different types of property as either attempts to perfect the initial allocation of rights or, more commonly, to reduce the transaction costs associated with those rights.[290]

In contrast to contractual rights that bind only the parties to an agreement, property creates rights against the whole world. Thus, as Thomas Merrill and Henry Smith have argued, property rights attached to a "thing" impose "an informational burden" on all those who are likely to interact with that "thing."[291] Merrill and Smith argue further that the broad application of the informational burden of property rights explains the tendency of these rights to come in a fixed menu of forms. Thus the law reduces transaction costs by limiting property rights to a set of standardized packages that the layperson can understand at low cost.[292]

Consideration of information costs suggests a possible divergence between property in tangible and intangible objects – whereas the informational burden of tangible property is limited by physical proximity, those same burdens can multiply almost infinitely in the case of intellectual property. Only those walking past Blackacre need to worry where it boundaries are; every musician in the world needs to worry that their new composition might have inadvertently copied from any one of thousands of pop songs their brains have absorbed over the years.[293] Another difference is also worth noting – physical objects suggest at least a core definition congruent with their physical attributes, whereas property rights in intangibles are purely a legal construct.[294] In other words, while the rights attached to real property and chattels might be fuzzy at the edges, the rights attached to copyrighted expression (as opposed to the piece of paper on which that expression resides) are fuzzy all the way through to the core.[295] Accordingly, the

---

[289] Ronald Coase, *The Nature of the Firm*, 4 ECONOMICA 386 (1937) (observing that differences in transaction costs explain variation in organizational hierarchy).

[290] The classic work in this area being, Guido Calabresi & A. Douglas Melamed, *Property Rules, Liability Rules and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089 (1972) (discussing the transaction costs implications of different remedial structures).

[291] Thomas W. Merrill & Henry E. Smith, *What Happened to Property in Law and Economics?*, 111 YALE L.J. 357, 358 (2001). See also, Thomas W. Merrill & Henry E. Smith, *The Property/Contract Interface*, 101 COLUM. L. REV. 773 (2001) (explaining property and contract law in terms of the information costs of *in rem* and *in personam* rights); Thomas W. Merrill & Henry E. Smith, *Optimal Standardization in the Law of Property: The Numerus Clausus Principle,* 110 YALE L.J. 1, 3-9 (2000) (arguing that the standardized forms of property reduce transaction costs).

[292] *Id.*

[293] In one noteworthy case, George Harrison was found to have "subconsciously plagiarized" the 1963 hit *He's So Fine* in his 1970 single, *My Sweet Lord*. The alleged similarities escape this author. Bright Tunes Music v. Harrisongs Music. 420 F. Supp. 177 (S.D.N.Y. 1976).

[294] See, WILLIAM LANDES & RICHARD POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 16 (2003) (noting that transaction costs tend to be higher in intellectual property because it is frequently difficult to identify such property because by definition it has no unique physical site).

[295] As much was acknowledged by Judge Learned Hand when he said that "as soon as literal appropriation ceases to be the test [for copyright infringement], the whole matter is necessarily at large" Nichols v. Universal Pictures Corp., 45 F.2d 119, 121 (2d Cir. N.Y. 1930).

definition of intellectual property rights must be even more sensitive to transaction costs, not just those between willing parties, but those imposed on the rest of the world.[296]

Numerous legal commentators have offered transaction cost and information cost explanations for specific attributes of intellectual property law. In *The Economic Structure of Intellectual Property Law*, William Landes and Richard Posner explain a diverse range of intellectual property doctrines as efficient responses to transaction costs problems. With respect to copyright they argue that the need to keep transaction costs low explains the idea-expression distinction, the limited duration copyright, and the derivative work right.[297] In a more explicitly comparative vein, Clarisa Long argues that many differences between patent and copyright law stem from divergences in the information costs and coordination problems associated with expressive works and inventions.[298] In a recent article, Henry Smith also applies information-cost theory to explain certain differences between copyright and patent law, arguing that the former is more tort-like and the later more property-like.[299]

