

2009

Copyright and Copy-Reliant Technology.

Matthew Sag

Loyola University Chicago, msag@luc.edu

Follow this and additional works at: <http://lawcommons.luc.edu/facpubs>



Part of the [Law Commons](#)

Recommended Citation

Sag, Matthew, Copyright and Copy-Reliant Technology, 103 NU L.R. 1607 (2009)

This Article is brought to you for free and open access by LAW eCommons. It has been accepted for inclusion in Faculty Publications & Other Works by an authorized administrator of LAW eCommons. For more information, please contact law-library@luc.edu.

COPYRIGHT AND COPY-RELIANT TECHNOLOGY

*Matthew Sag**

INTRODUCTION.....	1607
I. COPY-RELIANT TECHNOLOGIES AND THE INTERNET.....	1610
A. <i>New Technologies, Copyright Markets, and Copyright Law</i>	1611
B. <i>Four Case Studies of Copy-Reliant Technology</i>	1616
II. THE DOCTRINAL IMPLICATIONS OF NONEXPRESSIVE USE.....	1624
A. <i>The Principle of Nonexpressive Use</i>	1624
B. <i>Doctrinal Incorporation of Nonexpressive Use</i>	1639
C. <i>Fair Use and Nonexpressive Use</i>	1645
III. THE DOCTRINAL SIGNIFICANCE OF TRANSACTION COSTS	1657
A. <i>Transaction Costs and Copy-Reliant Technologies</i>	1657
B. <i>Transaction Costs and Property Rights</i>	1668
C. <i>The Significance of Opt-Outs in Fair Use Analysis</i>	1675
CONCLUSION.....	1681

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 have only just begun to manifest.¹ By expanding the breadth, diversity and sheer

* Assistant Professor, DePaul University College of Law. This Article benefited enormously from Justin Hughes' extensive and insightful comments and from the diligent editorial assistance of David Pekarek Krohn and Gautam Huded. Thanks also to David Fagundes, Brett Frischmann, Elizabeth Townsend Gard, Andrew Gold, Eric Goldman, Bobbi Kwall, Tonja Jacobi, Adam Mossoff, Dotan Oliar, Miquel Peguera Poch, Glen Robinson, Mark Schultz, Christopher Sprigman, Rebecca Tushnet and to the faculties of DePaul University College of Law and the University of Virginia School of Law. An earlier version of this article was presented at the American Intellectual Property Law Association Annual Meeting 2007, the Intellectual Property Scholars Conference 2008, the University of Virginia School of Law Faculty Workshop and the DePaul University College of Law Faculty Workshop. Special thanks to the University of Virginia School of Law Library staff and to Adam Cieslak for their invaluable research assistance. Please address comments to msag@depaul.edu.

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. See HILARY W. POOLE ET AL., *THE INTERNET: A HISTORICAL ENCYCLOPEDIA* 206 (2005). The first TCP/IP-wide area network was operational by January 1, 1983, when the U.S. National Science Foundation (NSF) constructed a university network backbone that later became the NSFNet. *Id.* at 145–46.

number of copyrighted works in existence, the Internet has fundamentally changed the nature of copyright markets. This transformation is most significant in the context of what I term “copy-reliant technologies”—technologies that copy expressive works for nonexpressive ends. Copy-reliant technologies, such as Internet search engines and plagiarism detection software, do not read, understand, or enjoy copyrighted works, nor do they deliver these works directly to the public. They do, however, necessarily copy them in order to process them as grist for the mill, raw materials that feed various algorithms and indices.

Other scholars have considered separately the copyright implications of Internet search engines, plagiarism detection software, reverse engineering of software, and the recently settled Google Book Project controversy.² This Article attempts to provide a unifying theoretical framework for these issues, recognizing them as subparts of a broader phenomenon: the emergence of copy-reliant technology.

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: (1) the copying of expressive works for nonexpressive uses, (2) a high volume of transactions, and (3) the use of technologically enabled opt-out mechanisms to reduce transaction costs. The business models that employ these technologies often inherently require these traits.

The rise of copy-reliant technologies exposes seemingly novel questions. First, should a nonexpressive use, which nonetheless requires copying the entirety of a copyrighted work, be found to infringe the exclusive rights of the copyright owner? Our historical intuition is that works are copied to communicate at least some part of the work’s original expression: books are copied to be read, not to serve as paperweights; and compact discs are copied to be played, not to function as drink coasters. This Article concludes that because expressive communication to the public implicitly defines and limits the extent of the copyright owner’s exclusive rights, acts

² On search engines, see Urs Gasser, *Regulating Search Engines: Taking Stock and Looking Ahead*, 8 YALE J. L. & TECH. 201 (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. Horowitz, *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, for example, 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); Rebecca Tushnet, *My Library: Copyright and the Role of Institutions in a Peer-to-Peer World*, 53 UCLA L. REV. 977 (2006); Siva Vaidyanathan, *The Googlization of Everything and the Future of Copyright*, 40 U.C. DAVIS L. REV. 1207 (2007).

of copying that do not communicate the author's original expression to the public do not generally constitute copyright infringement.

The second important question raised by copy-reliant technologies relates to the opt-out mechanisms built into many copy-reliant technologies. The architects of these technologies have chosen to build in these mechanisms to preserve the autonomy of the copyright owner. These mechanisms, however, switch the default position from "no copying without permission" to one in which copyright owners must affirmatively opt out of specific uses of their works. Accordingly, we face the question of whether this modification of the usual copyright default is justified from either a doctrinal or a utilitarian perspective.

The challenge presented by copy-reliant technology inevitably interacts with the issue of fair use. Technically, the fair use doctrine renders certain otherwise infringing actions relating to copyrighted works noninfringing.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, any examination of copyright and new technology inevitably becomes a reflection on the nature of fair use. This Article explains the correct application of the fair use doctrine in the context of nonexpressive uses. Furthermore, this Article explores the application of fair use 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 technology 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, Inc. v. Amazon.com, Inc.*, centers on the creation and display of thumbnail representations of copyrighted photographs 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 fragments of books as part of a menu of search results.⁷ The final case study looks at plagiarism detection software, which also addresses the use

³ 17 U.S.C. § 107 (2006) ("[T]he 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*, 11 MICH. TELECOMM. TECH. L. REV. 381 (2005).

⁵ 412 F. Supp. 2d 1106 (D. Nev. 2006); *see infra* Part I.B.1.

⁶ 487 F.3d 701 (9th Cir. 2007); *see infra* Part I.B.2.

⁷ Class Action Complaint, Authors Guild v. Google, No. 05-CV-8136 (S.D.N.Y. Sept. 20, 2005); *see infra* Part I.B.3.

of copyrighted works to generate metadata.⁸ These case studies illustrate copy-reliant technology in a number of contexts that are further developed in Parts II and III.

Part II explores the doctrinal implications of the nonexpressive 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 noncommunicative and nonexpressive. A careful review of existing copyright doctrine suggests that the rights of copyright owners do not typically encompass nonexpressive uses of their works. I argue that recognizing a principle of nonexpressive use resolves many questions relating to copy-reliant technologies. It also reconciles many puzzling features of the fair use doctrine more broadly. Part II concludes by addressing the doctrinal incorporation of this principle of nonexpressive use through the application of the fair use doctrine.

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 analyzes 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 section 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 recalibrating copyright law as technology and market conditions change. This framework forms the essential theoretical background for understanding the significance of the copy-reliant technology. Section B describes four case studies of copy-reliant technology that 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 Parts II and III. As the case studies illustrate, copy-reliant technologies tend to raise certain recurring legal issues: the copying of expressive works for nonexpressive uses, the potential for high transaction costs, and the role of opt-out mechanisms in addressing these transaction cost problems.⁹

⁸ *A.V. v. iParadigms*, 544 F. Supp. 2d 473 (E.D. Va. 2008); see *infra* Part I.B.4. Space constraints preclude additional case studies covering areas such as software reverse engineering.

⁹ See *infra* Parts II and III for the further development of these issues.

A. *New Technologies, Copyright Markets, and Copyright Law*

From the printing press to the photocopier, from the piano-roll to the mp3 player, new technologies have fundamentally altered copyright law.¹⁰ Photography, motion pictures, sound recording, and broadcasting have each demanded and (eventually) received accommodation from copyright law.¹¹ 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.¹²

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 allow users to copy and distribute existing works, such as sound recording and motion pictures, at virtually no cost.¹³ Unlocking content from physical delivery has facilitated more than just piracy; it has also enabled legal digital music services that have made more music available at a lower cost than ever before.¹⁴

Advances in technology have also opened up new possibilities of creative production by reducing the cost of sound and video editing. The type of video editing software used to create the *Phantom Edit*¹⁵—a fan edited version of *Star Wars Episode I* without the much-reviled 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.¹⁶ 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

¹⁰ See generally 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 of 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).

¹¹ See generally LITMAN, *supra* note 10 (discussing the legislative history of U.S. copyright law in the twentieth century).

¹² See LAWRENCE LESSIG, *CODE AND OTHER LAWS OF CYBERSPACE*, 22–23 (1999); WILLIAM LANDES & RICHARD POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* (2003).

¹³ See generally Matthew Helton, *Secondary Liability for Copyright Infringement: BitTorrent as a Vehicle for Establishing a New Copyright Definition for Staple Articles of Commerce*, 40 COLUM. J. L. & SOC. PROBS. 1 (2006) (explaining the technological and legal implications of the nearly costless replication of copyrighted data).

¹⁴ See CHRIS ANDERSON, *THE LONG TAIL* 139 (2006) (noting the effect on price).

¹⁵ 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 EPISODE 1.1: THE PHANTOM EDIT (2001), and its creation).

¹⁶ See Wikipedia, *List of Video Editing Software*, http://en.wikipedia.org/wiki/List_of_video_editing_software (last visited July 27, 2009), for a list of video editing software, including several free and open source modules.

medium. The disruptive effect of political blogs on the agenda-setting power of newspapers and television reporting is just one example.¹⁷

To regard these changes, and their implications for copyright law, as a mere continuation of past technological changes would risk overlooking a crucial transformation.¹⁸ Digital technology and the Internet have significantly expanded the scope, diversity, and 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 negligible. Thus, the Internet has not merely induced an increase in copying, but an exponential increase. This is not just a difference in degree, but a difference in kind.¹⁹ 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.²⁰ 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 we should discard the substantial body of copyright law that has developed over the past two centuries. To the contrary, many of the principles and distinctions derived from pre-Internet cases are equally applicable online. Often, the mere fact that copying took place online is of little or no relevance. For example, the legality of the 655,000 self-described parody videos hosted on YouTube²¹ largely depends on the amount of copyrighted material taken by the parodist,²² and on whether courts reasonably perceive the work as a genuine parody of the

¹⁷ See Kevin Wallsten, *Agenda Setting and the Blogosphere: An Analysis of the Relationship Between Mainstream Media and Political Blogs*, 24 REV. POL'Y RES. 567, 567–87 (2007) (finding a complex, bidirectional relationship between mainstream media coverage and blog discussion rather than a unidirectional media or blog agenda-setting effect); see also 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 COMM. Q. 451, 451–66 (2007) (finding no observable differences between the credibility of blogs and that of more traditional media).

¹⁸ See Menell, *supra* note 10, at 64 (discussing the relationship between new technology and new modes of expression).

¹⁹ P.W. Anderson, *More Is Different: Broken Symmetry and the Nature of the Hierarchical Structure of Science*, 177 (4047) SCIENCE 393 (1972).

²⁰ See LITMAN, *supra* note 10.

²¹ This figure is based on a YouTube.com search for the term parody, http://www.youtube.com/results?search_type=&search_query=parody&aq=f (search performed July 1, 2009).

²² *Berlin v. E.C. Publ'ns, Inc.*, 329 F.2d 541 (2d Cir. 1964).

copyright owner's work.²³ These considerations are the same today as they were in 1994, when the Supreme Court last addressed the issue.²⁴

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 digital technology and the Internet have enabled new forums and new ways to interact with copyrighted material, often by copying portions of it.²⁵ For example, while cutting out an article from a newspaper and sending it to a friend through the U.S. postal service does not implicate copyright, posting the contents of the same article on an Internet discussion board or blog necessitates copying and thus raises the specter of copyright infringement. The fair use doctrine has also become more important to the public as the sharing of user-generated content online enables copyright owners to detect infringement more easily.²⁶ The scope of fair use in relation to YouTube videos, blogs, and other forms of user-generated content is an important issue, but outside the scope of this Article. Instead, this Article explores a different set of issues that augurs a more fundamental change in the way we think about copyright and fair use. Indeed, this Article seeks to refocus the copyright debate, shifting it away from the perennial concern about freedom of expression and onto the important topic of nonexpressive use.²⁷ Specifically, this Article addresses the operation of copyright law in relation to copy-reliant technologies, such as Internet search engines, electronic archives, plagiarism detection software, and other applications that rely on copying expressive works for nonexpressive ends.

These technological changes are significant for copyright because by enabling more people to produce a greater range of copyrighted material, 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 to give authors an incentive to create those works in the first place.²⁸ However, these same exclusive rights raise the cost for consumers to acquire those works, and for subsequent authors to create further expressive works. In

²³ See *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994); see also *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001); *Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc.*, 109 F.3d 1394, 1399 (9th Cir. 1997).

²⁴ *Campbell*, 510 U.S. 569.

²⁵ See *L.A. Times v. Free Republic*, 2000 U.S. Dist. LEXIS 5669 (C.D. Cal. Mar. 31, 2000).

²⁶ See Molly Shaffer Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535, 1537–39 (2005).

²⁷ See, e.g., Neil Weinstock Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1, 42 (2001).

²⁸ U.S. CONST. art. I, § 8, cl. 8.

the world of tangible objects, these costs become price signals that ensure the efficient allocation of goods to those who value them most. Given, however, 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 deadweight loss because those who are unwilling to pay the higher price are forced to go without the work in question.²⁹

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 could, 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, it forces those who are unwilling to pay the higher price to go without the work in question.³⁰ 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.³¹

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" has long been understood.³² 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.³³ In many situations, authors can license their creations with relative ease and the theoretical loss of exclusion is minimal.³⁴ 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,"³⁵ while other times it is more like a cudgel. For example, Stephen Joyce has been accused of attempting to control access to unpublished ma-

²⁹ 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).

³⁰ This assumes, realistically, the absence of perfect price discrimination. See generally 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).

³¹ See, e.g., STEVEN SHAVELL, *FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW* 11–23 (2004) (describing the effect of exclusion on resource allocation).

³² *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).

³³ See Sag, *supra* note 29, at 208–15 (discussing the relationship between copyright scope and the effectiveness of private ordering).

³⁴ See GOLDSTEIN, *supra* note 10, at 5 (discussing product differentiation through versioning in the book publishing and motion picture industries).

³⁵ *Id.*

terial in order to influence historical and literary conceptions of his grandfather, James Joyce, whose literary estate he controls.³⁶ 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.³⁷ Biographers of Howard Hughes have faced similar difficulties.³⁸

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, reusing, or reinterpreting the facts and ideas embodied in the original work.³⁹ Statutory exemptions and compulsory licenses—such as the special reproduction rights of libraries and archives,⁴⁰ and the compulsory license for making and distributing audio recordings⁴¹—also provide some breathing space. Their scope, however, 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.

Fair use is a flexible standard that limits the scope of copyright protection and renders certain actions relating to copyrighted works noninfringing.⁴² 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.⁴³ Judges and legal scholars frequently attest to the importance of the fair use doctrine,⁴⁴ but in spite of its partial codification in the Copyright Act of 1976, the exact nature of fair use remains elusive and resists straightforward definition.⁴⁵

³⁶ See *Shloss v. Sweeney*, 515 F. Supp. 2d 1083 (N.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 (providing an account of Stephen Joyce's various threats of copyright litigation).

³⁷ *Shloss*, 515 F. Supp. 2d 1083.

³⁸ *Rosemont Enters., Inc. v. Random House, Inc.*, 366 F.2d 303 (2d Cir. 1966).

³⁹ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984); see also *Warner Bros., Inc. v. Am. Broad. 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”).

⁴⁰ 17 U.S.C. § 108(a) (2006).

⁴¹ 17 U.S.C. § 115.

⁴² 17 U.S.C. § 107.

⁴³ See U.S. COPYRIGHT OFFICE, REPORT OF THE REGISTER OF COPYRIGHTS ON THE GENERAL REVISION OF THE U.S. COPYRIGHT LAW (1961).

⁴⁴ See, e.g., *Ty, Inc. v. Publ'ns. Int'l., Ltd.*, 292 F.3d 512, 518 (7th Cir. 2002) (Posner, J.) (noting that fair use doctrine plays an essential role in copyright law).

⁴⁵ See *Dellar v. Samuel Goldwyn, Inc.*, 104 F.2d 661, 662 (2d Cir. 1939) (describing fair use as one of the most troublesome doctrines in the entire law of copyright).

