

COPYRIGHT AND COPY-RELIANT TECHNOLOGIES

Matthew Sag*

Abstract

This Article demonstrates that the technological and social changes of the Internet era are also changes to the underlying structure of copyright markets. These changes necessitate certain adjustments in copyright doctrines that are themselves contingent on market conditions; most notably in relation to the fair use doctrine. In particular, this article studies the rise of copy-reliant technologies such as Internet search engines and plagiarism detection software. Although these technologies do not read, understand or enjoy copyrighted works, they necessarily copy them in order to process them. This article addresses two seemingly novel questions in the context of copy-reliant technology. First, whether a non-expressive use that nonetheless requires copying the entirety of a copyright work be found to infringe the exclusive rights of the copyright owner. Second, whether the transaction costs associated with copy-reliant technologies justify switching copyright's default rule that no copying may take place without permission to one in which copyright owners must affirmatively opt-out of specific uses of their works.

Please note that this paper is very much a work in progress. Parts I and II are included herein and are substantially complete. Many of the citations are incomplete. Part III is not included in this version of the manuscript.

* Visiting Assistant Professor, University of Virginia School of Law & Assistant Professor, DePaul University College of Law. This **limited circulation draft** is dated March 20, 2008 and is subject to revision by the author. Please address comments to msag@depaul.edu.

CONTENTS

Introduction..... 3

I. Copy-Reliant Technologies and The Internet 5

 A. The Effect of New Technologies on Copyright Markets..... 5

 B. The Role of Changes in Copyright Markets in Copyright Law..... 7

 C. Copy-Reliant Technologies 11

 D. Recurring Issues in Copy-Reliant Technologies..... 18

II. The Doctrinal Implications of Non-Expressive Use..... 23

 A. The Non-Expressive Use of Copyrighted Works by Copy-Reliant Technologies 24

 B. The Implications of Copying for a Non-Expressive Use..... 28

 C. Doctrinal Incorporation..... 38

 D. Fair Use and Non-Expressive Use 41

III. The Doctrinal Significance of Transaction Costs**Error! Bookmark not defined.**

INTRODUCTION

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

The rise of copy-reliant technologies exposes seemingly novel questions. First, should a non-expressive use that nonetheless requires copying the entirety of a copyright work be found to infringe the exclusive rights of the copyright owner? Second, do the transaction costs associated with copy-reliant technologies justify switching copyright’s default rule that no copying may take place without permission to one in which copyright owners must affirmatively opt-out of specific uses of their works?

The copyright implications of Internet search engines, plagiarism detection software, reverse engineering of software and the emerging Google Book Project controversy have been separately considered by other scholars.² This article is the first to provide a unifying theoretical framework for the analysis of the issues raised, largely because it is the first to recognize them as sub-parts of a broader phenomenon relating to copy-reliant technologies.³

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 every key stage of their operation. Search engines and other copy-reliant technologies display three significant traits: the copying of expressive works for non-expressive uses; clearance costs that are potentially astronomical because of the very large volume of transactions they engage in; and the use of technologically enabled opt-out mechanisms to address those transaction costs problems.

One of the purposes of this Article is to demonstrate that the technological and social changes of the Internet era are also changes to the underlying structure of copyright markets. These changes necessitate certain adjustments in copyright doctrines that are

¹ While the Internet itself dates back to October 29, 1969, 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. The first packet-switching node of what would later be called the ARPANET went live on October 29, 1969. The first TCP/IP-wide area network was operational by January 1, 1983, when the United States’ National Science Foundation (NSF) constructed a university network backbone that would later become the NSFNet.

² [Collect citations]

³ [Some further due diligence is still required to ensure that this claim is descriptively accurate.]

themselves contingent on market conditions; in particular the fair use doctrine. Scholars and commentators are sometimes confounded by the curious results produced in new technology cases.⁴ They should not be. To the extent that copyright markets function differently in the context of new technology, they may also produce very different applications of copyright doctrines to accommodate them.

The phenomenon of copy-reliant technology raises economic and doctrinal questions in relation to the significance of non-expressive uses of copyrighted works and the significance of opt-out mechanisms in fair use analysis. Our historical intuition is that when a work is copied it is copied to communicate at least some part of the work's original expression: books are copied to be read, not to serve as paper weights; compact discs are copied to be played, not to function as drink coasters.

Copy-reliant technology also forces us to reevaluate the significance and doctrinal implications of transaction costs. The transaction costs problems presented by copy-reliant technologies are quite different from other copyright contexts. The potential transaction costs faced by search engines and similar copy-reliant technologies are a function of the sheer magnitude of copyrighted works that comprise the Internet, the decentralization of the relevant actors, and the diverse nature of those actors. The nature of these transaction costs and the role of technologically enabled opt-out mechanisms in addressing them have significant implications to the application of the fair use doctrine.

The structure of this Article is as follows: Part I introduces the phenomenon of copy-reliant technologies and explains the relationship between technology, market structure and copyright law. Part I also introduces the four case-studies used throughout this Article and previews the issues of non-expressive use, transaction costs and private ordering in the context of these 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 Second, *Perfect 10 v. Amazon* centers on the creation and display of thumbnail representations of copyrighted photographs used by image-based search engines.⁶ The conduct challenged in the third case-study, the Google Book Project, relates both to the generation of metadata and to the display of uncopyrightable fragments of books as part of a menu of search results.⁷ The final case-study is that of plagiarism detection software which also addresses the use of copyrighted works to generate metadata.⁸

Part II explores the doctrinal implications of the non-expressive use of copyrighted works. This Part demonstrates that extending the rights of copyright owners to encompass non-expressive uses of their works by copy-reliant technologies would constitute a significant departure from existing copyright principles. The doctrinal incorporation of this principle of non-expressive use through the application of the fair use doctrine is then addressed in detail.