Transaction costs are not only important in establishing certain features of copyright doctrine, an assessment of transaction costs is also a key internal feature of specific copyright doctrines, most notably fair use. The central purpose of the fair use doctrine is to permit certain uses that would otherwise be infringing. Indeed, as Wendy Gordon has shown, the presence of high transaction costs and other market failures provides a useful framework for understanding the entirety of the fair use doctrine.[300] The fair use doctrine plays a critical role where the copyright owner withholds permission for reasons that we as a society find unacceptable, such as to stifle parody, criticism, or social debate.[301] However, the doctrine is not limited to the suppression of criticism and social debate. Fair use is necessary even when copyright owners are purely commercially motivated because licensing and other private ordering mechanisms do not provide a solution for cases involving high exchange costs, high information costs and strategic behavior. The fair use doctrine is particularly important in situations where the costs of obtaining permission outweigh the benefits of the use. Thus, according to one court at least, the fair use

[296] See, *Id*. at 21. (concluding that intellectual property rights tend to be more costly than rights in physical property).
[297] *Id*. at 92-93, 21, and 111.
[298] Clarisa Long, *Information Costs in Patent and Copyright*, 90 VA. L. REV. 465 (2004).
[299] Henry E. Smith, *Intellectual Property as Property: Delineating Entitlements in Information*, 116 YALE L.J. 1742 (2007). See also, Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783 (2007) (making a transaction costs argument against the imposition of injunctions in cases where courts cannot easily tailor injunctions to forbid only the prohibited conduct); Samson Vermont, *Independent Invention as a Defense to Patent Infringement*, 105 MICH. L. REV. 475 (2006) (arguing that a defense of independent invention would reduce information costs associated with the patent system); Christopher A. Cotropia, *Patent Claim Interpretation and Information Costs*. 9 LEWIS & CLARK L. REV. 57 (2005) (concluding that full use of the patent specification early in the claim interpretation process would minimize information costs); Paul J. Heald, *A Transaction Costs Theory of Patent Law,* 66 OHIO ST. L.J. 473 (2005) (justifying patent law based on private transaction costs savings rather than the more conventional incentive-based rationale).
[300] Gordon, *supra* note 000
[301] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994); SunTrust Bank v. Houghton Mifflin Co., 268 F.3d 1257 (11th Cir. 2001).

doctrine protects book reviews because in the absence of a fair-use doctrine, most publishers would disclaim control over the contents of reviews in any event.[302] Consequently the fair use doctrine economizes on transaction costs by making such disclaimers unnecessary.[303]

While the role of fair use in addressing high costs of exchange is ground familiar to most copyright academics, the significance of fair use as a remedy to strategic behavior is less familiar.[304] Law and economics scholars are used to thinking about the implications of strategic behavior in divided ownership contexts such as oil field unitization or corporate governance situations.[305] The basic problem is that where several parties posses a veto right that can block some profitable enterprise – a new stadium, oil well, or corporate merger – each has an incentive to "hold out" for a disproportionate share of the gains to be had from that enterprise. Both experience and theory suggest that the mere presence of a surplus to be divided does not ensure that the parties will in fact agree on how that division should proceed.[306] Furthermore agreements on division are constrained by the costs of enforcement given that once a deal has been agreed, "[a]ccording to strictly wealth-maximizing behavioral assumptions, a party to exchange will cheat, steal, or lie when the payoff to such activity exceeds the value of the alternative opportunities."[307]

The problems attending strategic behavior are particularly relevant to copyright because all copyrighted works are built on previous works to some extent. Musicians attempting to clear samples often face license demands from the original copyright owners that effectively seek to expropriate the entire value of the newly created work.[308] Strategic behavior may prevent parties who would other have much to gain from cooperating if multiple clearances are required because it is quite rational for the players to adopt

[302] Ty, Inc. v. Publ'ns Int'l, 292 F.3d 512, 517 (7th Cir. 2002)

[303] *Id.* In this context fair use could also be seen as the solution to a collective action problem because it allows publishers to credibly commit to not censoring reviews. Landes and Posner explain the fair use status of book reviews in terms of implied consent, but collective action and credible commitment are more convincing explanations. Landes & Posner, *supra* note 000 at 120-121.

[304] *See,* Sag, Beyond Abstraction, *supra* note 000 at 250 (criticizing doctrinal recommendations which aim to optimize copyright scope in the abstract but do not account for the effect uncertainty or strategic behavior).