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.⁴⁶ 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 society finds unacceptable. For example, a copyright owner usually cannot deny permission to copy in order to stifle parody, criticism, or social debate.⁴⁷

B. Four Case Studies of Copy-Reliant Technology

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

Search technology is clearly a significant public policy issue.⁴⁸ The Internet has become integral to modern existence. For many, it is the dominant medium of communication, research, entertainment, social interaction, and political participation.⁴⁹ Search technology drives the Internet.⁵⁰ Without reliable search technology, the world's 1.7 billion Internet users⁵¹ 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.⁵² Search engines allow users to sift through massive amounts of data to find the specific information that is of particular interest to them. The In-

⁴⁶ See Sag, *supra* note 29, at 250 (criticizing doctrinal recommendations which aim to optimize copyright scope in the abstract but do not account for the uncertain effects or strategic behavior); see also Gideon Parchomovsky & Kevin A. Goldman, *Fair Use Harbors*, 93 VA. L. REV. 1483 (2007) (discussing the necessity of fair use).

⁴⁷ See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994); see also *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001).

⁴⁸ Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000).

⁴⁹ See Press Release, Pew Res. Ctr., Social Networking and Online Videos Take Off, Internet's Broader Role in Campaign 2008 (Jan. 11, 2008), available at <http://people-press.org/reports/pdf/384.pdf> (summarizing survey data).

⁵⁰ 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)*, CORP. FOR NAT'L RES. INITIATIVES (Dec. 1999), available at http://www.cnri.reston.va.us/what_is_internet.html.

⁵¹ Internet World Stats, World Internet Usage and Population Statistics, <http://www.Internetworldstats.com/stats.htm> (last visited Nov. 22, 2009).

⁵² See *infra* note 275 and accompanying text.

ternet would function very differently and much less efficiently without search engines.⁵³ Indeed, users might not find a great deal of content on the Internet without search engines, which begs the question of whether anyone would have the incentive to create or post content in the first place.⁵⁴

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. Automated software agents of the search engines continuously “crawl” across the Internet copying web pages. These copies form the raw data underpinning these search engines, which is later analyzed and cataloged.⁵⁵ As part of this process, search engines both copy and index each web page they find. They then store the HTML code from those pages in a temporary repository called a cache.⁵⁶

Search engines direct users to particular websites based on the relationship of their search term to the index of pages maintained by the search engine provider.⁵⁷ Typically, search engines display search results in a list that features both the title of the relevant web page and a short “snippet” or extract from the targeted web page. In the popular Google search engine, two hyperlinks follow the snippet. One goes to the actual web page, and the other goes 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 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 engines vulnerable to

⁵³ Web directories which list websites by category and subcategory offer an alternative to search engines. However, general directories require an elaborate system of categorization that tends to ossify and become redundant as the context for information retrieval and the relevance of existing information changes.

⁵⁴ Introna & Nissenbaum, *supra* note 48; see generally JOHN BATTLE, *THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE* (2005) (providing an account of the history and significance of Internet search).

⁵⁵ It is important for the discussion that follows to note that caching (i.e., copying web pages) is distinct from, and precedes, indexing. See Sergey Brin & Lawrence Page, *The Anatomy of a Large-Scale Hypertextual Web Search Engine*, 30 *COMPUTER NETWORKS AND ISDN SYSTEMS* 107 (1998), available at <http://infolab.stanford.edu/~backrub/google.html>.

⁵⁶ 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, <http://searchenginewatch.com/showPage.html?page=3628341> (estimating Google’s market share at 58.4 percent, Yahoo!’s at 22.9 percent, and Microsoft’s at 9.8 percent) (last visited Aug. 12, 2009).

⁵⁷ See, e.g., U.S. Patent No. 6,285,999 (filed Jan. 9, 1998) (“Method for Node Ranking in a Linked Database.”).

claims of copyright infringement. Whether these claims are spurious or well founded is the central topic of this Article.

The four case studies that follow provide a brief illustration of the vulnerability of copy-reliant technologies and their associated business models to claims of copyright infringement. The sections that follow briefly describe these cases. The remainder of this Article then explores them in more detail.

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.⁵⁸ 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.⁵⁹

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.⁶⁰ The differences such comparisons reveal can have important political, educational, and legal ramifications.⁶¹ Second, the availability of cached links enables users to understand why the search engine returned a seemingly irrelevant web page in response to their 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 the original copyrighted work could not serve alone.⁶² Likewise, referring to the cache to understand better the relationship between a particular page and a particular search term is also a use that the original copyrighted work could not serve alone.⁶³

⁵⁸ See *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006).

⁵⁹ *Id.* at 1115.

⁶⁰ *Id.* at 1112 (noting that by examining Google's copy of the page, people can identify subtle but potentially significant differences between the current and cached versions of a page).

⁶¹ See generally Matthew Fagan, "Can You Do a Wayback on That?" *The Legal Community's Use of Cached Web Pages In and Out of Trial*, 13 B.U. J. SCI. & TECH. L. 46 (2007) (exploring the implications of web caching for the legal community); see also *infra* note 63.

⁶² *Field*, 412 F. Supp. 2d at 1119.

⁶³ The Internet Archive is also subject to a similar set of copyright issues. The Internet Archive, <http://archive.org>, 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 snapshots of websites as they exist on various days. The Wayback Machine does not direct a user to a live website; instead, it presents the user with a static archived version of the website retrieved from the Internet Archive'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's author. See, e.g., *Healthcare Advocates, Inc. v. Harding, Earley, Follmer & Frailey*, 497 F. Supp. 2d 627, 629 (E.D. Pa. 2007) (holding that viewing and printing archived web pages retrieved from the Wayback Machine was fair use).

Although not expressed in this terminology, the essence of the trial court's finding in *Field* was that the nonexpressive use of the works in the cache did not interfere with the rights accorded to Field as an author.⁶⁴ Field's apparent ability to opt out of inclusion in the Google search engine also impacted the court's decision.⁶⁵ The significance of both of these rationales will be addressed below.⁶⁶

2. *Displaying Search Results: Perfect 10 v. Amazon.*—Google's search engine technology was also at the core of another recent case, *Perfect 10, Inc. v. Amazon.com, Inc.*⁶⁷ Perfect 10 produces and sells copyrighted images of nude models.⁶⁸ It does so primarily through a subscription website.⁶⁹ 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 Perfect 10's authorization.⁷⁰

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 browser takes the user to the original location to retrieve the full-scale image.⁷¹ 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.⁷² The district court 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.⁷³ 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 nonexpressive use paradigm advanced in this Article.⁷⁴

⁶⁴ 412 F. Supp. 2d at 1119.

⁶⁵ *Id.*

⁶⁶ See *infra* Parts II and III.

⁶⁷ 487 F.3d 701 (9th Cir. 2007).

⁶⁸ *Id.* at 710.

⁶⁹ *Id.*

⁷⁰ *Id.* at 711.

⁷¹ *Id.*

⁷² *Id.*

⁷³ *Id.*

⁷⁴ *Id.* at 721 (citation and internal quotation marks omitted); see also *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 819 (9th Cir. 2003) (finding that thumbnails are transformative).

3. *Generating Metadata: The Google Book Project.*—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.⁷⁵ Likened to the Library of Alexandria,⁷⁶ the Google Book Search 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—searchable by anyone with an Internet connection.⁷⁷ To create this search index, Google is currently in the process of digitizing vast collections of books, one page at a time.

Google Book allows users to search inside the text of captured books and to generate a list of books relevant to the user’s search terms.⁷⁸ 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 to such access.⁷⁹ In the default scenario—where Google has not received permission and the book does not appear to be in the public domain—Google Book presents a user who clicks on a book title with bibliographic data about the target book and a small extract or “snippet” of the relevant page containing her search terms.⁸⁰ It also presents users with additional information about the books targeted by their search term, including links to online bookstores and links to nearby libraries where the user can obtain the book.⁸¹

⁷⁵ Google Corporate Information: Company Overview, <http://www.google.com/corporate> (last visited Apr. 7, 2009).

⁷⁶ 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.

⁷⁷ Bob Thompson, *Search Me?; Google Wants to Digitize Every Book. Publishers Say Read the Fine Print First*, WASH. POST, Aug. 13, 2006, at D1.

⁷⁸ Answer, Jury Demand, and affirmative defenses of defendant Google Inc. ¶ 19, *Authors Guild v. Google Inc.*, No. 05-CV-8136 (S.D.N.Y. Nov. 30, 2005); Answer, Jury Demand, and affirmative defenses of defendant Google Inc. ¶ 4, *McGraw-Hill Cos., Inc. v. Google Inc.* No. 05-CV-8881 (S.D.N.Y. Nov. 8, 2005). Unless otherwise stated, discussion of the Google Book project in this Article does not take into account the substantial changes envisaged under the revised class action settlement. At the time this Article went to press, the settlement had yet to be finally approved by the district court. The proposed settlement is available at <http://www.googlebooksettlement.com/>.

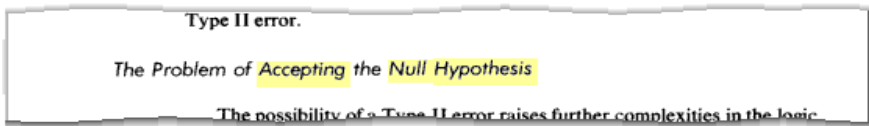
⁷⁹ Answer, Jury Demand, and affirmative defenses of defendant Google Inc. ¶ 4, *McGraw-Hill Cos., Inc. v. Google Inc.* No. 05-CV-8881 (S.D.N.Y. Nov. 8, 2005). Google has several agreements with publishers to do just that. See Complaint ¶ 30, 31, *McGraw-Hill Cos., Inc. v. Google Inc.* No. 05-CV-8881 (S.D.N.Y. Oct. 19, 2005); Answer, Jury Demand, and affirmative defenses of defendant Google Inc. ¶ 30, *McGraw-Hill Cos., Inc. v. Google Inc.* No. 05-CV-8881 (S.D.N.Y. Nov. 8, 2005). Amazon.com’s Search Inside feature also offers similar functionality for the much smaller collection of works for which Amazon.com has been able to obtain permission from the relevant publishers. See *Amazon.com’s Responses And Objections To Subpoena Served By Google Inc., McGraw-Hill Cos., Inc. v. Google Inc. and Authors Guild v. Google Inc.*, No. 05-CV-8136 and No. 05-CV-8881 (W.D.Wash. Oct. 20, 2006).

⁸⁰ Google, What You’ll See When You Search on Google Book Search, <http://books.google.com/googlebooks/screenshots.html> (last visited Apr. 7, 2009).

⁸¹ *Id.*

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."⁸² Entering the search term generates a menu of books containing the term. Selecting any one book leads to a fuller set of information about the book, including snippets from the digitized book 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.

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, or to find the book in a lending library. A second click generates a list of libraries ranked in order of geographic proximity. A third click actually retrieves the call number from, for instance, the University of Virginia Law Library.⁸³ 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.⁸⁴ It might be hyperbolic to suggest that "all the books in the world [will] become a single liquid fabric of interconnected words and ideas,"⁸⁵ but perhaps great advances in human knowledge deserve a little hyperbole.

Not everyone is so enamored. Google has recently settled two significant lawsuits in relation to Google Book.⁸⁶ The first is by the American As-

⁸² Search conducted by the author on April 7, 2009 using the Google Book search engine at <http://books.google.com>.

⁸³ 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* 149 (2d ed. 1980).

⁸⁴ See Edward Wyatt, *Google Adds Library Texts to Search Database*, N.Y. TIMES, Nov. 3, 2005, at C11.

⁸⁵ Kevin Kelly, *Scan This Book!*, N.Y. TIMES, May 14, 2006, §6 (Magazine), at 42.

⁸⁶ See *supra* note 78.

sociation of University Presses.⁸⁷ The second is a class action representing published authors and The Authors Guild.⁸⁸ Both suits sought declaratory and injunctive relief and money damages. The copyright challenge to Google Book focused primarily on the way Google is building its search engine, rather than the output of the search engine per se.⁸⁹ The information contained in the search results of any one Google Book search is not, by itself, likely to infringe the copyright of any author for two reasons. First, most of the information Google Book generates falls into the category of facts about books, which are not protectable by copyright.⁹⁰ Second, even the snippets of material that Google directly copies from the print version of a book will not amount to copyright infringement because the amounts taken are too fragmented and insignificant to constitute a substantial reproduction of the original work.⁹¹

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 web pages as part of the indexing process, the Google Book project requires the routine, automatic, and indiscriminate copying of printed library books whether they are likely to be protected by copyright or not. 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 consistent with 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 have the ability to opt out.⁹² The implications of both of these features are discussed in more detail in Parts II and III.

⁸⁷ 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. See *McGraw-Hill Cos., Inc. v. Google Inc.*, Civil Action No. 05-CV-8881 (S.D.N.Y. Oct. 19, 2005).

⁸⁸ The Authors Guild filed a lawsuit in relation to Google's scanning and digitizing of library books on September 20, 2005. See *Class Action Complaint, Authors Guild v. Google Inc.*, No. 05-CV-8136 (S.D.N.Y. Sept. 20, 2005).

⁸⁹ Admittedly, the Authors Guild's Class Action Complaint is not so precise. See *First Amended Class Action Complaint* ¶¶ 3-4, *Authors Guild v. Google Inc.*, No. 05-CV-8136 (S.D.N.Y. July 24, 2006).

⁹⁰ See *infra* Part II.B.

⁹¹ See *infra* Part II.B. This may not be so with respect to poetry, dictionaries, drug reference guides, price guides, and books of quotations. Google intends to provide snippet previews of such works only with authorization from the rightsholder. See Google, *Google Book Settlement*, <http://www.googlebooksettlement.com/help/bin/answer.py?answer=118722> (last visited Aug. 4, 2009).

⁹² The mechanics of the opt-out mechanism have been considerably refined in the proposed Google Book Settlement. See Matthew Sag, *The Google Book Settlement and the Fair Use Counterfactual* (July 22, 2009) (working paper), available at <http://ssrn.com/abstract=1437812>.

4. *Turnitin.com: Plagiarism Detection Software*.—Several different types of educational institutions have turned to technological solutions to combat the threat of plagiarism.⁹³ Harvard University,⁹⁴ the International Baccalaureate program,⁹⁵ and thousands of high schools across the United States⁹⁶ use plagiarism detection software to detect and deter cheating by their students. Plagiarism detection services, such as those available at 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.⁹⁷

This technology has obvious benefits for educators and for students. However, like other copy-reliant technologies, antiplagiarism software also has its share of critics.⁹⁸ In 2006, students at McLean High School in Virginia objected when the school mandated the compulsory use of antiplagiarism software.⁹⁹ The students took umbrage to both the implied accusation of cheating and to the fact that a commercial software company would be able to make use of their works by adding them to a reference database.¹⁰⁰ Two McLean High School students followed up their protest with a copyright infringement lawsuit against iParadigms, the company that provides the Turnitin.com service.¹⁰¹ 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.¹⁰² Some other plagiarism detection services avoid similar disputes by allowing students to opt out of inclusion in their reference databases.

⁹³ 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, 21 (2007) (citing various studies of academic integrity, including a 1999 survey finding 50 percent of students admitted to Internet plagiarism).

⁹⁴ See *Nation In Brief*, WASH. POST, Nov. 3, 2006, at A15.

⁹⁵ See S. Mitra Kalita, *Schools Turn to Software to Help Stop Plagiarism*, WASH. POST, Apr. 15, 2004, at T4.

⁹⁶ See Andy Dehnart, *The Web's Plagiarism Police*, SALON.COM, June 14, 1999, <http://www.salon.com/tech/feature/1999/06/14/plagiarism> (reviewing several different services); Maria Glod, *McLean Students Sue Anti-Cheating Service; Plaintiffs Say Company's Database of Term Papers, Essays Violates Copyright Laws*, WASH. POST, Mar. 29, 2007, at B5; Brock Read, *Anti-Cheating Crusader Vexes Some Professors*, 54 CHRON. HIGHER EDUC. Issue 25, Feb. 29, 2008, at A1, available at <http://chronicle.com/free/v54/i25/25a00101.htm>.

⁹⁷ See Turnitin.com, *Proprietary Matching Technology*, <http://turnitin.com/static/plagiarism.html> (last visited Apr. 7, 2009).

⁹⁸ See, e.g., CONFERENCE ON COLLEGE COMPOSITION AND COMMUNICATION, *CCCC-IP CAUCUS RECOMMENDATIONS REGARDING ACADEMIC INTEGRITY AND THE USE OF PLAGIARISM DETECTION SERVICES* (2006), available at <http://cccip.org/files/CCCC-IP-PDS-Statement-final.pdf> (arguing somewhat incoherently that antiplagiarism software undermines 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).

⁹⁹ Maria Glod, *Score One for McLean High Students; Administration Amends Anti-Cheating Policy After Protests*, WASH. POST, Oct. 4, 2006, at B1.

¹⁰⁰ Glod, *McLean Students Sue Anti-Cheating Service*, *supra* note 96.

¹⁰¹ *Id.*; *A.V. v. iParadigms*, 544 F. Supp. 2d 473 (E.D. Va. 2008).

¹⁰² Glod, *McLean Students Sue Anti-Cheating Service*, *supra* note 96.