[305] *See,* Gary D. Libecap & James L. Smith, *The Economic Evolution of Petroleum Property Rights in the United States*, 31 J. LEGAL STUD. 589 (2002); George J. Mailath & Andrew Postlewaite, *Asymmetric Information Bargaining Problems with Many Agents,* 57 REV. ECON. STUD. 351 (1990); Zohar Goshen, *Voting (Insincerely) in Corporate Law*, 2 THEORETICAL INQUIRIES IN LAW 815, 820 (2001) (explaining the holdout problem in the corporate governance setting).

[306] Robert Cooter, *The Cost of Coase*, 11 J. LEGAL STUD. 1, 23 (1982) (arguing that disagreements as to how to divide the contractual surplus may prevent successful Coasean bargaining). See also, Michael A. Heller, *The Tragedy Of The Anticommons: Property In The Transition From Marx To Markets,* 111 HARV. L. REV. 621 (1998) (arguing that underuse results when too many owners hold rights of exclusion in a single resource).

[307] North, supra note 282 at 30.

[308] [See, Andrew Ross, *Princes Among Thieves: Sampling the '80s; Hip Hop Music,* ARTFORUM INTERNATIONAL, March 1, 2003 (reporting that only record companies or performers with deep pockets could viably sample, given the high cost of legal clearance).]

strategies that risk destroying the surplus in order to gain a larger share.[309] Even non-strategic parties sometimes cause hold-out problems because of the divergent valuations that result from egotism and other cognitive biases.[310] The fair use doctrine reduces transaction costs associated with strategic behavior by eliminating the hold-out power of the copyright owner in situations where her contribution is small compared to that of the defendant, or where some degree of copyright owner intransigence is effectively presumed.

The salient point to take away from this discussion so far is that the invocation of the property mantra does very little to tell us whether the rights of a copyright owner include the right not to allow a particular copy-reliant technology to interact with her work in a particular way. A further import point remains: even if that issue is resolved in the copyright owner's favor, the question of the form of that property right still remains at large. The relevant question for current purposes is not just the question of "property rules", "liability rules" and "inalienability rules".[311] Even if we accept that the copyright owners rights are to be protected by a veto right (i.e. a property rule), the question is, on what, if any conditions can this right be exercised? As Abraham Bell and Gideon Parchomovsky observe rather in their own memorable phrasing, entitlements are often dynamic in nature and "pliability" rules – contingent rules that provide an entitlement owner with either property rule or liability rule protection as long as some specified condition obtains – are quite common.[312]

The notion that the rights of the property owner can be protected under permissive default coupled with an opt-out is hardly new. Robert Ellickson famously describes the "fencing out" rule whereby cattle were allowed to roam freely on the property of others unless that property was fenced.[313] Landowners still maintained their property rights, subject to the burden of fencing out neighbors' cattle. Presumably if cattle could read, a sign not unlike the Robots Exclusion Protocol would have been sufficient. Cattle is just one example, indeed, beneath the visage of "property" one sees a variegated landscape with rules tailored accordingly to the differences between rights in blackacre, animals (wild and domestic), oil and gas, water rights (subject to multiple regimes depending on geography and land use), and air rights.[314] Specifically in relation to copyright, the suggestion that authors should be required to accept some cost before their rights can be vindicated is not unprecedented. As Michael Mattioli perceptively notes, "while formal registration and

[309] Robert Cooter & Steven Marks, *Bargaining In the Shadow of the Law: A Testable Model of Strategic Behavior*, 11 J. LEG. STUD. 225, 243 (1982) (arguing that private bargaining to redistribute external costs will not achieve efficiency unless there is an institutional mechanism to dictate the terms of the contract for dividing the stakes).
[310] Robert P. Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75, 89 (1994).
[311] Calabresi & Melamed, *supra* note 000.
[312] Abraham Bell & Gideon Parchomovsky, *Pliability Rules*, 101 MICH. L. REV. 1, 5 (2002).
[313] *See e.g.*, ROBERT C. ELLICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES 76 (1991). But note that here norms appeared to govern behavior regardless of the underlying law.
[314] Richard A. Epstein, *Intellectual Property: Old Boundaries and New Frontiers*, 76 IND. L.J. 803, 804–805 (2001).

deposit are no longer strict requirements for copyrightability, both are demanded of authors who wish to bring infringement suits."[315]

Although copyright is primarily a system of property rights, it has no uniform or immutable character. Just as the invocation of the property does not settle disputes over the scope of rights, in the final analysis it is also less than definitive about the form of those rights which are ultimately recognized. The centrality of transaction costs in modern property theory and the practical importance of opt-out mechanisms in reducing transaction costs for copy-reliant technologies make some analysis of the doctrinal implications of opt-outs essential. As in so many other cases, the fair use doctrine provides the most natural framework for that analysis. The next section specifically addresses this question by analyzing the implications of opt-out mechanisms in the context of copyright's fair use doctrine.

C. The Significance of Opt-outs in Fair Use Analysis

1. The "Purpose and Character" of Opt-outs

Assessing the relevance of an opt-out mechanism to the first statutory fair use factor – the "purpose and character of the use"– requires some kind of meta-theory as to what types of uses should be preferred.[316] The Copyright Act itself is not particularly instructive as to what uses should be preferred under this factor, thus courts must inevitably revert to the fundamental principles of copyright law itself.[317]

As discussed at length in Part II, one of these fundamental principles is that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. This follows from the essential observation that purpose of copyright is to protect authors from the unfairness of having their own original expression used in competition against them as a substitute for their work. Although the Supreme Court's most recent guidance stresses the question of transformativeness,[318] the transformative use doctrine is but one manifestation of the broader principle of expressive substitution.

In addition to this core concept of expressive substitution, courts should also consider institutional design of copyright. Copyright achieves its constitutional purpose – the promotion of progress in science and useful arts – "by establishing a marketable right to the use of one's expression";[319] this marketability not only encourages authorship, it decouples authorship from the corrupting influences of state subsidy and elite patronage.[320] Copyright is not the only way to encourage authorship, a system of state prizes might do that just as well. The advantage of copyright over other systems is that it

[315] Michael R. Mattioli, *Opting Out: Procedural Fair Use,* 12 Va. J.L. & Tech. 3**,** 20 (2007).
[316] 17 U.S.C. § 107(1).
[317] Sag, *God in the Machine*, supra note 000 at 385.
[318] *See* Campbell v. Acuff-Rose Music, 510 U.S. 569, 579 (1994). See also, Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1111 (1990).
[319] Eldred v. Ashcroft, 537 U.S. 186, 219 (2003) (citing Harper & Row 471 U.S., at 558).
[320] Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288 (1996).

not only encourages authorship, it also gives authors a degree of autonomy. Accordingly, in cases where fair use is arguable, but not clear-cut, courts should consider to what extent the defendant's conduct as a whole undermines or enhances the autonomy of the relevant class of authors. Put another way, when in doubt, courts should maximize choice by setting default rules that reduce transaction costs. In this context, the effect of opt-out mechanisms moves from the periphery of the fair use question to center stage.

Once judges appreciate that the effect of legal rules is to establish default positions which are then subject to modification, the effect of opt-out mechanisms becomes a significant consideration in fair use cases. A finding of fair use conditioned on the existence of a low cost opt-out mechanism poses far less risk to the autonomy of the author than a finding of fair use with no such mechanism. Furthermore, such a finding may be the only way to overcome certain information asymmetries and problems associated with strategic behavior. Consider the following scenario.

Assume that the majority of authors would actually benefit from the defendant's proposed use, but that a minority objects. Assume further that the cost of affirmatively clearing rights for the defendant is very high but the cost taking advantage of the opt-out mechanism provided by the defendant is very low. In this situation, if the court determines that the defendants use is fair, the majority's preferences are satisfied and the minority must either tolerate the defendant's use or negotiate with the defendant to abate the use. This is not a common outcome in copyright because in most situations it is unreasonable to expect that copyright owners would be able to contract around a default of permission given the multiplicity of potential users they would have to deal with.[321] If the court determines that the defendant's use is fair *subject to an opt-out*, the majority's preferences are satisfied and the minority must either tolerate the defendant's use or incur the cost of the opt-out in order to have their preferences satisfied. Under our assumption that the cast of the opt-out is relatively low, the autonomy of the majority and the minority is preserved, subject only to the cost of opting out. Finally, if the court determines that the defendants use is unfair, then it is the majority who must bear the burden of opting in. The cost of opting in might be especially high because of coordination problems in situations where the use only has value if it reaches a critical mass, or where copyright owners simply lack information about the needs of potential users. If the cost of opting in is effectively preclusive, a denial of fair use will bind the majority to the will of the minority. On the other hand, if the cost of opting in is *de minimums*, both the majority and the minority will have their preferences respected. From an economic perspective, the efficiency of an opt-in versus and opt-out system will depend on the ratio of those who prefer inclusion to those who do not and on the comparative costs of opting-in versus opting out.