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.¹⁰³

Plagiarism detection services rely on access to entire copies of student term papers and any works from which the student might have illicitly copied.¹⁰⁴ Yet the services do not necessarily cause any of the copyrighted content they process to be displayed to or read by end users.¹⁰⁵ As such, antiplagiarism software also presents the paradox of nonexpressive copying: the service copies copyrighted works in their entirety in order to compute a result, but the result itself contains none of the copyrighted expression of the original works.

II. THE DOCTRINAL IMPLICATIONS OF NONEXPRESSIVE USE

A. *The Principle of Nonexpressive Use*

Copyright protects only works that contain original creative expression. As such, copyrighted works are typically capable of enjoyment, appreciation, or at least comprehension by human actors. The enjoyment of watching a film, listening to music, or reading a book is derived from the creative expression contained within those objects. We, as viewers, listeners, and readers choose some films, songs, and books over others because of the quality of their expression. It is convenient to think of these experiential uses as “expressive” in that they relate to, and are motivated by, the expression embedded within a copyrighted work.¹⁰⁶ This observation, that expressive works are usually copied in contemplation of experiential expressive uses, extends to partial copies as well. Because meaning is derived from context,¹⁰⁷ sampling a segment of music might change what that music

¹⁰³ The fact that apparently the students themselves loaded their papers into the database after accepting a click-wrap agreement weakened the students’ copyright claim. See *A.V.*, 544 F. Supp. 2d 473, 480 (finding that the parties entered into a valid contractual agreement when plaintiffs clicked “I Agree” to acknowledge their acceptance of the terms of the defendant’s user agreement).

¹⁰⁴ The same issue arises in relation to automated copyright detection technology used for digital images. See U.S. Patent No. 7,120,274 (filed Dec. 31, 2002) (“Automated copyright detection in digital images.”).

¹⁰⁵ See *infra* notes 187–188 and accompanying text.

¹⁰⁶ To the extent that this definition of the “expressive use” of a copyrighted work departs from a conventional understanding, the reader should understand that this Article employs it as a term of art.

¹⁰⁷ See RICHARD BANDLER & 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.”); 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).

expresses, but the end product is expressive in the general sense nonetheless.¹⁰⁸

The distinction between expressive and nonexpressive *works* is already well recognized in copyright law as the gatekeeper to copyright protection—novels are protected by copyright, telephone books and other uncreative compilations of data are not.¹⁰⁹ This Article explores the importance of the same distinction in relation to potential *acts* of infringement. In brief, the argument is that nonexpressive uses of copyrighted works—i.e., acts of copying that do not communicate the author’s original expression to the public—should not generally be regarded as infringing.

The legal status of actual copying for nonexpressive uses was not a burning issue before digital technology: there simply was no commercially relevant total literal copying directed towards a nonexpressive end. To illustrate through absurdum, it would be both uncommon and nonsensical to photocopy *Gone With The Wind* and then to use it to light a fire. This is technically a nonexpressive use, but not one that factors heavily in any serious policy discussion. However, digital technology and the increasing value of metadata have combined to make the legality of nonexpressive copying arguably the most significant issue in copyright law today.

In a world of analog works, nonexpressive 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 involve copying the work in question, copyright owners have no legal right to object. Similarly, 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, incendiary uses of *Gone With The Wind* notwithstanding. 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 nonexpressive use of a copyright work that involved physically copying the entire work. However, given the significant role of nonexpressive copying in Internet search engines and other copy-reliant technologies, the legality of nonexpressive copying is an issue that copyright doctrine must now address.

¹⁰⁸ See Andrew Ross, *Princes Among Thieves: Sampling in the 80s*, ARTFORUM INT’L, Mar. 2003, at 249 (discussing the social meaning of sampling in American hip-hop music of the 1980s); see also David Hesmondhalgh, *Digital Sampling and Social Inequality*, 15 SOC. & LEGAL STUD. 53 (2006) (summarizing the literature and addressing the social and legal issues of music sampling as cultural “borrowing”); Thomas G. Schumacher, “*This Is a Sampling Sport*”: *Digital Sampling, Rap Music and the Law in Cultural Production*, 17(2) MEDIA, CULTURE AND SOC’Y 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”). The status of computer programs under this expressive–nonexpressive paradigm is considered below. See *infra* Part II.A.5.

¹⁰⁹ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).

The purpose of this section is to demonstrate three related propositions: one descriptive, one normative, and one prescriptive. The descriptive proposition is that for the vast majority of works, the copyright owner's exclusive rights are implicitly defined and limited in reference to expressive communication to the public.¹¹⁰ The normative proposition follows from the descriptive: acts of copying, which by their very nature cannot communicate the author's original expression to the public, should not generally be held to constitute copyright infringement.¹¹¹ The prescriptive proposition addresses the implementation of the broader normative claim and argues that the best way to apply the general principle of nonexpressive use is via the fair use doctrine. Specifically, the fair use analysis should incorporate the extent of the nonexpressive nature of the defendant's use. For the reasons detailed below, nonexpressive use is not precisely a free-standing defense to copyright infringement.¹¹² Instead, it is a consideration that permeates and informs the application of the fair use doctrine.¹¹³ These propositions are consistent with the goals of copyright generally and existing copyright doctrine.

The express purpose of the Copyright Clause in the U.S. Constitution is "to promote the Progress of Science and useful Arts."¹¹⁴ Copyright exists to encourage the creativity of authors and to promote the creation and dissemination of information.¹¹⁵ 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."¹¹⁶ Where the law strikes that balance 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.¹¹⁷ 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 ad-

¹¹⁰ Abraham Drassinower makes a similar point in his characterization of copyright as "an exclusive right of public presentation." Abraham Drassinower, *Authorship as Public Address: On the Specificity of Copyright Vis-à-vis Patent and Trade-Mark*, 2008 MICH. ST. L. REV. 199, 204.

¹¹¹ It is important to note that for these purposes, "the public" includes individual consumers and thus nonexpressive use is not a synonym for personal use.

¹¹² See *infra* Part II.B.

¹¹³ See *infra* Part II.C.

¹¹⁴ U.S. CONST. art. I, § 8, cl. 8.

¹¹⁵ *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

¹¹⁶ *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985) (quoting *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984)).

¹¹⁷ 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).

ditive contributions.¹¹⁸ 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 breathe.

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. The natural rights argument for copyright is primarily an extension of the Lockean framework of labor as the basis of property ownership of intangibles.¹¹⁹ However, this justification for property does little by itself to establish either its form or its limitations.¹²⁰ "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 person."¹²¹ In either case, the guiding principle of copyright is that one should generally not 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.¹²² These rights are defined, articulated, and limited by a number

¹¹⁸ *E.g.*, *Metro-Goldwyn-Mayer Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005); *Sony*, 464 U.S. at 479 ("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.").

¹¹⁹ See 2 JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* 27 (Peter Laslett ed., 1988). As Locke famously argued:

Every 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 [sic] to it something that is his own, and thereby makes it his Property.

Id. But note that contrary to a strict Lockean approach, copyright law in the United States requires a minimal threshold of creativity in addition to mere "sweat of the brow." For a recent application of Locke to intangibles, see Andrew S. Gold, *A Property Theory of Contract*, 103 NW. U. L. REV. 1 (2009).

¹²⁰ For two quite different views of Locke's implications for intellectual property, compare 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), which argues 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, with Justin Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L.J. 287 (1988), which argues that Locke's labor theory can be used to justify intellectual property without many of the problems that attend its application to physical property.

¹²¹ Hughes, *supra* note 120, at 330; see also Margaret Jane Radin, *Property and Personhood*, 34 STAN. L. REV. 957, 957 (1982) ("[T]o achieve proper self-development—to be a *person*—an individual needs some control over resources in the external environment.").

¹²² 17 U.S.C. § 106(1)–(6) (2006).

of initially judge-made doctrines, such as the idea–expression distinction, the threshold of substantial similarity, and the fair use doctrine.¹²³ As this section explores in more detail below, these doctrines typically limit copyright protection to the expressive aspects of original works of authorship in a way that confirms the place of public communication at the heart of copyright. This point is important because once it is understood that copyright’s primary function is to protect the author from the threat of expressive substitution, the case in favor of nonexpressive uses becomes almost self-evident. Standing alone, a nonexpressive use carries no threat of expressive substitution; such uses should thus fall outside the scope of an author’s entitlement.¹²⁴

Copyright’s focus on acts that have the potential to communicate the author’s original expression is immediately apparent in the exclusion of facts and ideas from protection by copyright.¹²⁵ 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. 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.¹²⁶ However, it is not immediately apparent that communication to the public is a particularly useful lens through which to view the protection of computer software under copyright law. These topics, including the exceptional status of computer software, are now explored in detail.

I. The Exclusion of Nonexpressive 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 the work describes, explains, or illustrates them.¹²⁷ This principle, often simply abbreviated to the “idea–expression

¹²³ The Copyright Act of 1976 also reflects the idea–expression distinction and the fair use doctrine. See 17 U.S.C. §§ 102(b), 107. But these doctrines remain essentially common law features of the copyright system.

¹²⁴ Note as an analogy the case of *Smith v. United States*, in which Justice Scalia concludes that “[t]o use an instrumentality ordinarily means to use it for its intended purpose.” 508 U.S. 223, 227 (1993) (Scalia, J., dissenting) (holding that an individual who traded his gun for drugs had not “used” a firearm for the purposes of the enhanced criminal sentencing provisions of 18 U.S.C. § 924(c)(1)). See also Andrew S. Gold, *Absurd Results, Scrivener’s Errors, and Statutory Interpretation*, 75 U. CIN. L. REV. 25, 43 (2006).

¹²⁵ See *infra* Part II.A.1.

¹²⁶ See *infra* Parts II.A.2–4.

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

distinction,” is longstanding at common law and was expressly incorporated into the 1976 revision of the Copyright Act.¹²⁸

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.”¹²⁹ 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 author’s expression of those concepts.¹³⁰ In *Selden*, for example, the plaintiff had developed a novel and useful method of bookkeeping, the practice of which created value regardless of how and from what source a bookkeeper learned the method.¹³¹ Nonetheless, the plaintiff’s copyright in his instructional material was limited to the expression of his useful methods and did not encompass the methods themselves.¹³² 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 the expression of facts, providing no protection for the former and only limited protection for the latter.¹³³ In *Feist Publications, Inc. v. Rural Telephone Service Co., Inc.*, the Supreme Court ruled that copying listings from a telephone directory did not infringe the copyright in that directory because the information itself was not copyrightable.¹³⁴ As the Court explained, facts—whether they are telephone numbers and addresses or the details of historical occurrences—are not “original” to the author.¹³⁵ The author’s copyright,

¹²⁸ 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.”

¹²⁹ *Baker v. Selden*, 101 U.S. 99, 102 (1879). “Art” and “illustrate” are not meant in the aesthetic sense in this context.

¹³⁰ *Id.* Note also that the copyright protection available for maps is somewhat thin as a result. See 1 NIMMER ON COPYRIGHT § 2.08[A] and the cases cited therein.

¹³¹ *Id.* at 99–100. *Selden*’s system may well have been patentable under today’s standards. See *State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 1373 (Fed. Cir. 1998) (holding that a patent on a data processing system is valid). But see *Lab. Corp. of Am. Holdings v. Metaboli Labs., Inc.*, 548 U.S. 124, 136 (2006) (per curiam) (Breyer, J., dissenting) (noting that the Supreme Court has never endorsed the Federal Circuit’s “useful, concrete, and tangible result” test for patentable processes). See generally Pamela Samuelson, *Why Copyright Law Excludes Systems and Processes from the Scope of Its Protection*, 85 TEX. L. REV. 1921, 1924 (2007) (arguing that thin copyright protection for computer programs is especially appropriate given the availability of patent protection for program innovations).

¹³² *Baker*, 101 U.S. at 103-04.

¹³³ See *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340, 349-50 (1991) (holding that facts are not copyrightable and that the copyright in a factual compilation is thin).

¹³⁴ *Id.* at 362-63.

¹³⁵ *Id.* at 348 (“[C]opyright protection may extend only to those components of a work that are original to the author.”).

therefore, did not cover the facts themselves.¹³⁶ The *Feist* Court further held that the expression of those facts was not protectable, because the selection and alphabetical arrangement of those facts in the telephone directory was “so mechanical or routine as to require no creativity whatsoever.”¹³⁷

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, reusing, or reinterpreting 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 preventing the exploitation of their writings and society’s competing interest in the free flow of ideas, information, and commerce.¹³⁸

The exclusion of facts and ideas from the ambit of copyright protection applies with equal force to nonexpressive 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 a 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 newspaper article, or ideas and processes from an instructional text, without infringing the author’s copyright.¹³⁹ Whether Congress should, or even could, alter the traditional contours of copyright by extending its protection to facts and ideas merits debate, but there is no doubt that copyright law currently offers no such protection.¹⁴⁰

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 holding that the copying of facts and ideas alone does

¹³⁶ *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 556 (1985) (“No author may copyright his ideas or the facts he narrates.”).

¹³⁷ *See Feist*, 499 U.S. at 362 (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 & Co., Inc., v. West Publ’g Co.*, 158 F.3d 674 (2d Cir. 1998) (holding that West’s factual enhancements to judicial opinions could be reasonably viewed as obvious, typical, and lacking even minimal creativity).

¹³⁸ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984); *see also* *Warner Bros., Inc. v. Am. Broad. Cos., Inc.*, 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”).

¹³⁹ *See Feist*, 499 U.S. 340; *Baker v. Selden*, 101 U.S. 99 (1879).

¹⁴⁰ *See generally* Jonathan Band & Makoto Kono, *The Database Protection Debate in the 106th Congress*, 62 OHIO ST. L.J. 869 (2001) (discussing legislative proposals to provide database protection).

not constitute infringement. Preserving the functional force of the idea-expression distinction in the digital context requires a slightly different application: copying for purely nonexpressive 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. The Act, however, 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.¹⁴¹ 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 *New York Times v. Tasini*.¹⁴² In that case, six freelance authors sued a group of publishers, including the Times Company, for placing articles written by the authors into third party 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 authors' contributions as part of a revision to a collective work.¹⁴³

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 communication to the public in determining the rights of the copyright owner. In *Tasini*, the Supreme Court rejected the Times Company's 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 presented," they did not qualify as part of a revision to the original collective work.¹⁴⁴

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 Section 201(c) of the Act.¹⁴⁵ From the Times Company's perspective, electronic storage was no different from the conversion of newsprint to microfilm.¹⁴⁶ As with the electronic database, microfilm required archiving the entire issue to facilitate later retrieval of specific articles. The Court, however, held that what mattered was not how the articles were stored, but how they were retrieved and displayed to us-

¹⁴¹ 17 U.S.C. § 201(c) (2006).

¹⁴² *N.Y. Times Co. v. Tasini*, 533 U.S. 483 (2001).

¹⁴³ *Id.* at 491-92.

¹⁴⁴ *Id.* at 487.

¹⁴⁵ *Id.* at 499.

¹⁴⁶ *Id.* at 501.

ers.¹⁴⁷ Unlike microfilm files, the database presented the individual articles to the user devoid of their initial context.¹⁴⁸ The Court's view was that only user perception mattered and that whether the articles were stored in their initial sequence was irrelevant to both readers and authors alike.¹⁴⁹ 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."¹⁵⁰

Although *Tasini* is not an instance of nonexpressive use, it nonetheless supports the proposition that acts of copying that do not communicate the author's original expression to the public do not constitute copyright infringement. By defining the scope of the publishers' collective works privilege in terms of what is communicated to the public and dismissing the relevance of unseen uses within the defendants' databases, the Court reinforced that expressive communication to the public is the touchstone of copyright infringement.

3. *Substantial Similarity*.—The centrality of expressive communication to the public is inherent in the tests applied by the courts to determine the threshold of infringement—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.¹⁵¹

The copyright owner's exclusive right to "reproduce the copyrighted work in copies" extends to both exact and inexact reproductions.¹⁵² 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 "nonliteral infringement,"—where the accused work is not an exact copy of the copyright owner's work—courts assess whether

¹⁴⁷ *Id.* at 503–04 ("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).

¹⁴⁸ *Id.* at 501.

¹⁴⁹ *Id.* at 501 n.9.

¹⁵⁰ *Id.* at 499 (internal quotation marks and citations omitted).

¹⁵¹ "Substantial similarity" in this sense should not be confused with what is more properly termed "probative similarity," which is the circumstantial evidence of copying founded upon the unlikely similarity between the accused work and the purported original. See *Stillman v. Leo Burnett Co., Inc.*, 720 F. Supp. 1353, 1358 (N.D. Ill. 1989).

¹⁵² 17 U.S.C. § 106(1); *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2d Cir. 1930) ("[T]he question is whether the part so taken is substantial.") (internal quotation marks omitted).

the allegedly infringing work possesses a “substantial similarity” to the copyrighted work.¹⁵³

Courts define the threshold of substantial similarity in reference to the perspective of the ordinary observer.¹⁵⁴ 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.”¹⁵⁵ 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.¹⁵⁶ For example, in *Newton v. Diamond*, the plaintiff alleged that the Beastie Boys had infringed his copyright in a musical composition by including a six second sample of a sound recording in their own musical creation, “Pass the Mic.”¹⁵⁷ 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 their limited use and the sparse nature of Newton’s original composition.¹⁵⁸

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.¹⁵⁹ In cases of fragmented literal similarity, courts determine whether the copying amounts to infringement “by considering the qualitative and quantitative significance of

¹⁵³ See *Tufenkian Import/Export Ventures, Inc. v. Einstein Moomjy, Inc.*, 338 F.3d 127, 131 (2d Cir. 2003); *Laureyssens v. Idea Group, Inc.*, 964 F.2d 131, 140 (2d Cir. 1992).