Field v. Google illustrates the particular relevance of these criteria to copy-reliant technologies. In the *Field* case, as in many copy-reliant technology cases, taking advantage of the opt-out mechanism was virtually costless. Indeed, the district court found that disabling the cache functionality for any of the pages on Field's website would

[321] The inability of the copyright owner to buy the silence of her critics is of course desirable in many cases where the fair use in question takes the form of political or social commentary.

have taken him a matter of seconds.[322] The legal significance of this finding was directly tied to the question of autonomy. In the court's view, given the easy availability of the opt-out mechanism, it was in fact the plaintiff copyright owner and not Google who effectively controlled whether cached links would appear in relation to his web-pages.[323] Thus by placing control in the hands of site owners, the "character and purpose" of Google's use of the copyrighted material was not unfair. *Field v Google* also illustrates the evidentiary value of an established opt-out mechanism where the copyright owner claims that although the use in question may appear to be non-expressive, it nonetheless poses some hypothetical danger of expressive substitution. As the *Field* court observed: "[t]he fact that the owners of billions of Web pages choose to permit these links to remain is further evidence that they do not view Google's cache as a substitute for their own pages."[324] In other words, the presumed acquiescence of a large number of copyright owners who could very cheaply opt-out indicates that expressive substitution is unlikely.

In sum, the relevance of the existence of a low cost opt-out mechanism to the first fair use factor is that it affects the purpose and character of the defendants use in certain situations. Courts should consider whether (i) the defendant's proposed use is one which the majority of effected copyright owners would actually favor; (ii) the costs of taking advantage of the opt-out are sufficiently small such that the autonomy of the minority is preserved; and (iii) the costs of opting-in would be high enough to threaten the autonomy of the majority under an opt-in rule. Under these circumstances allowing the defendant to proceed subject to an opt-out will significantly reduce transaction costs, thus benefiting the defendant, a majority of effected copyright owners and preserve the autonomy of the minority. In choosing rules that facilitates private ordering through opt-outs, judges can stay true to copyright's basic design by maintaining the autonomy of the author and allowing breathing space for later generations to make their own contributions.

2. The Market Effect of Opt-outs

The presence of an opt-out mechanism is also potentially relevant under the fourth fair use factor, "the effect of the use upon the potential market for or value of the copyrighted work."[325] The primary considerations here are similar to those stated above. If the cost of taking advantage of an opt-out mechanism provided by the defendant is very low, then it is hard to see how a finding of fair use subject to an opt-out could have a harmful effect on the "potential market for or value of the copyrighted work." Indeed, it seems strange at first blush that any copyright owner would bother to object to a permissive default coupled with an opt-out rather than simply exercising the opt-out.

There is in fact a logical explanation for this behavior. A rational copyright owner will insist on a veto right rather than the right to opt-out under either one of two conditions: (i) where the expected costs of obtaining and exercising a veto are lower than the expected

[322] Field v. Google Inc., 412 F. Supp. 2d 1106, 1119 (D. Nev. 2006).
[323] *Id.*
[324] *Id.*
[325] 17 U.S.C. § 107(1).

costs of taking advantage of the equivalent opt-out; or (ii) where the expected benefits of exercising a veto are greater than those that can be obtained by merely opting out.

The first condition is easily illustrated. It would, for example, place an intolerable burden on the average mystery writer if she had to contact each book publisher in the United States and inform them that she did not wish them to publish her manuscript. In that case, the expected costs of a veto over publication are substantially lower than the costs of an alternative opt-out regime. In the context of copy-reliant technology, however, the opposite often holds true. For example, the average burden of the opt-out default policed by the Robots Exclusion Protocol is extremely slight because the copyright owner need only attach one notice to communicate to all comers.

The second condition in which a copyright owner would insist on a veto right rather than the right to opt-out is where the benefits of exercising a veto are greater than those that can be obtained by merely opting out. This condition can be met, as the *Perfect 10* case illustrates, where the right to opt-out is ineffective because the copyright owner has lost control of the uses of her works by infringing third parties. Perfect 10 is both a beneficiary and a victim of the open end-to-end architecture of the Internet. The openness of the Internet gives Perfect 10 access to an enormous market unconstrained by geography and zoning laws – the later being especially important to the "adult" content market. However, the openness of the Internet also enables third parties to infringe Perfect 10's copyrights in ways that can be hard to detect or enforce. By taking advantage of the Robots Exclusion Protocol, Perfect 10 has opted out of inclusion in image based search engines but elected to remain visible to traditional text oriented searches. In this way, Perfect 10 is a direct beneficiary of default rules and opt-out mechanisms that prevail on the Internet *vis-à-vis* traditional search engines. Nonetheless, in a series of court battles, Perfect 10 has chosen to attack this same institutional setup in relation to image based searching. The reason is simple, although opting out of image based searching is low-cost for Perfect 10, it is also ineffective because Perfect 10 has not been able to prevent third parties from infringing its works. Perfect 10 thus illustrates the problem of a minority hold-out to a transaction cost reducing mechanism. The decision for the court in this case was to determine how the benefits of the default of inclusion weighed against the costs imposed on minorities such as Perfect 10. By suing Google rather than the websites that illegally hosted its photos, Perfect 10 was effectively asking the courts to shift the costs of copyright enforcement onto Google and the public at large that benefits from image-based searching. Seen in this light, the court's conclusion that the benefits of the open default outweighed the limited costs to Perfect 10 is quite understandable.[326]

The second condition can also be met, as the *Field* case illustrates, where the copyright owner believes that she can strategically use a veto right to extract some of the surplus value in a joint enterprise contributed by authors who consent to the use of their works or the independent investment of the defendant. The trial court found that Field had no

[326] *Id.* at 725. ("In this case, Google has put Perfect 10's thumbnail images (along with millions of other thumbnail images) to a use fundamentally different than the use intended by Perfect 10. In doing so, Google has provided a significant benefit to the public.").

genuine objection to the default rules and opt-out mechanisms that prevail on the Internet. Indeed, by his own admission, Field's objection was purely a strategic attempt to extract rents from Google.[327] Field argued that Google's caching functionality harmed the market for his works by depriving him of revenue he could have obtained by charging Google for the right to present cached to his web-pages. The court rejected this transparently circular argument, noting that "the fourth fair use factor is not concerned with such syllogisms."[328] As discussed in Part II, courts limit the potential circularity of the fourth factor by limiting the market for potential derivative uses in a number of ways. The reason that the court excluded the copyright owner's bootstrapping claim of a market effect in this case was that he was not seeking to extract the value that Google derived from access to his works – which was almost certainly nil – rather he was attempting to extract value based on the permission costs he could impose on Google in relation to other copyright owners.

The *Field* case raises an issue of more general application: How should courts treat strategic rent-seekers in copyright disputes? In the ordinary course, a copyright owner should be entitled to hold-out for whatever she thinks the use of her particular work is worth. That kind of rent seeking is the mechanism through which copyright provides an incentive to creativity in the first place. However, in the context of copy-reliant technologies at least, courts should be disinclined to allow one copyright owner to expropriate the value added by other copyright owners. A copyright owner might argue that the effect of one copyright owner's veto on other copyright owners is irrelevant to the fourth factor because the particular language of the statute refers to "the effect of the use upon the potential market for or value of *the* copyrighted work." However, such a narrow reading of the fourth factor is unsustainable in light of the purpose and structure of copyright law.