¹⁵⁴ See *Shine v. Childs*, 382 F. Supp. 2d 602, 614 (S.D.N.Y. 2005) (summarizing authorities).

¹⁵⁵ *Arnstein v. Porter*, 154 F.2d 464, 473 (2d Cir. 1946) (footnotes omitted); see also *Warner Bros., Inc. v. Am. Broad. Cos., Inc.*, 720 F.2d 231, 240 (2d Cir. 1983).

¹⁵⁶ 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 NIMMER ON COPYRIGHT § 13.03[A][2][b]; *Bridgeport Music, Inc. v. Dimension Films*, 410 F.3d 792 (6th Cir. 2005).

¹⁵⁷ *Newton*, 388 F.3d at 1191.

¹⁵⁸ 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.* Note that sound recordings and their underlying compositions are separate and distinct copyrighted works. See 17 U.S.C. § 102(a)(2), (7).

¹⁵⁹ 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.”).

the copied portion in relation to the plaintiff's work as a whole."¹⁶⁰ 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."¹⁶¹

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.¹⁶² 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 communication of protected expression to the public defines and limits the copyright owner's exclusive rights. It also demonstrates that acts of copying that do not communicate the author's original expression to the public do not generally constitute copyright infringement.

4. *Allegations of Intermediate Copying in Hollywood.*—Meritless claims of copyright infringement are a recognized cost of doing business in Hollywood.¹⁶³ Some of these claims are merely opportunistic, while others are 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 latter phenomenon.¹⁶⁴ In that case, the author of a one-page poem about a land of monsters who are afraid of human children

¹⁶⁰ *Newton*, 388 F.3d at 1195 (citing *Worth v. Selchow & Righter Co.*, 827 F.2d 569, 570 n.1 (9th Cir. 1987)); see also *Jarvis v. A&M Records*, 827 F. Supp. 282, 289–90 (D.N.J. 1993); 4 NIMMER ON COPYRIGHT § 13.03[A][2][a], at 13-57–58.

¹⁶¹ *Newton*, 388 F.3d at 1195.

¹⁶² *Id.* at 1193 ("The principle that trivial copying does not constitute actionable infringement has long been a part of copyright law."); *id.* at 1195 ("[T]he dispositive question is whether the copying goes to trivial or substantial elements.")

¹⁶³ 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 for defending against a claim that his screenplay for the movie *Twister* had stolen from another screenplay, Michael Crichton said: "I hope it will usher in a new era where studios fight these frivolous charges and don't treat it as a cost of doing business").

¹⁶⁴ *Madrid v. Chronicle Books*, 209 F. Supp. 2d 1227 (D. Wyo. 2002).

alleged that the Pixar film *Monsters, Inc.* infringed her copyright.¹⁶⁵ The court held, however, that the inverted plot of monsters afraid of children was generic.¹⁶⁶ Other cases involve works with similar themes,¹⁶⁷ similar descriptive titles applied to the same general subject,¹⁶⁸ or similarities discernable only to the plaintiffs themselves.¹⁶⁹

Confronted with motions for summary judgment, plaintiffs often urge the courts to allow them to scrutinize every single draft of the defendant's screenplay, in the hope that some earlier version of the work will disclose a greater resemblance to their own copyrighted work than the finished film does.¹⁷⁰ Courts invariably deny these requests.¹⁷¹ The reasons behind the denials provide an important insight into the structure of copyright law.

Courts refuse to entertain discovery with respect to early drafts of a noninfringing final work precisely because infringement requires at least

¹⁶⁵ *Id.* at 1234.

¹⁶⁶ *Id.* Two antecedents spring immediately to mind. First, E.T. hiding in the cupboard from Elliot. E.T.: THE EXTRA-TERRESTRIAL (Universal Pictures 1982). Second, Max's dominion over the fearful monsters in *Where The Wild Things Are*. MAURICE SENDAK, *WHERE THE WILD THINGS ARE* (1963).

¹⁶⁷ For example, in *Litchfield v. Spielberg* the author of a musical play about two aliens stranded at the North Pole unsuccessfully accused the producers of the motion picture *E.T.: The Extra-Terrestrial* of infringement. See 736 F.2d 1352, 1358 (9th Cir. 1984) (finding no substantial similarity between the sequences of events, mood, dialogue, and characters of the two works); see also *Warner Bros., Inc. v. Am. Broad. Cos., Inc.*, 720 F.2d 231, 235, 243 (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; as the court observed, "[i]n the genre of superheroes, Hinkley follows Superman as, in the genre of detectives, Inspector Clouseau follows Sherlock Holmes").

¹⁶⁸ In *Davis v. United Artists, Inc.*, the author of the 1972 Vietnam novel entitled "Coming Home" failed to demonstrate copyright infringement in relation to a 1978 Vietnam film also titled "Coming Home." 547 F. Supp. 722, 725 (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* unsuccessfully alleged that the Time Life film *Fort Apache, The Bronx* amounted to copyright infringement. 615 F. Supp. 430, 435 (S.D.N.Y. 1985) (holding that no reasonable observer could find substantial similarity and that "[a]ny similarity that may exist is either trivial, abstract or non-protectible as a matter of law").

¹⁶⁹ For example, in *Stromback v. New Line Cinema*, the author of a screenplay outline about a callous reporter who brings down a corrupt governor accused the writers of the film *Little Nicky* of copyright infringement; *Little Nicky* is a comedy about the Devil and his three sons, one of whom, played by Adam Sandler, has a speech impediment. See 384 F.3d 283, 297 (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 of the screenplay *Amoral Dilemma*, the rather grim story of a disaffected young Manhattan insurance attorney who knowingly corresponds with a death row prisoner. See 388 F. Supp. 2d 274, 279–81 (S.D.N.Y. 2005).

¹⁷⁰ See, e.g., *Stromback*, 384 F.3d 283; *Flaherty v. Filardi*, 2007 U.S. Dist. LEXIS 69202, at *8–9 (S.D.N.Y. Sept. 19, 2007) (copyright claim to interim drafts of a published non-infringing final work dismissed as a matter of law); *Walker*, 615 F. Supp. at 434 (request to discover drafts denied).

¹⁷¹ See *Walker*, 615 F. Supp. at 435 (noting that courts routinely reject requests to consider earlier drafts of screenplays).

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."¹⁷² 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."¹⁷³

The refusal of courts to entertain copyright infringement allegations in relation to unpublished drafts and preliminary scripts demonstrates the practical importance of a focus on expressive substitution. Because the copyright owner's rights are generally limited to the communication of their original expression to the public, even if it were not in the public domain, a filmmaker would be perfectly entitled to start with Jane Austen's *Emma* and rework the plot over and over again until she comes out with *Clueless*.¹⁷⁴ 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.

5. *Computer Software: An Exception?*—Copyright protection for computer software has long been a source of controversy and disquiet.¹⁷⁵ Although the broad language of the statutory definition of "literary works" in the Copyright Act includes computer programs,¹⁷⁶ treating software as a work of literature presents something of a contradiction. The 1976 Copy-

¹⁷² *Davis*, 547 F. Supp. at 724 n.9 (citing *Fuld v. Nat'l Broad. Co., Inc.*, 390 F. Supp. 877, 882 n.4 (S.D.N.Y. 1975)); see also *Stromback*, 384 F.3d at 299 ("In deciding infringement claims, courts have held that only the version of the alleged infringing work presented to the public should be considered."); *Madrid v. Chronicle Books*, 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.") (internal quotation marks omitted); *Walker*, 615 F. Supp. at 434 ("The Court considers the works as they were presented to the public.")

¹⁷³ *Warner Bros., Inc. v. Am. Broad. Cos., Inc.*, 720 F.2d 231, 241 (2d Cir. 1983) (citing 3 NIMMER ON COPYRIGHT § 13.03[B] at 13-38.1 to -38.2; *Eden Toys, Inc. v. Marshall Field & Co.*, 675 F.2d 498, 501 (2d Cir. 1982); *Durham Indus., Inc. v. Tomy Corp.*, 630 F.2d 904, 913 & n.11 (2d Cir. 1980)).

¹⁷⁴ *CLUELESS* (Paramount 1995).

¹⁷⁵ For a range of opinions, see Jane C. Ginsburg, *Four Reasons and a Paradox: The Manifest Superiority of Copyright over Sui Generis Protection of Computer Software*, 94 COLUM. L. REV. 2559 (1994); Peter S. Menell, *An Analysis of the Scope of Copyright Protection for Application Programs*, 41 STAN. L. REV. 1045 (1989); and Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994).

¹⁷⁶ 17 U.S.C. § 101 (2006) ("literary works" includes works "expressed in words, numbers, or other verbal or numerical symbols or indicia"); see also NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS, FINAL REPORT OF THE NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS 16 (1978) (concluding that "it was clearly the intent of Congress to include computer programs within the scope of copyrightable subject matter in the Act of 1976"); but see Pamela Samuelson, *CONTU Revisited: The Case Against Copyright Protection for Computer Programs in Machine-Readable Form*, 1984 DUKE L.J. 663.

right Act clearly states that copyright protection does not extend to any “process, system, [or] method of operation”¹⁷⁷ And yet, thanks to an amendment made in 1980, the Act clearly extends copyright protection to computer programs, which are defined as “a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result.”¹⁷⁸ A “set of instructions” used “in order to bring about a certain result” appears to be the very essence of the “process, system, method of operation” exclusion under Section 102(b).

With this contradiction in mind, it is hardly surprising that the general theory of copyright advanced in this Article—the centrality of expressive substitution—does not fit perfectly to software.¹⁷⁹ Users do not typically copy copyrighted computer programs so that they can imbibe the artistry of the programmer’s expression. In most cases no human being ever directly observes that expression. Instead, the copyrighted software sends a series of commands to a computer (usually via a software platform and an operating system) which brings about a certain result such as launching a word processing program or a video game. Aspects of the word processor or the video game may themselves be expressive, but it is something of a stretch to argue that the underlying computer program is itself used expressively. To be clear, in terms of copyrighted subject matter, a computer program should be considered to be expressive. However, the ordinary *use* of a computer program is more functional than expressive. The distinction between expressive and nonexpressive uses is not intended to eviscerate copyright protection for computer software.¹⁸⁰ As the preceding discussion makes clear, the rational justification for copyright is generally that it protects the author against expressive substitution—the anomalous nature of computer software points to a different basis for attaching copyright protection and thus does not admit a defense of nonexpressive use to the same extent. In sum, computer software should continue to be treated as exceptional—nonexpressive use should not be regarded as a defense to ordinary acts of software piracy.¹⁸¹

¹⁷⁷ 17 U.S.C. § 102(b). Exclusive rights in processes and methods of operation are generally left to the patent system. See 35 U.S.C. § 101 (2006).

¹⁷⁸ Computer Software Copyright Act of 1980, Act of Dec. 12, 1980, Pub. L. 96-517, sec. 10, 94 Stat. 3028; 17 U.S.C. § 101, as amended by Pub. L. 96-517, Sec. 10(a).

¹⁷⁹ The same objections could be raised with respect to the copyright protection of architectural plans, and the following discussion applies *mutatis mutandis* to that subject matter. The Berne Convention Implementation Act (1988) and the Architectural Works Copyright Protection Act (1990) recognize two separate forms of protection for architectural works, one for architectural plans and the other for structures based on such plans. For an overview, see 1 NIMMER ON COPYRIGHT § 2.08.

¹⁸⁰ As will be explained in more detail below, a use that effectively undermines the rationale of copyright protection for a particular class of work can hardly be regarded as fair. See *infra* Part II.C.

¹⁸¹ However, as noted below, the nonexpressive use analysis still provides a useful framework for understanding software reverse engineering cases. See *infra* note 254 and accompanying text.

The difference between copyright in software and in “normal” literary works is apparent from the case law relating to intermediate copying. As discussed in the previous subsection, in the context of the motion picture industry, courts have refused to entertain the notion that intermediate scripts that never see the light of day could result in copyright infringement. In contrast, in software reverse engineering cases, courts take the allegation of infringement via intermediate copying seriously as a potential basis for infringement.¹⁸² Nonetheless, in the case of computer software, the intermediate copying required for reverse engineering has invariably been found to constitute fair use.

We should not be particularly troubled that statutory accretions to the copyright system such as computer software and architectural plans do not conform to the general theory outlined in this Article. While the introduction of these non-native species has certainly altered copyright’s landscape, their mere existence should not prevent us from articulating a coherent theory with respect to subject matter indigenous to copyright. The general theory that expressive substitution is fundamental to copyright infringement necessarily makes an exception for works that consist of the “process,” “system,” or “method of operation” designed to bring about a certain result. This is the difference between a general theory and a universal one.

* * *

This section has explored the centrality of expressive substitution to a variety of doctrines and applications: the idea–expression distinction, substantial similarity, the collective work right, and finally, the refusal of courts to entertain infringement actions solely based on unpublished intermediate drafts of literary works. With the exception of anomalous subject matter such as software, copyright law appears to embrace a general concept of expressive substitution. To the extent that communication of original expression to the public is the touchstone of copyright infringement, it follows that copyright liability should not ordinarily be found in circumstances where the use in question is incapable of giving rise to any expressive communication. Once the pivotal nature of expressive substitution to copyright infringement is properly understood, the implications for copy-reliant technologies crystallize: the nonexpressive use of a copyrighted work should not ordinarily result in copyright infringement. To be clear, nonexpressive uses are merely a subset of uses that do not create any risk of expressive substitution; the advance of digital technology has made that subset increasingly important, however.

¹⁸² See, e.g., *Sony Computer Entm’t, Inc. v. Connectix Corp.*, 203 F.3d 596, 599 (9th Cir. 2000); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1518–19 (9th Cir. 1992) (holding that files and printouts created during the reverse engineering process meet the requirements for being a “copy” and “therefore fall[] squarely within the category of acts that are prohibited”).

B. Doctrinal Incorporation of Nonexpressive Use

As discussed above, the general principle of nonexpressive 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 generally include copying that offers no possibility of expressive substitution because it is entirely nonexpressive in nature. As the doctrinal survey in section A demonstrates, authors possess a set of limited and largely economic rights to control the expressive uses of their works. Except in already anomalous cases such as computer software, extending those rights to encompass non-expressive uses would constitute a significant departure from existing copyright principles. The preceding discussion has established the general principle of nonexpressive use; we now turn to the prescriptive implications of that principle. Specifically, this section demonstrates *why* the principle of nonexpressive use should be applied in the context of copyright’s fair use doctrine and not as a freestanding defense to copyright infringement. The next section then explains *how* nonexpressive use fits within a traditional fair use analysis.

How does the principle of nonexpressive use relate to copy-reliant technology? As set forth in more detail below, the copying at issue in the case studies discussed in Part I was manifestly different from the usual copyright infringement scenario. 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 metadata about the works. Explicit recognition of this principle of nonexpressive use would significantly clarify the legal status of copy-reliant technologies such as Internet search engines, plagiarism detection software, and the Google Book project. However, acknowledging the principle of nonexpressive use raises the subsidiary question of how to implement the principle in the law.

Although a court could rule that the use of a copyrighted work that does not communicate its expressive content is *per se* not a violation of the copyright owner’s rights under Section 106 of the Copyright Act, this Article does not recommend that course for three reasons. First, the statutory text of Section 106 does not directly support the concept of nonexpressive use. For example, Section 106(1) gives copyright owners the exclusive rights “to reproduce the copyrighted work in copies.”¹⁸³ The Act defines “[c]opies” as “material objects . . . in which a work is fixed . . . and from which the work *can* be perceived, reproduced, or otherwise communi-

¹⁸³ 17 U.S.C. § 106(1) (2006).

cated.”¹⁸⁴ Thus, the requirement for a copy is only that the work is *capable* of being perceived, not that anyone actually perceives it and uses it expressively. The second reason not to interpret Section 106 as providing a blanket exclusion for nonexpressive use is that, as noted above, the principle may not apply to some of copyright’s more irregular subject matter, such as computer software and architectural plans. The exceptional status of computer software in particular suggests that the principle requires a more context-specific implementation. The third reason not to adopt a per se rule with respect to nonexpressive use is that in many contexts the concept is ambiguous. Where the validity of a defendant’s claim that a particular use is nonexpressive is contestable, courts may find that adopting a categorical rule that nonexpressive uses are noninfringing simply shifts the focus of argument from substantive questions to questions of category definition.

Although the principle of nonexpressive 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 nonexpressive use, in that it does not relate to the expression of these facts, but to the facts themselves.¹⁸⁵ Similarly, generating factual information about a work should also be categorized as a nonexpressive 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 because this information about the work is independent of the expressive value of the work.¹⁸⁶

Similarly, the nonexpressive use of some copy-reliant technologies is fairly clear-cut. Plagiarism detection software illustrates one of the less ambiguous applications of the nonexpressive use principle. These services rely on access to entire copies of student term papers and any works from which a student might have copied them; yet the services do not necessarily display any of the copyrighted content they process to the end users. Plagiarism detection software works by comparing strings of text in new works to strings of text in existing works.¹⁸⁷ If the software finds a match, it indi-

¹⁸⁴ 17 U.S.C. § 101 (emphasis added).

¹⁸⁵ 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.”).

¹⁸⁶ *Moby Dick* is in the public domain in the United States and is available at Project Gutenberg at <http://www.gutenberg.org/etext/2701>.

¹⁸⁷ 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, at A1. 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’afar, *The JK System to Detect Plagiarism*, 6(2) J. COMPUTER SCI. & TECH. 66 (2006). The Turnitin software uses statistical techniques originally designed to analyze brain

cates 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.¹⁸⁸ Of course, most plagiarism software is also programmed to display the source file from which the work being scrutinized was allegedly copied. This optional feature may indeed be regarded as expressive, although not necessarily without fair use protection. Nonetheless, many other significant copy-reliant technologies present some degree of ambiguity as to whether they should be regarded as expressive or nonexpressive.

In both *Field* and *Perfect 10*, the courts effectively found that the primary purpose of the copying at issue was nonexpressive. In *Field*, the court held that although allowing users to retrieve web pages from the search engine cache also allowed them to be read, the primary use of the cache was nonexpressive and thus noninfringing.¹⁸⁹ 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.¹⁹⁰ The court relied 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.¹⁹¹ Although the court did not employ the terminology set forth in this Article, its reasoning illustrates that the relevant distinction was that Google's use was nonexpressive. 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 ramifications.¹⁹² As the court noted: “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.”¹⁹³ In addition, the court also noted that the availability of cached links enables users to understand why the search engine indicated that a seemingly irrelevant web page was responsive to their query.¹⁹⁴

waves to compare the fingerprints of student papers to more than a billion documents that have been fingerprinted in a similar fashion. See *Plagiarise. Let No One Else's Work Evade Your Eyes*, THE ECONOMIST, Mar. 14, 2002 (U.S. Edition).

¹⁸⁸ 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.

¹⁸⁹ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1115 (D. Nev. 2006).

¹⁹⁰ *Id.* at 1118.

¹⁹¹ *Id.* at 1118–19.

¹⁹² *Id.*

¹⁹³ *Id.* at 1119.

¹⁹⁴ *Id.*

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 not served by the original copyrighted work alone. Likewise, referring to the cache to understand better the relationship between a particular page and a particular search term is also a use not served by the original copyrighted work alone.¹⁹⁵

Perfect 10 presents a similar ambiguity and a similar resolution.¹⁹⁶ In *Perfect 10*, the thumbnail representations were clearly visible to the public. The court of appeals, however, found that the thumbnails did not fulfill a demand for the originals as expressive works.¹⁹⁷ 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.¹⁹⁸ However, the court found that Google's creation of thumbnail representations did not infringe Perfect 10's rights, in large measure because the purpose of the copying the plaintiff complained of was a different use, a nonexpressive use.¹⁹⁹ The court of appeals distinguished the use of thumbnails by Google as pointing devices from the use of pictures as expressive works themselves. In the court's words, Google's replication of the original works as thumbnails "served a different function" unrelated to "artistic expression."²⁰⁰ The court further 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."²⁰¹ 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 possibility that the search engine copying could function as an expressive substitute for the copyright owners' original works. Nonetheless, in both cases the courts found that the copying at issue did not fulfill a demand for the originals as

¹⁹⁵ Not all assertions of nonexpressive use deserve equal credence. Reproducing an entire film, with additional commentary, to explain its nomination for an Academy Award would not likely qualify as a nonexpressive use. In this example, the fact that the amount of the underlying work copied grossly exceeds what was necessary gives the lie to the claim that the use was nonexpressive. See *infra* Part II.C.3.

¹⁹⁶ See *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701 (9th Cir. 2007).

¹⁹⁷ *Id.* at 721.

¹⁹⁸ *Id.* at 719.

¹⁹⁹ *Id.* at 725 (reversing the district court's ruling that the use of thumbnails was not fair use).

²⁰⁰ *Id.* at 721 (citation and quotation omitted); see also *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 819 (9th Cir. 2003).

²⁰¹ *Perfect 10*, 487 F.3d at 721.

expressive works.²⁰² 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 copyrighted works, given that the predominant uses of the cached content were unrelated to the expressive function of the original works.²⁰³ 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.²⁰⁴ In the court's opinion, because the search engine used thumbnail representations to show users which websites contained images relevant to their search terms, they were not substitutes for the originals.²⁰⁵ The potential for expressive and nonexpressive uses to sit side by side in cases such as *Field* and *Perfect 10* highlights the limits of a strictly categorical approach. This in turn suggests the need to integrate the issue of nonexpressive use into a fair use analysis, as explained in section C below.

The Google Book search engine litigation also illustrates the potential ambiguity of nonexpressive use. 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 tomes 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.²⁰⁶ The traditional (and now obsolete) library “card catalog” is an archetypal metadata repository—the card catalog contains information about the author, title, and subject matter, but it does not contain the volumes themselves.²⁰⁷ As the quantity of available information in-

²⁰² *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006); *Perfect 10*, 487 F.3d 701.

²⁰³ *Field*, 412 F. Supp. 2d at 1118–19.

²⁰⁴ *Perfect 10*, 487 F.3d at 721.

²⁰⁵ *Id.* at 724.

²⁰⁶ 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.” AM. LIBRARY ASS'N ALCTS CC:DA TASK FORCE ON METADATA, SUMMARY REPORT (June 1999), <http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>; see also Ganesan Shankaranarayanan & Adir Evan, *The Metadata Enigma*, 49 COMMS. ACM 88 (2006) (arguing that most definitions of metadata ignore its richness and complexity).

²⁰⁷ See BARBARA TILLET, WHAT IS FRBR?: A CONCEPTUAL MODEL FOR THE BIBLIOGRAPHIC UNIVERSE (Library of Congress Cataloging Distribution Service 2004), available at <http://www.loc.gov/cds/downloads/FRBR.pdf>; IFLA STUDY GROUP ON THE FUNCTIONAL

creases, so too 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 along with search results. The construction of the Google Book database involves the actual copying of millions of expressive works for an intermediate purpose that is itself entirely nonexpressive. In this regard, it is exactly analogous to plagiarism detection software.²⁰⁸ However, while the process of generating data 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? Although the search engine displays expressive snippets of books to end users, in most cases those snippets are too fragmented and insubstantial to amount to infringing copies of the books themselves. It is conceivable that a three line snippet of a haiku could be infringing by itself, but for most books, the notion that any three lines could meet the test of substantial similarity is risible.²⁰⁹ However, if Google Book's snippets were less abbreviated, they might be coherent and substantial enough to be infringing. As long as Google Book's final expressive use does not infringe, it should not detract from the nonexpressiveness of an intermediate use.²¹⁰

In sum, although Internet search engines (and possibly even the Google Book project) are strong candidates for nonexpressive use, the extent to which that label actually fits may depend on a detailed assessment of specific facts. For example, the claim of nonexpressive use in relation to an image search engine that reproduced full-scale images as opposed to thumbnails would be doubtful. The categorization of intermediate nonexpressive uses intertwined with infringing expressive uses is also ambiguous. To ameliorate this problem, courts should evaluate claims of nonexpressive use within the context of a fair use analysis rather than adopting a categorical rule. Section C addresses the links between nonexpressive use and fair use in detail.

REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS, FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS, FINAL REPORT (1998), <http://www.ifla.org/files/cataloguing/frbr/frbr.pdf>.

²⁰⁸ 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 4, at 425–28.

²⁰⁹ Note that Google Book treats works such as poetry, dictionaries, drug reference guides, price guides, and books of quotations differently for this very reason. *See Google Book Settlement, available at* <http://www.googlebooksettlement.com/help/bin/answer.py?answer=118722> (last visited Aug. 4, 2009).

²¹⁰ Whether a converse finding that an end product infringes requires treating an intermediate non-expressive use as expressive is a question for another day.

C. Fair Use and Nonexpressive Use

Copyright law does not require a radical reinterpretation to accommodate the principle of nonexpressive use. It merely requires applying the existing elements of fair use doctrine to recognize that acts of copying that do not communicate the author's original expression to the public do not typically 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 potential market for or value of the copyrighted work.²¹¹

In reality, these factors are neither complete²¹² nor individually or cumulatively determinative.²¹³ Even the notion that there are four factors is misleading: beneath the statutory factors lies an amalgamation of interconnected metafactors, subfactors, and presumptions. The implications of nonexpressive use in relation to fair use are explored below.

1. *The "Purpose and Character" of Nonexpressive Uses.*—The non-expressive nature of the defendant's 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."²¹⁴ Recognizing the copyright owner's exclusive rights as implicitly defined and limited in reference to expressive communication to the public makes sense of both expressive and nonexpressive fair uses. Indeed, recognition of this overarching principle may be the key to rescuing the concept of transformative use from elastic imprecision.

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

[W]hether the new use merely supersedes the objects of the original creation . . . or instead adds something new, with a further purpose or different

²¹¹ 17 U.S.C. § 107 (2006).

²¹² See *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 and internal quotation marks omitted); *Universal City Studios, Inc. v. Sony Corp. of Am.*, 480 F. Supp. 429, 448 (C.D. Cal. 1979) ("The factors are illustrative, not definitive.").

²¹³ See *Sag*, *supra* note 4, at 434; see also Michael J. Madison, *A Pattern-Oriented Approach to Fair Use*, 45 WM. & MARY L. REV. 1525, 1564 (2004); David Nimmer, "Fairest of Them All" and *Other Fairy Tales of Fair Use*, 66 LAW & CONTEMP. PROBS. 263, 281 (2003) ("Courts tend first to make a judgment that the ultimate disposition is fair use or unfair use, and then align the four factors to fit that result as best they can.").

²¹⁴ 17 U.S.C. § 107(1).

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.²¹⁵

Traditionally, courts apply the concept of transformative use to new expressive uses that “provide social benefit, by shedding light on an earlier work, and, in the process, creat[e] a new one.”²¹⁶ Transformative use is most obvious when the work is itself transformed; however, in many cases courts have held that the mere recontextualization of a copyrighted work from one expressive context to another is sufficient to sustain a finding of fair use—the work itself need not be altered.²¹⁷ Understanding the rationale for transformative use is the key to grasping the link between transformative use and nonexpressive use. The privileged status of transformative uses under the fair use doctrine allows for the creation of new works from old. This is not a sufficient explanation, however, because other doctrinal levers, such as a narrower understanding of the author’s exclusive right to make derivative works, could achieve the same effect.²¹⁸ Beyond a simple enthusiasm for new works, courts accord special status to transformative uses because they do not substitute for the author’s original expression—they do not merely supersede the objects of the original creation.²¹⁹ Because of this special status, the greater the extent of transformation, the less significant other factors weighing against fair use will become.²²⁰

Cognizant of the Supreme Court’s focus on transformative uses, some courts have simply equated nonexpressive with transformative. 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.”²²¹ This seems to

²¹⁵ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994) (citations and internal quotation marks omitted); see also Pierre N. Leval, Commentary, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990) (“I believe the answer to the question of justification turns primarily on whether, and to what extent, the challenged use is transformative.”).

²¹⁶ *Campbell*, 510 U.S. at 579.

²¹⁷ See, e.g., *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 609–10 (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”); *Mattel, Inc. v. Walking Mountain Prods.*, 353 F.3d 792, 796–98, 800–06 (9th Cir. 2003) (concluding that photos parodying Barbie by depicting “nude Barbie dolls juxtaposed with vintage kitchen appliances” was a fair use).

²¹⁸ See, e.g., 17 U.S.C. § 106(2) (2006).

²¹⁹ See, e.g., *Campbell*, 510 U.S. at 579.

²²⁰ *Id.*

²²¹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 721 (9th Cir. 2007) (holding further 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.” (citing *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818–19 (9th Cir. 2003))).

be stretching the concept of transformation beyond its natural utility. It would be better to recognize 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.

By construction, the more nonexpressive the use of a copyrighted work is, the less it substitutes for the author's original expression.²²² As such, courts should regard primarily nonexpressive uses as equivalent (but not identical) to highly transformative uses—their “purpose and character” is such that they do not merely supersede the objects of the original creation.²²³ In addition, the same logic that dictates that the more transformative a work is, the less significant the other factors become, also applies to non-expressive uses.²²⁴

2. *Nonexpressive 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.”²²⁵ Although the application of the fair use doctrine to 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.²²⁶ 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 nonexpressive, its commercial or noncommercial nature is irrelevant because non-expressive uses do not substitute for the author's original expression.

In both *Sony Corp. of America v. Universal City Studios, Inc.* and *Harper & Row Publishers, Inc. v. Nation Enterprises*, the Supreme Court indicated that commercial uses are disfavored under the fair use doctrine.²²⁷ Writing for the majority in *Sony*, Justice Stevens suggested that if Sony's

²²² The analysis in this section is subject to the caveat regarding computer software and other quasi-functional works discussed in Part II.A.5.

²²³ See *Campbell*, 510 U.S. at 583.

²²⁴ See *id.* at 579.

²²⁵ 17 U.S.C. § 107(1) (2006).

²²⁶ Indeed, the Ninth Circuit's approach to commerciality in *Napster* defines the concept exclusively in terms of market substitution. See *A&M Records, Inc. 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”).

²²⁷ *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449 (1984).

video cassette recorder “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.”²²⁸ 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.”²²⁹

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 v. Ashcroft*, copyright promotes the creation and publication of free expression “[b]y establishing a marketable right to the use of one’s expression.”²³⁰ 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.”²³¹ The virtues of creative production freed from the shackles of patronage and direct government control apply equally to all forms of private production, regardless of 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. Non-commercial uses may have other characteristics, such as a greater degree of spillovers, which justify fair use,²³² but there are no inherent differences between the uses of commercial and noncommercial actors. In a modern free market economy, most copyrighted 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 many 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,

²²⁸ *Sony*, 464 U.S. at 449. “Time-shifting” refers to the consumer practice of recording a program to be viewed at a later, more convenient time. *Id.* at 418.

²²⁹ *Harper & Row*, 471 U.S. at 562.

²³⁰ *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) (citing *Harper & Row*, 471 U.S. at 558).

²³¹ Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288 (1996).

²³² See 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).

teaching, scholarship, and research, since these activities ‘are generally conducted for profit in this country.’”²³³

The Court in *Campbell* rejected the notion that commerciality by itself had any “hard presumptive significance.”²³⁴ 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,” then “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.”²³⁵ 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, but rather “add[] something new, with a further purpose or different character, altering the first with new expression, meaning, or message.”²³⁶ Courts should treat commercial nonexpressive uses similarly: the more nonexpressive a use is, the less it is capable of substituting for the author’s original expression.

There is ready support for this position in the case law. 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.²³⁷ 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.²³⁸ 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.”²³⁹

3. *Nonexpressive Use and “Amount and Substantiality.”*—The degree to which a use is nonexpressive is also significant in terms of the third fair use factor, “the amount and substantiality of the portion used in relation

²³³ 510 U.S. 569, 584 (1994) (citing *Harper & Row*, 471 U.S. at 592 (Brennan, J., dissenting); 3 BOSWELL’S LIFE OF JOHNSON 19 (G. Hill ed. 1934)) (other citations omitted).

²³⁴ *Campbell*, 510 U.S. at 585.

²³⁵ *Id.* at 579 (citations omitted). 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. See Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978–2005*, 156 U. PA. L. REV. 549, 598 (2008).

²³⁶ *Campbell*, 510 U.S. at 579; see Leval *supra* note 215, at 1111 (“If . . . the secondary use adds value to the original—if the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings—this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society.”).

²³⁷ *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 815 (9th Cir. 2003).

²³⁸ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 724 (9th Cir. 2007).

²³⁹ *Id.* (citing *Campbell*, 510 U.S. at 591) (internal quotation marks omitted).

to the copyrighted work as a whole.”²⁴⁰ This 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 simply that a use properly categorized as nonexpressive does not substitute for the any of the expressive value of the author’s original expression.

Even in the realm of expressive uses, there is no linear relationship between the amount 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.²⁴¹ 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.²⁴² 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 defendant’s purpose required complete reproduction.²⁴³

²⁴⁰ 17 U.S.C. § 107(3) (2006). 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 (C.C. 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.* at 345.

²⁴¹ Compare *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1014–15 (9th Cir. 2001) (“Napster users are not fair users.”), with *Blanch v. Koons*, 467 F.3d 244 (2d Cir. 2006) (holding that the visual artist Jeff Koons’s use of fashion photography in collage was fair use).

²⁴² See *Recording Indus. Ass’n of Am. v. Diamond Multimedia Sys.*, 180 F.3d 1072, 1081 (9th Cir. 1999) (holding that personal digital music players are legal); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449–50 (1984) (holding that videotape recorders, primarily used for time shifting, are legal); cf. *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 971 (9th Cir. 1992) (noting that modifying the way in which an entirely copied video game is played is legal in just the same way as video recorders that allow the viewer to view the recorded work in a different sequence are legal).