To begin with, the very nature of common law adjudication demands that courts should consider the welfare of copyright owners beyond the plaintiff. Litigated cases not only settle disputes between parties, they also set rules and precedents that extend far beyond the specific parties to the litigation. Courts should therefore consider the likely market effect of their decisions on copyright owners generally, not merely the particular plaintiff before them;[329] in doing so they are more likely to set beneficial precedents of general application. The Supreme Court's admonition in *Campbell* – to consider the four statutory factors in light of the purposes of copyright – also requires courts to consider their decisions in light of their more general effect on the progress of science and useful arts. [330]

The second reason that courts should consider the welfare of copyright owners beyond the plaintiff relates back to the role of autonomy in copyright law. As discussed, copyright achieves its constitutional purpose – the promotion of progress in science and

[327] *Id.* at 1113.
[328] *Id.* at 1121. (quoting Campbell, 510 U.S. at 592.). *See also,* Religious Tech. Ctr., 907 F. Supp. at 1378 n.25; Mattel Inc. v. Walking Mt. Prods., 353 F.3d 792, 806 (9th Cir. 2003).
[329] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 562 (1985)
[330] Campbell 510 U.S. at 577-78.

useful arts by establishing a marketable right to in original expression.[331] The author's marketable right in her expression is not merely an instrument of incentive; it is also an instrument of autonomy because it leaves the author free to choose her own path, significantly less reliant on state or elite subsidy. If the autonomy of the author is a free-standing policy goal of the copyright system, then the effect of default rules on autonomy must be considered under the fourth factor. In a scenario where the majority of copyright owner's would consent to inclusion and but the costs of individualized permission are much greater than the costs of opting out, the failure of a court to find fair use may effectively bind the majority to the will of the minority. While this is by no means preclusive of a finding against fair use, it is clearly a significant consideration under the fourth factor.

Google Book: An Illustration. The Google Book controversy provides an excellent illustration of the ramifications of this broader view of market effect under the fourth factor. This section attempts to briefly map out the likely effects of a fair use ruling which allows Google to continue its project, subject to an opt-out. The first thing to consider is the nature and utilities of the authors who may be affected by a ruling in the Google Book case. In general, we need to consider three types of author: (1) those who favor inclusion in the Google Book database, (2) those who are opposed to inclusion, and (3) those that favor inclusion, but are inclined to strategically object in order to extract rents. These categories are not necessarily fixed, because an author's approval or opposition to inclusion may vary according to the number of other authors who take part.

If the courts do not find that Google's intermediate non-expressive copying constitutes a fair use, then it is likely that Google will only be able to build a shadow database consisting primarily of low permission cost works – i.e. works published before 1923 or after 2000.[332] Authors who strongly favor inclusion could still opt-in, but the costs of credibly making that commitment are likely to be high for works published before 2001.

Against this background, a finding of fair use, subject to the opt-out will have the following welfare effects on authors.[333] First, it will increase the welfare of authors who favor inclusion because their works will now be included without the (often preclusive) cost of opting-in. Those authors will not only benefit from inclusion, they will also benefit from the inclusion of other authors which increases the value of the network to all participants. Second, judicial approval of the opt-out will reduce the welfare of authors who truly disfavor inclusion in the database. Those authors will either suffer the cost of opting-out or, if their preference for isolation is sufficiently weak, some harm from inclusion that is less than the cost of opting-out. The important point to realize here is that the harm suffered by this second category of authors is effectively capped at the cost of

[331] See *supra* note 187 to 188 and accompanying text.
[332] In *Tasini*, the Supreme Court held in favor of freelance journalists in relation to the inclusion and display of their articles available in online databases. Rather than providing the freelance authors who it had already paid once for their articles with a new stream of revenue, the New York Times and other publishers simply removed most of the freelance pieces from their online databases. See generally, Amy Terry, Tasini *Aftermath: The Consequences of the Freelancers' Victory*, 14 DEPAUL-LCA J. ART & ENT. L. 231 (2004).
[333] The benefits to the public are taken as given.

the opt-out. Third, the welfare impact is uncertain for authors who favor inclusion, but are inclined to strategically object in order to extract rents. These authors will gain the same benefits of inclusion as the first group of authors – their works will be cited, borrowed and purchased more often. Their only potential loss will be from a failure to collect rents based on the permission costs they could otherwise have hoped to threaten to impose on Google and less strategic authors.

Should this peculiar kind of loss be remedial under copyright law? While there is no doubt that an author's failure to extract rent based on the value of her own work is cognizable harm under the Copyright Act; it is far from clear that there is any justification in regarding rent derived purely from the threat of imposing costs on the defendant in relation to the enterprise as a whole and thus capturing some of the surplus from the consenting authors in the same light. In this case, some of the value being extracted derives from consenting authors, some from Google's investments in the project and some from the inclusion of works that are in the public domain. Furthermore, if the cost of exercising an opt-out right is in fact small, an author can always collect the rent attributable to their individual value by agreeing, for valuable consideration, not to take advantage of the opt-out. Consequently, where an author declines to use the opt-out, demanding instead the right to brandish a veto, it suggests that she is looking for payment beyond the value of her own contribution. Of course, none of this analysis is meant to be determinative at this stage because the essential facts of the Google Book litigation are still being developed through the ongoing litigation process.

***

A combination of permissive defaults and opt-out mechanisms is a common feature of the copy-reliant technology. As this Article has shown, opt-outs are the primary mechanism through which copy-reliant technologies mitigate otherwise prohibitive transaction costs. Critics of particular copy-reliant technologies argue that the use of opt-outs should not be tolerated as a means of securing the rights of authors and that opt-outs are irrelevant to the fair use analysis. This Article takes the contrary position. A dogmatic insistence on that literary property extends to every conceivable use of the author's work is both inaccurate as a description of settled law and unhelpful in the context of novel questions at the fringe of copyright law. The Copyright Act itself requires courts to determine the content and form of the rights of authors in response to new developments and the fair use doctrine acts as an instrument of policy delegation in that regard. As this Part has shown, the central role of transaction costs in defining the scope and content of property rights and the specific statutory factors of the fair use doctrine each suggest that the defendant's compliance with an opt-outs regime must be a significant factor in this analysis.

CONCLUSION

In many ways, technology is the dog on copyright's leash. In theory, and occasionally in practice, copyright channels the direction of technological progress; but more often, technology simply drags the law in its wake, going where it will. The pull of recent technological change on copyright law has been demonstrated in this Article. Copy-

reliant technologies – technologies that necessarily copy expressive works in large quantities, but do so for non-expressive purposes – are vital to the operation of the Internet. And yet, because these technologies are so dependant on access to copyrighted works, they are also vulnerable to claims of copyright infringement.

Recognizing the common ground shared by search engines, electronic archives, plagiarism detection software and other copy-reliant technologies, sheds considerable light on the application of copyright law in the Internet era. In relation to the first core question posed by copy-reliant technology, the potential for copyright liability for the expressive use of copyrighted works, this Article has established that acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. To do so would conflict with decades of accumulated precedent that limit the rights copyright owners to those uses of their works that offer some threat of expressive substitution.

In spite of its centrality, the question of non-expressive use may not fully resolve all copyright disputes involving copy-reliant technologies. While the category of non-expressive use is conceptually neat, it may prove messy in implementation. Inevitably, courts will face cases where the line between expressive and non-expressive remains ambiguous. In such cases the effect of opt-out mechanisms offered by the defendant moves from the periphery to the center of legal analysis. Technologically enabled opt-out mechanisms such as the Robots Exclusion Protocol play an essential role in maintaining order on the digital frontier. Such devices are essential to overcoming the otherwise daunting transaction costs facing copy-reliant technologies. Accordingly, to treat the phenomenon of copy-reliant technology comprehensively requires addressing the significance of opt-outs under copyright law.

Copyright law is fluid by design, and nowhere is that fluidity more evident than in the development of the fair use doctrine. Even without the fair use doctrine, the mere invocation of literary property would not settle the scope of the copyright owner's rights or the nature of the remedies to which she is entitled. The fair use doctrine both allows and requires judges to consider market realities in determining the application of copyright law in novel circumstances. To the extent that other commentators have considered the doctrinal significance of transaction costs in relation to isolated issues such as the Google Book project, they have largely missed the point. Judges are not state planners; they should not attempt to use the fair use doctrine to achieve some static allocation of uses for a given set of copyrighted works. What judges should do is apply the fair use doctrine to fashion a set default rules which facilitate the kind of private ordering the copyright has traditionally embraced. In the high transaction costs environment of copy-reliant technologies, this may well mean finding in favor of the user who provides copyright owners with the choice to opt-out.