²⁴³ See *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”); *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); *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”); *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 at issue would have made the picture useless to the story).

Far from being linear or arithmetic in nature, proper application of the third factor is contingent upon the purpose and the 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."²⁴⁴ 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 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.²⁴⁵ Just as the extent of permissible copying varies according to purpose, it also varies according to effect. In *Harper & Row*, the defendant magazine, *The Nation*, copied only a few hundred words from a soon-to-be-published autobiography of former President Gerald Ford. The Supreme Court held, however, 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," had taken "the most interesting and moving parts of the entire manuscript," and had structured its article around these quoted excerpts.²⁴⁶ The Court's finding in relation to the third factor rests on the finding that *The Nation* had taken essentially the heart of the book's expressive value.²⁴⁷

Instead of relying on a mechanical quantification of the *amount* of the original work used, the third factor asks courts to assess how much of the *value* of the original work is present in the allegedly infringing work.²⁴⁸ Accordingly, the extent to which a use is nonexpressive plays a vital role in the assessment of the third fair use factor. A nonexpressive use does not generally substitute for the expressive value of the author's original expression, and therefore courts should view it as qualitatively insignificant under the third factor, even if it involves literal copying of an entire work.

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.²⁴⁹ Consistent with its earlier decision in *Kelly*, the court found that the representation of an entire photographic image was

²⁴⁴ *Campbell*, 510 U.S. at 586–87.

²⁴⁵ *Id.* at 588.

²⁴⁶ *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 565 (1985) (internal quotation marks omitted).

²⁴⁷ *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.").

²⁴⁸ See *Sag*, *supra* note 4, at 391.

²⁴⁹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 721–22 (9th Cir. 2007) ("Here, Google uses Perfect 10's images in a new context to serve a different purpose.").

reasonable in light of the purpose of an image search engine.²⁵⁰ As the court explained, while a user can identify relevant text by seeing merely a fraction of it, understanding the relevance of an image requires seeing a representation of the complete image.²⁵¹ In *Perfect 10*, as in *Kelly*, the court found that the third fair use factor did not weigh in favor of either party.²⁵²

Likewise, in *Field*, the court found that making entire web pages available in the search engine cache served a purpose that could not be effectively accomplished by using only portions of the web pages. The court found that Google's nonexpressive uses of the cached pages—such as verifying the authenticity of live pages and assessing the relevance of search queries—required caching complete reproductions of the plaintiff's web 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.”²⁵³

Furthermore, the numerous copyright cases dealing with the practice of reverse engineering computer software also support the proposition that a nonexpressive use is qualitatively insignificant under the third factor. In *Sony v. Connectix*, for example, the court acknowledged that Connectix had copied an entire section of Sony's 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.”²⁵⁴

The third factor requires a holistic assessment of the extent of appropriation of a work's expressive value, measured against the need and justification of the defendant in appropriating it. Accordingly, courts frown upon nontransformative expressive uses of a copyrighted work, but grant considerable latitude to transformative expressive uses. Similarly, nonexpressive uses, even those that require total copying, should be generally 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 Nonexpressive Uses.*—The fourth fair use factor is “the effect of the use upon the potential market for or value of the copyrighted work.”²⁵⁵ The *Harper & Row* Court described the fourth fair

²⁵⁰ *Id.* at 724 (citing *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003)).

²⁵¹ *Id.*

²⁵² *Id.*; *Kelly*, 336 F.3d at 821.

²⁵³ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1121 (D. Nev. 2006). Presumably, the court would have come to a different conclusion if it found that the primary use of cached links was to substitute for the original web page.

²⁵⁴ *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 606 (9th Cir. 2000) (internal quotation marks omitted).

²⁵⁵ 17 U.S.C. § 107(41) (2006).

use factor as “undoubtedly the single most important element of fair use.”²⁵⁶ Barton Beebe, in contrast, concludes that the fourth factor is “no factor at all.”²⁵⁷ As detailed below, although the fourth factor risks collapsing into circularity because everything is a potential market effect, courts have in fact avoided this nadir by applying certain limiting principles that 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 to the extent that a use is nonexpressive, it typically has no *cognizable* market effect under the fourth factor.

The first step in ascertaining the market effect of an unauthorized use is to define 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 circularity.²⁵⁸ 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.”²⁵⁹ The aggregation of any harm that is likely to result from widespread use is reasonable in evaluating the fourth factor. The Court, however, offers no particular reason to presume that *all* uses will become widespread.²⁶⁰

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.²⁶¹ Taken at face value it becomes impossible for a defendant to prove that her particular use, if widely replicated, would not displace

²⁵⁶ *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 566 (1985).

²⁵⁷ Beebe, *supra* note 235, at 620–21 (“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.”) (footnote omitted).

²⁵⁸ 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 is unfair, there is axiomatically at least one potential licensee (the defendant), and thus the copyright owner’s market has been adversely affected.

²⁵⁹ *Harper & Row*, 471 U.S. at 568 (internal quotation marks omitted).

²⁶⁰ 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.” S. REP. NO. 94-473, at 65 (1975) (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. See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994).

²⁶¹ 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. See *id.*

some potential future market in some derivative of the copyright owner's work. As the Second Circuit noted in *American Geophysical Union v. Texaco Inc.*, "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."²⁶²

Courts avoid this potential circularity by adopting a number of limiting principles in relation to the fourth factor. First, the unlikelihood of a derivative market may limit its relevance: "The market for potential derivative uses includes only those that creators of original works would in general develop or license others to develop."²⁶³ 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 that courts refuse to recognize illustrate again that 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.

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

[W]hen 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.²⁶⁴

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. Courts have consistently held that making unauthorized copies of a computer program, as a necessary step in reverse engineering, is fair use.²⁶⁵ For

²⁶² 60 F.3d 913, 929 n.17 (2d Cir. 1994) (citations omitted); *see also* Leval, *supra* note 215, at 1124 ("By definition every fair use involves some loss of royalty revenue because the secondary user has not paid royalties.").

²⁶³ *Campbell*, 510 U.S. at 592.

²⁶⁴ *Id.* at 591-92 (quoting, in part, *Fisher v. Dees*, 794 F.2d 432, 438 (9th Cir. 1986); BENJAMIN KAPLAN, AN UNHURRIED VIEW OF COPYRIGHT 69 (1967)) (internal quotation marks and citations omitted).

²⁶⁵ *See, e.g., Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 606 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000) (holding that Connectix's copying of Sony's copyrighted basic input-

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 platforms capable of playing Sony Playstation games, the Virtual Game Station was a "legitimate competitor" in that market.²⁶⁶ 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."²⁶⁷

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."²⁶⁸ Consistent with its decision in *Sega Enterprises v. Accolade, Inc.*,²⁶⁹ the Ninth Circuit held in *Sony* that intermediate copying of software is fair use if the copying was necessary to gain access to the functional elements of the software.²⁷⁰ The court based its ruling firmly on the importance of maintaining the idea-expression distinction: "We drew this distinction because the Copyright Act protects expression 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."²⁷¹ 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 nonexpressive uses such as reverse engineering, courts have consistently held that the protection that copyright affords is limited to certain cognizable markets. Trans-

output system (BIOS) during reverse engineering, used by Connectix to develop a software program that emulates the functioning of the Sony PlayStation console for regular computers, was fair use); *Atari Games Corp. v. Nintendo of Am., Inc.*, 975 F.2d 832, 842–43 (Fed. Cir. 1992) (observing that Atari's reverse engineering of Nintendo's 10NES program would have been a fair use of the program, except that Atari did not possess an authorized copy of the work); *Sega Enter. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1520 (9th Cir. 1992) (holding that Accolade's reverse engineering of Sega's video game programs in order to figure out how to make its own games compatible with Sega's Genesis system is a fair use); see also David A. Rice, *Copyright and Contract: Preemption After Bowers v. Baystate*, 9 ROGER WILLIAMS U. L. REV. 595, 601 n.19 (2004) (collecting cases). Circumventing encryption for the purpose of reverse engineering is also allowed under the safe harbor provisions of the DMCA. See 17 U.S.C. § 1201(f) (2006).

²⁶⁶ 203 F.3d at 607; see also *Sega*, 977 F.2d at 1522–23.

²⁶⁷ *Sony*, 203 F.3d at 607; see also *Sega*, 977 F.2d at 1523–24.

²⁶⁸ 203 F.3d at 598.

²⁶⁹ 977 F.2d 1510.

²⁷⁰ *Sony*, 203 F.3d at 607.

²⁷¹ *Id.* at 603 (citing *Sega*, 977 F.2d 1510).

formative expressive uses do not usually 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 cases of nonexpressive use because, to the degree that a particular use is nonexpressive, it has literally no potential substitution effect on a cognizable copyright market.

* * *

As established earlier in this Part, the exclusive rights of the copyright owner typically hinge upon the communication of original expression to the public. Acts of copying which do not communicate the author's original expression to the public should not generally be held to constitute copyright infringement. The application of this principle to anomalous copyright subject matter must be considered carefully. Rightly or wrongly, Congress has extended copyright protection to computer software and architectural plans in order to provide incentives for the development of these primarily functional objects. As already conceded, while computer programs are treated as expressive literary works, their expressive elements may be secondary to the nonprotectable functional output of the program—i.e., what it actually does. In consequence, the everyday use of a computer program is nonexpressive, but that does not suggest that copyright protection for software should be effectively dismantled. Instead, courts must exercise caution when dealing with anomalous copyright subject matter so as not to negate the very protection Congress intended.

The most appropriate method of doctrinal incorporation of the principle of nonexpressive use is through the fair use doctrine. The role of expressive substitution is not merely compatible with the fair use doctrine; it is actually necessary to make sense of much 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 the doctrine favors acts of copying that are unlikely to substitute for the copyright owner's original expression explains the majority of cases. Like transformative expressive uses, primarily nonexpressive uses should generally be classified as fair uses because, by their very nature, they do not substitute for the author's original expression. Accordingly, like transformative use, nonexpressive use should be favored under the first, third, and fourth factors—such uses are nonsubstitutive 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.²⁷²

²⁷² As is so often the case, the second statutory factor does not appear to have much bite in the context of nonexpressive uses, and thus does little to “separat[e] the fair use sheep from the infringing goats.” *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 586 (1994).

III. THE DOCTRINAL SIGNIFICANCE OF TRANSACTION COSTS

Copyright disputes involving copy-reliant technology may be completely resolved once a court determines that the use in question is nonexpressive. However, there are nonetheless two reasons for delving further into the issues of transaction costs and opt-outs that preoccupy this final Part. First, opt-outs are a common feature of copy-reliant technology and their ubiquity deserves some explanation. Second, although the correct application of the nonexpressive 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. Section 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. Section B discusses the relationship between transaction costs and the form and content of property rights generally. Section 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. Often, the sheer number of transactions that must typically be accommodated by copy-reliant technologies makes the transaction costs problems they face unique. The irony of copy-reliant technology is that, while technology has helped reduce the per unit transaction cost in relation to 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 does not necessarily offer a solution in the context of copy-reliant technologies due to the scale, decentralization, and heterogeneity of the 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-1990s. 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, musical

works, manuscripts, and other printed materials.²⁷³ The volume of material available on the Internet has dwarfed this number in a very short period. There are now an estimated 1.7 billion Internet users worldwide.²⁷⁴ 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.²⁷⁵ This 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 cost 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 she 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, but has radically decentralized it as well. The Gartner consulting firm estimates that around 100 million writers actively maintained a personal website or blog in 2007.²⁷⁶ Even as the mainstream press continues to consolidate into fewer and fewer media empires, the Internet has decentralized news production and increased both its volume and its diversity.²⁷⁷ Not only are these

²⁷³ See The Library of Congress, About the Library, <http://www.loc.gov/about/facts.html> (last visited July 29, 2009). Fewer than 32 million of these items are books. *Id.*

²⁷⁴ Internet World Stats, World Internet Usage and Population Statistics, <http://www.Internetworldstats.com/stats.htm> (last visited Nov. 22, 2009).

²⁷⁵ The Internet Archive is a nonprofit organization 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. The Internet Archive, About the Internet Archive, <http://www.archive.org/about/about.php> (last visited June 22, 2009).

²⁷⁶ Antony Savvas, *Gartner's Top 10 Forecasts for 2007 and Beyond*, COMPUTERWEEKLY.COM, Dec. 15, 2006, available at <http://www.computerweekly.com/Articles/2006/12/15/220726/gartners-top-10-forecasts-for-2007-and-beyond.htm>.

²⁷⁷ 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).

actors decentralized, they are also diverse.²⁷⁸ The “new media” and distributed production exemplified by blogs and social networking sites are characterized by (1) a blurring of the lines between producers and consumers, (2) a significant degree of interaction between participants who are both producers and consumers, and (3) low monetary costs, at least for the majority of participants.²⁷⁹ 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 cost problems if they intend to clear rights before unleashing their automated processes.

The high transaction cost environment is not limited to Internet search engines. Plagiarism detection software, which must search the whole Internet for possible sources of plagiarism, faces a similar problem of scale. The transaction cost 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 of Google’s partner libraries, and each one of these needs to be pulled off a shelf, scanned, and reshelved.²⁸⁰ The average cost of scanning each book is estimated at around \$10.²⁸¹ 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.²⁸² 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 (4) negotiate to obtain the permission of the owner(s).

Google’s clearance costs will vary according to the book in question: broadly speaking, 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).²⁸³ The clearance costs for very old and very new works are quite

²⁷⁸ See generally BENKLER, *supra* note 277 (describing information production in a networked information economy and its inclusion of a broader range of participants).

²⁷⁹ *Id.* at 91–132.

²⁸⁰ Brian Lavoie, Lynn Silipigni Connaway & Lorcan Dempsey, *Anatomy of Aggregate Collections: The Example of Google Print for Libraries*, 11(9) D-LIB MAG., Sept. 2005, available at <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html>.

²⁸¹ 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, at F1. (“[P]ress 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-TRIB., Aug. 10, 2006, at A4 (reporting that the University of California estimates that it would cost it \$30 to \$40 per book to scan its collection).

²⁸² Google’s recent class settlement with The Authors Guild, Inc., the Association of American Publishers, Inc., and the broad class of authors and publishers they represent effectively resolves the rights clearance issue for Google Book, as with respect to U.S. copyright interests. A draft of the proposed Settlement Agreement is available at <http://books.google.com/books/rightsholders/agreement-contents.html>.

²⁸³ The significance of each of these dates is explained below.

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.²⁸⁴ 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.²⁸⁵ Recall *Tasini*, where the Supreme Court held in favor of freelance journalists in relation to the inclusion and display of their articles in online databases.²⁸⁶ Rather than providing the freelance authors who it had already paid once for their articles with a new stream of revenue, the Times Company and other publishers simply removed most of the freelance pieces from their online databases. Since *Tasini*, it has become industry practice to insist on very broad rights in relation to electronic storage and retrieval.²⁸⁷

However, for the vast numbers of books published between 1923 and 2001, the applicable clearance costs are likely to be quite high.²⁸⁸ 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 twenty-eight years, plus a renewal period of twenty-eight years.²⁸⁹ This renewal period was subsequently extended to forty-seven years, and then again to sixty-seven years.²⁹⁰ The result is that if a work was published in the United States before 1963 and not renewed, it is in the public domain.²⁹¹ The Copyright Renewal Act of 1992 made renewal automatic for works first copyrighted between January 1, 1964, and December 31, 1977.²⁹² Likewise, certain works may also be in the public

²⁸⁴ Note that a work created but not published prior to 1978 may not be in the public domain. See generally Elizabeth Townsend Gard, *January 1, 2003: The Birth of the Unpublished Public Domain and Its International Implications*, 24 CARDOZO ARTS & ENT. L.J. 687 (2006); R. Anthony Reese, *Public but Private: Copyright's New Unpublished Public Domain*, 85 TEX. L. REV. 585 (2007).

²⁸⁵ Google and Amazon.com have each established cooperative agreements with publishers that allow them to display considerable portions of books in a searchable database. See Google, Google Books Partner Program, <http://books.google.com/support/partner/bin/answer.py?answer=106167> (last visited July 1, 2009); Amazon.com, Search Inside! Participation Agreement, http://www.amazon.com/gp/html-forms-controller/SITB_Publisher_Signup_Form (last visited July 1, 2009).

²⁸⁶ See *supra* Part II.A.2; N.Y. Times Co. v. Tasini, 533 U.S. 483 (2001).

²⁸⁷ See Amy Terry, *Tasini Aftermath: The Consequences of the Freelancers' Victory*, 14 DEPAUL-LCA J. ART & ENT. L. 231, 238–39 (2004).

²⁸⁸ 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* 11 (Tulane Public Law Research Paper No. 07-09, 2007), available at <http://ssrn.com/abstract=1015575>.

²⁸⁹ 1909 Act, §§ 23–24, 35 Stat. 1075, 1080–81.

²⁹⁰ 17 U.S.C. § 304(a) (1992) (current version at 17 U.S.C. §304(a) (2006)).

²⁹¹ Thanks to the combined efforts of Carnegie Mellon Universal Library Project, Project Gutenberg, Distributed Proofreaders, and a Google software engineer named Jarkko Hietaniemi, the paper-based renewal record maintained by the U.S. Copyright Office can now be searched electronically. See Jon Orwant, *U.S. Copyright Renewal Records Available for Download*, INSIDE GOOGLE BOOKS, June 23, 2008, <http://booksearch.blogspot.com/2008/06/us-copyright-renewal-records-available.html>.

²⁹² Copyright Renewal Act of 1992, Pub. L. No. 102-307 (codified as amended at 17 U.S.C. § 304 (2006)).

domain because they were published in the United States without the appropriate copyright notice.²⁹³ However, the notice requirement only applies to works first published before March 1, 1989.²⁹⁴ 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.²⁹⁵ Furthermore, in some circumstances it is possible for the author to correct a notice defect.²⁹⁶ 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 person created the work in her official capacity.²⁹⁷ The status of unpublished works and the works of foreigners adds several additional layers of complexity, including the possibility that some works may have been removed from the public domain by the Uruguay Round Agreements Act of 1994.²⁹⁸

Second, even if a work remains subject to copyright, the ownership of those rights can be highly uncertain. The basic principle of copyright law is that copyright vests initially in the author or authors of the work.²⁹⁹ However, those rights may be assigned in an infinite chain of transactions, bankruptcies, and inheritances. 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,³⁰⁰ even if that domicile is in a foreign country.³⁰¹ Furthermore, unvested renewal rights under the 1909 Copyright Act and rights of termination under the 1976 Copyright Act pass according to complex statutory provisions, which may, or may not, mirror the author's

²⁹³ In this regard, the requirement for copyright notice functions as an opt-in. See Christopher Sprigman, *Reform(aliz)ing Copyright*, 57 STAN. L. REV. 485, 487 (2004) (discussing the role of copyright formalities in "facilitat[ing] licensing by lowering the cost of identifying rightsholders").

²⁹⁴ 17 U.S.C. §§ 401–02 (1976) (current version at 17 U.S.C. §§ 401–02 (2006)) (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.

²⁹⁵ Communication "to a definitely selected group, and for a limited purpose, without the right of diffusion, reproduction, distribution or sale . . ." without a copyright notice does not inject a work into the public domain. *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) (quoting NIMMER ON COPYRIGHT § 4.13[A]). Publication can be especially ambiguous for copies of letters deposited with a library.

²⁹⁶ 17 U.S.C. § 405 (1976) (current version at 17 U.S.C. § 405 (2006)) (prior to the Berne Convention Implementation Act).

²⁹⁷ See 17 U.S.C. § 105 (2006).

²⁹⁸ The Uruguay Round Agreements Act of 1994 restores copyright protection to certain foreign works which had fallen into the public domain due to failure to comply with formalities, but which would otherwise now be eligible for protection in the United States. See 17 U.S.C. § 104A(d)(3)(A)(i) (2006).

²⁹⁹ 17 U.S.C. § 201(a). This is subject to the work for hire doctrine and the possibility of joint authorship. *Id.* at § 201(b), (c).

³⁰⁰ 17 U.S.C. § 201(d)(1).

³⁰¹ *Brecht v. Bentley*, 185 F. Supp. 890, 893 (S.D.N.Y. 1960) ("It by no means follows that because a proprietor under American Copyright Law is given no rights against an infringing foreign publication that his rights of ownership may not descend by the law of a foreign domicile.").

will or the default positions of her domicile.³⁰² Beyond the name of the initial author, the work discloses almost none of the facts relevant to determining the current ownership of copyright. The records of the U.S. Copyright Office also fail to contain the necessary information. As the work ages, the 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.³⁰³

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—they are only held by one of the participating libraries.³⁰⁴ It is estimated that slightly less than 20% of these works were published before 1923 and thus likely present no copyright issues.³⁰⁵ That leaves about 8.4 million books with some potential copyright constraint. Even if the average clearance cost (the cost of determining the status of the book, finding the relevant copyright owners, and negotiating a license) were as little as \$200, the total cost of rights clearance before any royalties have been paid would be over a billion dollars. It is easy to imagine that clearance costs could be in the thousands, not merely the hundreds, in which case the total cost of proactively 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 searching. Such advances, however, 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 1990s. Specifically, the prediction was that digital rights management technology (DRM) would allow copyright owners to automatically enforce their rights and to prevent uses that were once considered fair.³⁰⁶ Those who embraced the death of fair use online argued that DRM

³⁰² See 17 U.S.C. § 304(a); see also *Stewart v. Abend*, 495 U.S. 207, 209, 221 (1990) (holding that even if an author has assigned renewal rights, “the assignee holds nothing if the author dies before the commencement of the renewal period”).

³⁰³ See *infra* note 383 and accompanying text.

³⁰⁴ Lavoie, Connaway & Dempsey, *supra* note 280.

³⁰⁵ *Id.*

³⁰⁶ See generally Dan L. Burk & Julie E. Cohen, *Fair Use Infrastructure for Rights Management Systems*, 15 HARV. J.L. & TECH. 41 (2001). DRM refers to access control technologies that limit the use

would allow copyright owners to define the permissions associated with their works and make it possible to charge different prices to different users, thus reducing the need for fair use.³⁰⁷ Those who feared the death of fair use made the same prediction, arguing that that the control facilitated by DRM would enable an end run around the public policy values embedded in copyright law.³⁰⁸

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 at least two reasons. First, once the encryption on any one copy of a work is broken, that copy can be used to propagate an infinite number of unencrypted copies. Second, users tend to gravitate toward unrestricted formats precisely because they offer fewer restrictions.³⁰⁹ The network effects of this preference for unrestricted formats mean 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 CDs have been

of digital media or devices. See INFORMATION INFRASTRUCTURE TASK FORCE, INTELLECTUAL PROPERTY AND THE NATIONAL INFORMATION INFRASTRUCTURE: THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS 177–200 (1995).

³⁰⁷ 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 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, 580–84 (1998) (same); Trotter Hardy, *Property (and Copyright) in Cyberspace*, 1996 U. CHI. LEGAL F. 217, 233 (same); Edmund W. Kitch, *Can the Internet Shrink Fair Use?*, 78 NEB. L. REV. 880, 881 (1999) (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, 130–32 (1997) (same). See also Jane C. Ginsburg, *Authors and Users in Copyright*, 45 J. COPYRIGHT SOC'Y U.S.A. 1, 15 (1997) (“[T]he primary justification for exempting private copying as fair use has been transaction costs, but these are much attenuated in the digital world.”).

³⁰⁸ See Burk & Cohen, *supra* note 306; 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).

³⁰⁹ Both Apple and Amazon have adopted this strategy. See Press Release, Apple.com, Apple Unveils Higher Quality DRM-Free Music on the iTunes Store, (Apr. 2, 2007), available at <http://www.apple.com/pr/library/2007/04/02itunes.html>; Evan Blass, *Amazon Announces DRM-Free MP3 Music Store*, ENGADGET, May 16, 2007, <http://www.engadget.com/2007/05/16/amazon-announces-drm-free-mp3-music-store>. Note also that consumers may distrust DRM 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 Seán Byrne, *Microsoft Postpones MSN Music DRM Server Shut-Off*, CDFREAKS.COM, June, 19, 2008, <http://www.myce.com/news/Microsoft-postpones-MSN-Music-DRM-server-shut-off-14762>.

somewhat disastrous.³¹⁰ To the extent that copyright owners have had any success with DRM, these successes look more like “sufficient” control within the traditional contours of copyright law, not the Orwellian vision of a digital lockdown predicted over a decade ago.³¹¹

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 premillennial 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 adoption of new technologies has reduced the former, but not the latter.

3. *Collective Rights Management and Copy-Reliant Technologies.*— Collective management is the exercise of copyright and related rights by organizations acting on behalf of the owners of rights.³¹² 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 pa-

³¹⁰ 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, 81 (discussing consumer class action lawsuits and law enforcement proceedings against Sony resulting from an attempt to prevent audio CD copying by consumers); J. Alex Halderman & Edward W. Felten, *Lessons from the Sony CD DRM Episode*, 15 USENIX SECURITY SYMPOSIUM PROCEEDINGS 77 (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) (arguing that there are very good reasons to doubt the meaningful impact of DRM anytime soon).

³¹¹ See Jane C. Ginsburg, *The Pros and Cons of Strengthening Intellectual Property Protection: Technological Protection Measures and Section 1201 of the U.S. Copyright Act* 24 (Columbia Pub. Law Research Paper No. 07-137, 2007) (“To date, ‘digital lock-up’ persists in spectral guise, a grim, yet untranspired, anticipation.”).

³¹² 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 (last visited Apr. 12, 2009).

tent pools and joint ventures, such as the 3G Patent Platform Partnership.³¹³ The success of collective rights management in some fields demonstrates that market based solutions can overcome high transaction costs in situations where the individual management of rights is impossible or impractical.³¹⁴ However, this particular type of private ordering solution may not 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.

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.³¹⁵ 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 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 into that system. Given the billions of works at issue and the hundreds of millions of rights owners involved, similar levels of participation seem unlikely, at least for search engines. On the other hand, it is actually somewhat surprising that high schools and universities have not established a centralized body to facilitate plagiarism detection. These institutions have a great interest in advancing plagiarism software and could easily make participation in an antiplagiarism database a condition of entry. In this and other areas, even where the prerequisites for effective collective rights management appear to exist, rights holders and the relevant intermediaries have been slow to take advantage of the potential savings collective action offers.³¹⁶

³¹³ 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. See generally Reiko Aoki and Aaron Schiff, *Promoting Access to Intellectual Property: Patent Pools, Copyright Collectives and Clearinghouses*. 38(2) R&D MANAGEMENT 189–204 (2008) (reviewing and comparing patent pools, intellectual property clearinghouses, and copyright collectives as alternative systems for promoting efficient access to licensable intellectual property).

³¹⁴ Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CAL. L. REV. 1293, 1293–94 (1996). But note that the continuing court supervision of ASCAP pursuant to an antitrust decree indicates that it is not the best example of a purely private market solution. See *United States v. Am. Soc’y of Composers, Authors and Publishers*, No. Civ.A. 42-245, 1950 WL 42273, at *4 (S.D.N.Y. Mar. 14, 1950).

³¹⁵ *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.”).

³¹⁶ 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).

4. *Private Ordering Through Opt-Outs.*—It has been established thus far that transaction costs present a formidable potential obstacle to copy-reliant technologies, one that is unlikely to be overcome by either DRM technology or collective rights management. How then do copy-reliant technologies continue to function? The answer is that copy-reliant technologies typically rely on a combination of well understood default rules and technologically enabled opt-out mechanisms to reduce transaction costs.

Grasping 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 that allows any endpoint (usually a computer) to communicate with any other endpoint through a set of standard protocols.³¹⁷ 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, a password, or both. They can also control how search engines interact with their copyrighted material by employing a technological device known as the Robots Exclusion Protocol.³¹⁸ 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 for three reasons. First, the initial design of the Internet and its basic protocol for the exchange of information embedded an open architecture.³¹⁹ The open, minimalist, and neutral design of TCP/IP has enabled an unparalleled diversity of social and technological innovations.³²⁰ Open systems and end-to-end architecture were fundamental early Internet technical standards. The default of open access therefore 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 seen and their emails 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.

³¹⁷ JACK GOLDSMITH & TIM WU, WHO CONTROLS THE INTERNET? ILLUSIONS OF A BORDERLESS WORLD 23 (2006).

³¹⁸ See *infra* notes 320–324 and accompanying text.

³¹⁹ See Vinton G. Cerf & Robert E. Kahn, *A Protocol for Packet Network Intercommunication*, 22(5) IEEE TRANSACTIONS ON COMM. 637, 637–48 (1974), available at <http://www.cs.princeton.edu/courses/archive/fall06/cos561/papers/cerf74.pdf> (the original specification of the “Transmission Control Protocol”); CHRISTOS J. P. MOSCHOVITIS ET AL., HISTORY OF THE INTERNET: A CHRONOLOGY, 1843 TO THE PRESENT 80–82 (1999) (an excerpt from a 1974 paper discussing the architecture of the Internet); GOLDSMITH & WU, *supra* note 317, at 23.

³²⁰ TCP/IP stands for Transmission Control Protocol/Internet Protocol. Together, these two networking protocols largely control the movement of data across the Internet. See GOLDSMITH & WU, *supra* note 317, at 22–24.

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.³²¹ 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 to a website is trivial and there are a number of widely available free tools for automatically generating a robots.txt file.³²² To disallow all robots from a website simply requires two lines of code:

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

Adding these instructions to the robots.txt file at the root level of a website³²³ will block all compliant search engine robots and other information harvesting software agents.³²⁴ As explained in section C below, software architects who deliberately disregard the instructions contained in a robots.txt file and thus do not facilitate an opt-out regime will be less able to rely on a fair use defense than those who do.

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.³²⁵ The Google search engine, for example, is designed to allow site owners to prevent individual

³²¹ In this context, a “robot” is synonymous with a web crawler. 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* In a Nutshell, <http://www.robotstxt.org/robotstxt.html> (last visited Apr. 12, 2009). 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>.

³²² 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 is available at <http://support.microsoft.com/kb/217103>. Other websites that provide additional tools include <http://robots.googletoad.com/> and <http://www.seochat.com/seo-tools/robots-generator/>.

³²³ For example: “<http://www.example.com/robots.txt>.”

³²⁴ See TARA CALISHAIN & RAEL DORNFEST, *GOOGLE HACKS: 100 INDUSTRIAL-STRENGTH TIPS & TOOLS* 309 (2003); THOMAS A. POWELL, *WEB DESIGN: THE COMPLETE REFERENCE* 247–49 (2000).

³²⁵ For example: “Disallow: /nameofdirectory/.”

pages, sections of a website, or an entire website from being indexed.³²⁶ In the event that a website owner changes its preferences by activating the robots exclusion standard after Google has already indexed the content, the Google search engine will remove this content the next time Google “crawls” it.³²⁷ Google’s implementation of the Robots Exclusion Protocol is also highly customizable: among other things, site owners can also remove either or both of the snippets and images that appear in the search results.³²⁸

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 manageable, if not trivial.³²⁹ The Robots Exclusion Protocol and similar opt-outs do not actually clear rights in the sense of negotiating licenses, but they substitute for rights clearance where copyright owners would price access at zero—those who are not content with a zero price still retain the option to negotiate for something more than the default. *Field*, *Perfect 10*, and *Google Book* are interesting partly because the plaintiffs in those cases chose to object to the default rule instead of simply opting out.³³⁰

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.”³³¹ The rallying cry of property rights here is more rhetorically than analytically useful. When a new office building casts a shadow over a hotel swimming pool or obstructs the air current so as to impede the operation of a windmill, it is pointless for the affected parties to simply proclaim their

³²⁶ See CALISHAIN & DORNFEST, *supra* note 324, at 315; Google, Removing My Own Content From Google’s Index, <http://www.google.com/support/webmasters/bin/answer.py?answer=35301&topic=8459> (last visited July 29, 2009).

³²⁷ Google, *supra* note 326.

³²⁸ *Id.*

³²⁹ See John S. Sieman, *Using the Implied License to Inject Common Sense into Digital Copyright*, 85 N.C. L. REV. 885, 891 (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.”).

³³⁰ The motivations for this discontent are explored further in Part III.C.2, *infra*.

³³¹ Nick Taylor, *Letter from the President—Q4 2005*, AUTHORS GUILD BULLETIN, Fall 2005, *available at* <http://www.european-writers-congress.org/upload/312006102637.pdf> at 14.

property rights are sacrosanct.³³² 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 owners assert conflicting claims? Both the doctrinal and welfare economics answers to this question must be resolved with reference to transaction costs. This section briefly summarizes some of the significant literature on transaction costs and explains why the mere assertion of “property rights” does very little to tell us whether the rights of a copyright owner include the right not to allow copy-reliant technology to interact with her work.

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.³³³ The allocation and definition of property rights determines both which individuals have the authority to decide how a specific resource is used and to whom the costs and benefits of that use will flow.³³⁴ 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.³³⁵ 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.³³⁶

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 that enable private exchange.³³⁷ On this view, the law should not devote considerable resources to optimizing initial allocation; it should just ensure that all the rights worth specifying are allocated.³³⁸ A second view is that, because substantial transaction costs persist even after private rights have been allo-

³³² These examples come from R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 8 (1960).

³³³ 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.

³³⁴ NICHOLAS MERCURO & STEVEN G MEDEMA, ECONOMICS AND THE LAW: FROM POSNER TO POSTMODERNISM AND BEYOND 249 (2d ed. 2006).

³³⁵ Coase, *supra* note 332.

³³⁶ NORTH, *supra* note 333.

³³⁷ ROBERT COOTER & TOMAS ULEN, LAW AND ECONOMICS 97 (4th ed. 2004).

³³⁸ *See id.* (“By lubricating bargaining, the law enables the private parties to exchange legal rights, thus relieving lawmakers of the difficult task of allocating legal rights efficiently.”).

cated, lawmakers should allocate property rights to their best initial use so as to minimize the harm caused by inevitable failures to reach private agreements.³³⁹

These contending implications of the *Coase Theorem* are frequently offered as a binary choice. Because the applicability of either depends on the exact nature of the transaction costs at issue, however, courts should accept neither prescription as dogma. In fact, just as the decision of a firm to either make or buy is determined by relative transaction costs, so too is the structure of property rights.³⁴⁰ In this context, it should be noted that the form that property rights take can play a significant 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.³⁴¹

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.”³⁴² 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.³⁴³

Consideration of information costs suggests a possible divergence between property in tangible and intangible objects—whereas physical proximity limits the informational burden of tangible property, those same burdens can multiply almost infinitely in the case of intellectual property. Only those walking past Blackacre need to worry where its boundaries are; every musician in the world needs to worry that their new composition might have been inadvertently copied from any one of thousands of pop

³³⁹ See *id.* at 97–98 (discussing this theory as the “Normative Hobbes Theorem”).

³⁴⁰ See generally R.H. Coase, *The Nature of the Firm*, 4 *ECONOMICA* 386 (1937) (observing that differences in transaction costs explain variation in organizational hierarchy).

³⁴¹ 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 implications of transaction costs for different remedial structures).

³⁴² Thomas W. Merrill & Henry E. Smith, *What Happened to Property in Law and Economics?*, 111 *YALE L.J.* 357, 358–59 (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).

³⁴³ See Merrill & Smith, *Optimal Standardization*, *supra* note 342.

songs their brains have absorbed over the years.³⁴⁴ 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.³⁴⁵ In other words, while the rights attached to real property and chattels might be fuzzy at the edges, the ambiguity of rights attached to intangible expression is usually more pronounced.³⁴⁶ Accordingly, the 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.³⁴⁷

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.³⁴⁸ 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.³⁴⁹ 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 like tort and the latter more like property.³⁵⁰

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 cen-

³⁴⁴ 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*. See *Bright Tunes Music Corp. v. Harrison*, 420 F. Supp. 177 (S.D.N.Y. 1976). The alleged similarities escape this author.

³⁴⁵ See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 16 (2003) (noting that transaction costs tend to be higher in intellectual property as it is frequently difficult to identify such property because by definition it has no unique physical site).

³⁴⁶ 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. 1930).

³⁴⁷ See LANDES & POSNER, *supra* note 345, at 21 (concluding that intellectual property rights tend to be more costly than rights in physical property).

³⁴⁸ *Id.* at 21, 92–93, 111.

³⁴⁹ Clarisa Long, *Information Costs in Patent and Copyright*, 90 VA. L. REV. 465 (2004).

³⁵⁰ Henry E. Smith, *Intellectual Property as Property: Delineating Entitlements in Information*, 116 YALE L.J. 1742 (2007); see also 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); Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783 (2007) (making a transaction cost argument against the imposition of injunctions in cases where courts cannot easily tailor injunctions to forbid only the prohibited conduct).

tral 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.³⁵¹ Fair use plays a critical role where copyright owners attempt to use their rights to stifle criticism or prevent the articulation of contrary viewpoints.³⁵² However, the doctrine is not limited to facilitating 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. As the market failure paradigm demonstrates, the fair use doctrine is particularly important in situations where the costs of obtaining permission outweigh the benefits of the use. Thus, according to at least one court, the fair use 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.³⁵³ Consequently, the fair use doctrine economizes on transaction costs by making such disclaimers unnecessary.³⁵⁴

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.³⁵⁵ 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.³⁵⁶ The basic problem is that where several parties possess 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

³⁵¹ Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600 (1982).

³⁵² See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994) (discussing parody as having social value); *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257, 1278 (11th Cir. 2001) (discussing parody as entertainment and social and literary criticism).

³⁵³ *Ty, Inc. v. Publ'ns Int'l*, 292 F.3d 512, 517 (7th Cir. 2002).

³⁵⁴ *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, LANDES & POSNER, *supra* note 345, at 120–22, but collective action and credible commitment are more convincing explanations.

³⁵⁵ See Sag, *supra* note 29, at 250 (criticizing doctrinal recommendations which aim to optimize copyright scope in the abstract but do not account for the effect of uncertainty or strategic behavior).

³⁵⁶ See Zohar Goshen, *Voting (Insincerely) in Corporate Law*, 2 THEORETICAL INQUIRIES L. 815, 820 (2001) (explaining the holdout problem in the corporate governance setting); 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, 351 (1990) (citing examples such as “oil field unitization, . . . a group of workers deciding on whether to leave their firm and establish a new venture . . . , [and] the construction of a production facility generating pollution”).

experience and theory suggest that the mere presence of a divisible surplus does not ensure that the parties will in fact agree on how that division should proceed.³⁵⁷ Furthermore, agreements on division are constrained by the costs of enforcement given that once a deal has been reached, “[a]ccording to a 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 available to the party.”³⁵⁸

The problems attending strategic behavior are particularly relevant to copyright because all copyrighted works build on previous works to some extent. Musicians attempting to clear samples can face license demands from the original copyright owners that effectively seek to expropriate the entire value of the newly created work.³⁵⁹ Strategic behavior may prevent parties who would otherwise have much to gain from cooperating if multiple clearances are required, because it is quite rational for the players to adopt strategies that risk destroying the surplus in order to gain a larger share.³⁶⁰ Even nonstrategic parties sometimes cause holdout problems because of the divergent valuations that result from egotism and other cognitive biases.³⁶¹ The fair use doctrine reduces transaction costs associated with strategic behavior by eliminating the holdout power of the copyright owner in situations where her contribution is comparatively small to that of the defendant, or where some degree of copyright owner intransigence is effectively presumed.

The salient point 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 copy-reliant technology to interact with her work.³⁶² A further important point remains: even if that issue is resolved in the copyright owner’s favor, the form of that property right still remains an open question. It is not enough to simply determine the appropriate “property rules,” “liabili-

³⁵⁷ See 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).

³⁵⁸ NORTH, *supra* note 333, at 30.

³⁵⁹ See DONALD S. PASSMAN, ALL YOU NEED TO KNOW ABOUT THE MUSIC BUSINESS: REVISED AND UPDATED FOR THE 21ST CENTURY 306–08 (Simon & Schuster 2000) (1991); Ross, *supra* note 108, at 282–83 (discussing the costs of licensing fees of nonparody sampling in rap and hip hop music as prohibitively high for all but “established artists and the media Goliaths”).

³⁶⁰ Robert Cooter & Steven Marks with Robert N. Mnookin, *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).

³⁶¹ Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75, 89 (1994).

³⁶² Oren Bracha makes a similar point in relation to Google Book. See Bracha, *supra* note 2.

ty rules,” and “inalienability rules.”³⁶³ Even if we accept that the copyright owners’ rights are to be protected by a veto right (i.e., a property rule), the question remains: under what conditions, if any, can this right be exercised? As Abraham Bell and Gideon Parchomovsky observe 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.³⁶⁴

The notion that the rights of the property owner can be protected under permissive default rules 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.³⁶⁵ 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 are just one example. Indeed, beneath the visage of “property,” one sees a variegated landscape with rules tailored according 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.³⁶⁶ Specifically in relation to copyright, the suggestion that authors should be required to accept some cost before vindicating their rights is not unprecedented. As Michael Mattioli perceptively notes, “while formal registration and deposit are no longer strict requirements for copyrightability, both are demanded of authors who wish to bring infringement suits.”³⁶⁷

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, it is also inconclusive as to the form of those rights. The centrality of transaction costs in modern property theory and the practical importance of opt-out mechanisms in reducing

³⁶³ Calabresi & Melamed, *supra* note 341.

³⁶⁴ Abraham Bell & Gideon Parchomovsky, *Pliability Rules*, 101 MICH. L. REV. 1, 5 (2002) (commenting on Calabresi and Melamed’s methods).

³⁶⁵ See, e.g., ROBERT C. ELICKSON, *ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES* 76 (1991). But note that here norms appeared to govern behavior regardless of the underlying law.

³⁶⁶ Richard A. Epstein, *Intellectual Property: Old Boundaries and New Frontiers*, 76 IND. L.J. 803, 804–05 (2001). The Illinois Right of Publicity Act contains a special provision allowing professional photographers to exhibit photos which might otherwise infringe upon a person’s right of publicity “unless the exhibition is continued by the professional photographer after written notice objecting to the exhibition has been given by the individual portrayed.” See Illinois Right of Publicity Act, 765 ILL. COMP. STAT. 1075/35(b)(5) (2009).

³⁶⁷ Michael R. Mattioli, *Opting Out: Procedural Fair Use*, 12 VA. J.L. & TECH. 1, 20–21 (2007); see also 17 U.S.C. § 411(b) (2006) (“[N]o civil action for infringement of the copyright in any United States work shall be instituted until preregistration or registration of the copyright claim . . .”); *id.* § 412 (precluding statutory damages for infringements before the effective date of registration in many circumstances).

transaction costs for copy-reliant technologies make some analysis of the doctrinal implications of opt-out mechanisms essential. As in many other cases, the fair use doctrine provides the most natural framework for that analysis. The next section specifically addresses this question.

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”³⁶⁸—requires some determination of what types of uses should be preferred. 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.³⁶⁹

As discussed at length in Part II, one of these principles is that acts of copying that do not communicate the author’s original expression to the public should generally not be held to constitute copyright infringement. This follows from the essential observation that the 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,³⁷⁰ 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—“[b]y establishing a marketable right to the use of one’s expression”;³⁷¹ this marketability not only encourages authorship, it decouples authorship from the corrupting influences of state subsidy and elite patronage.³⁷² Copyright is not the only way to encourage authorship; a system of state prizes might do that just as well.³⁷³ The advantage of copyright over other systems is that it not only encourages authorship, it also gives authors a degree of autonomy. Accordingly, when evaluating a claim of fair use, courts should consider to what extent the defendant’s conduct as a

³⁶⁸ 17 U.S.C. § 107(1) (2006).

³⁶⁹ Sag, *supra* note 4, at 385.

³⁷⁰ See *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994); see also Leval, *supra* note 215, at 1111.

³⁷¹ *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) (citing *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985)).

³⁷² Netanel, *supra* note 231, at 288.

³⁷³ See, e.g., Steven Shavell & Tanguy van Ypersele, *Rewards Versus Intellectual Property Rights*, 44 J.L. & ECON. 525 (2001) (concluding that intellectual property rights do not possess a fundamental social advantage over reward systems and that an optional reward system—under which innovators choose between rewards and intellectual property rights—is superior to intellectual property rights).

whole undermines or enhances the autonomy not just of the author in the case, but of the class of authors the decision will affect. Put another way, when in doubt, courts should attempt to 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 legal rules 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 of taking advantage of the opt-out mechanism provided by the defendant is very low.³⁷⁴ In this situation, if the court determines that the defendant's 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; 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.³⁷⁵ 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 cost 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 defendant's use is unfair, then the majority must bear the burden of opting in. The cost of opting in might be especially high. This could be because of coordination problems in situations where the use requires 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 minimis*, both the majority and the minority will have their preferences respected. From an economic perspective, the efficiency of an opt-in versus an opt-out system will depend on the

³⁷⁴ Note that the copyright owner's monitoring costs are a significant element of the cost of the opt-out; accordingly, for this scenario to be true, monitoring costs must also be very small.

³⁷⁵ 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.

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.³⁷⁶ 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 have taken him "a matter of seconds."³⁷⁷ The legal significance of this finding tied directly 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.³⁷⁸ 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* 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 nonexpressive, 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."³⁷⁹ In other words, the presumed acquiescence of a large number of copyright owners who could very inexpensively 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 defendant's use in certain situations. Courts should consider whether (1) the defendant's proposed use is one which the majority of affected copyright owners would actually favor; (2) the costs of taking advantage of the opt-out are sufficiently small such that the autonomy of the minority is preserved; and (3) 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 and a majority of affected copyright owners while preserving the autonomy of the minority. In choosing rules that facilitate 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 ef-

³⁷⁶ *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006).

³⁷⁷ *Id.* at 1119.

³⁷⁸ *Id.*

³⁷⁹ *Id.* The court further remarked that "Google's alleged copying and distribution of Field's Web pages containing copyrighted works was transformative." *Id.*

fect of the use upon the potential market for or value of the copyrighted work.”³⁸⁰ 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.”³⁸¹ 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 in this scenario.

There is, however, a logical explanation for such behavior. A rational copyright owner will insist on a veto right rather than the right to opt-out under either one of two conditions: (1) where the expected costs of obtaining and exercising a veto are lower than the expected costs of taking advantage of the equivalent opt-out; or (2) 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 affirmatively tell each publisher in the United States that she did not wish them to publish her manuscript.³⁸² 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 may hold 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.³⁸³ *Perfect 10* both benefits from and is 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 latter 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 employing certain technical devices, *Perfect 10* has opted out of inclusion in image-based search engines but elected to remain visible to traditional text-oriented searches.³⁸⁴ In this way,

³⁸⁰ 17 U.S.C. § 107(1) (2006).

³⁸¹ *Id.*

³⁸² Of course, the © and “all rights reserved” notices serve this function as well.

³⁸³ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 725–29 (9th Cir. 2007).

³⁸⁴ The *Perfect 10* website also remains accessible through the Internet Archive’s Wayback Machine.

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: for Perfect 10, opting out of image-based searching is ineffective because it has not been able to prevent third parties from infringing its works. Perfect 10 thus illustrates the problem of a minority holdout 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 court to shift the costs of copyright enforcement onto Google and the public at large, which 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.³⁸⁵

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.³⁸⁶ The trial court found that Field had no 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.³⁸⁷ 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 caches of his web pages. The court rejected this transparently circular argument, noting that "the fourth fair use factor is not concerned with such syllogisms."³⁸⁸ 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—

³⁸⁵ See 487 F.3d at 725 ("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.").

³⁸⁶ See *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).

³⁸⁷ *Id.* at 1113 ("Field decided to manufacture a claim for copyright infringement against Google in the hopes of making money from Google's standard practice."). It should be noted that even if Google had not prevailed on the issue of fair use, the result in this particular case would have been unchanged because Field's invitation to Google's search robots amounted to an implied license.

³⁸⁸ *Id.* at 1121 n.9 (quoting *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 592 (1994)); see also *Religious Tech. Ctr. v. Netcom On-Line Comm'n. Servs., Inc.*, 907 F. Supp. 1361, 1378 n.25 (N.D. Cal. 1995) ("[T]here could hardly be a market for licensing the temporary copying of digital works onto computer screens to allow browsing."); *Mattel Inc. v. Walking Mt. Prods.*, 353 F.3d 792, 806 (9th Cir. 2003) (noting that "Mattel was unlikely to develop or license others to develop a product in the 'adult' doll market").

which was almost certainly nil. Instead, 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."³⁸⁹ 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 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 on the particular plaintiff before them;³⁹⁰ 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.³⁹¹

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 useful arts—by establishing a marketable right to an original expression.³⁹² 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, one that is significantly less reliant on state or elite subsidy. If the autonomy of the author is a freestanding 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 owners would consent to inclusion and the costs of individualized permission are much greater than

³⁸⁹ 17 U.S.C. § 107(4) (2006).

³⁹⁰ See *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 562 (1985).

³⁹¹ *Campbell*, 510 U.S. at 577–78.

³⁹² See *supra* notes 230–231 and accompanying text.

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.

* * *

A combination of permissive defaults and opt-out mechanisms is a common feature of copy-reliant technology. As this Article has shown, it is often critical for copy-reliant technologies to mitigate otherwise prohibitive transaction costs through opt-outs. Critics of particular copy-reliant technologies may argue that the use of opt-outs is irrelevant to the fair use analysis.³⁹³ This Article takes the contrary position. A dogmatic insistence 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 a low cost opt-out 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. This Article has demonstrated the pull of recent technological change on copyright law. Copy-reliant technologies—technologies that necessarily copy expressive works in large quantities, but do so for nonexpressive purposes—are vital to the operation of the Internet. And yet, because these technologies are so dependent on access to copyrighted works, they are also vulnerable to claims of copyright infringement.

An exploration of 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. The mere fact that common ground exists does not suggest that the copyright issues affecting these technologies must necessarily be resolved uniformly. There are also significant differences between the

³⁹³ See, e.g., 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, 67 (2006) (doubting the validity of the opt-out argument).

copy-reliant technologies surveyed in this Article. 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 generally be held to constitute copy-copyright infringement. In relation to traditional copyright subject matter at least, to do so would conflict with decades of accumulated precedent that limit the rights of copyright owners to those uses of their works that offer some threat of expressive substitution.

In spite of its centrality, the question of nonexpressive use may not fully resolve all copyright disputes involving copy-reliant technologies. Inevitably, courts will face cases where the line between expressive and nonexpressive use remains ambiguous. In many 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 help maintain order on the digital frontier. Such devices can play an essential role in 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-out mechanisms 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. 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 of default rules: these default rules should encompass the distinction between expressive and nonexpressive uses articulated in this Article; they should also take into account the role of transaction costs and facilitate the kind of private ordering that copyright has traditionally embraced.