⁴ [Cite to Paul Goldstein's recent criticism of recent Internet search cases.]

⁵ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006), see *infra* Part I-C-1.

⁶ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701 (9th Cir. 2007), see *infra* Part I-C-2.

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

⁸ *Av et al v. Iparadigms, Llc*, US District Court Civil Docket 1:07cv293, ee *infra* Part I-C-4.

Part III studies the doctrinal implications of high transaction costs in relation to copy-reliant technologies and the use of opt-out mechanisms to resolve those transaction costs.

I. COPY-RELIANT TECHNOLOGIES AND THE INTERNET

This Part introduces and expands upon the concept of copy-reliant technologies in the context of the relationship between copyright law and technological change in general. The general framework is discussed in Parts I-A and I-B. Part I-A establishes how the technological and social changes of the Internet era should also be understood as changes to the underlying structure of copyright markets. Part I-B then addresses the responsiveness of copyright law to market conditions. As discussed in more detail below, in many instances, copyright's balance of rights and freedoms is struck with an eye toward market structure – on how effectively we expect initial authors and subsequent users of their works to cooperate and compromise. Much of the doctrinal analysis in the later parts of this Article is devoted to the fair use doctrine, because it is primary way in which copyright law adjusts to changes in market conditions.

Moving from the general to the specific, Parts I-C and I-D establish the concept of copy-reliant technology and address the specific issues raised by such technologies. Part I-C describes four case-studies of copy-reliant technology which serve to illustrate the concept and its application. I return to these case-studies throughout the Article. Based on these descriptions, Part I-D then reviews the recurring legal issues raised by copy-reliant technologies: the copying of expressive works for non-expressive uses; the potential for high transaction costs associated with copy-reliant technologies; and the role of opt-out mechanisms in addressing these transaction costs problems. These issues are then addressed more fully in Parts II and III.

A. The Effect of New Technologies on Copyright Markets

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

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

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

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

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 exemplifies how digital technology and online distribution allows existing works, such as sound recording and motion pictures, to be copied and distributed at virtually no cost.¹² Unlocking content from physical delivery has facilitated more than just piracy; it has also enabled legal digital music services which have made more music available at a lower cost than ever before.¹³

The same technology has also opened up new possibilities of creative production by reducing the cost of sound and video editing. The video editing software that was used to create *the Phantom Edit*¹⁴ – an edited version of *Star Wars I* without the much-reviled character Jar Jar Binks character – used to be reserved for Hollywood studios alone; it is now widely available for less than the cost of a new television.¹⁵ These new possibilities have done more than simply lower costs for existing producers; they have introduced new participants and in some cases dramatically changed the medium. Just as newspapers and television reporting have been changed by the rise of political blogs,¹⁶ advances in digital technology have created whole new genres of entertainment, such as the mash-ups which blend the vocal tracks from one song over the instrumental or rhythm track of another.¹⁷

Regarding these changes and their implications for copyright law as a mere continuation of past technological changes risks missing the significant transformation that has occurred.¹⁸ Digital technology and the Internet have significantly expanded the scope, diversity and sheer number of copyrighted works in existence. The cost of reproducing and disseminating digital works has not merely fallen; in many cases it has become entirely trivial. Thus, the Internet has seen not just an increase in copying, but an exponential increase. Similarly, copyright policy in the Internet age requires more than the inclusion of one or two neglected interest groups, copyright law now reaches deep inside the home and must take account of a much broader set of stakeholders than ever before.¹⁹ 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.

¹² [list cases involving peer-to-peer file sharing]

¹³ [Add cite for Apple iTunes and its effect on the price of music] See generally, CHRIS ANDERSON, *THE LONG TAIL* 139 (2006).

¹⁴ See, Amy Harmon, 'Star Wars' Fan Films Come Tumbling Back to Earth, N.Y. TIMES, Apr. 28, 2002, § 2 (Arts & Leisure), at 28. (discussing *Star Wars 1.1: The Phantom Edit* and its creation).

¹⁵ See, WIKIPEDIA, *List of video editing software*, http://en.wikipedia.org/wiki/List_of_video_editing_software for a list of video editing software, including several free and open source modules.

¹⁶ [Add cites on the effect of blogs on the mainstream media]

¹⁷ See, Roberta Cruger, *The Mash-Up Revolution*, Salon.com (http://dir.salon.com/story/ent/music/feature/2003/08/09/mashups_cruger/)

¹⁸ See also, Menell, *supra* note 000 at 64 (discussing the relationship between new technology and new modes of expression).

¹⁹ See, Litman, *Digital Copyright*, *supra* note 000; Pamela Samuelson, *Preliminary Thoughts on Copyright Reform* (working paper).

B. The Role of Changes in Copyright Markets in Copyright Law

By expanding the breadth, diversity and sheer number of copyrighted works in existence, the Internet has fundamentally changed the nature of copyright markets. To understand the significance of these changes for copyright law, it is first necessary to examine the economic function of copyright.

Copyright creates exclusive rights in certain forms of expression in order to give authors an incentive to create those works in the first place. However, these same exclusive rights raise the cost of acquiring works for both consumers and subsequent authors. In the world of tangible objects, these costs become price signals that ensure the efficient allocation of goods to those who value them most; however, given that expressive works can be consumed again and again by different people without diminishing their value, the exclusive rights established by copyright also result in some dead weight loss because those who are unwilling to pay the higher price are forced to go without the work in question.²⁰

The author's exclusive rights under copyright law provide a buffer against price competition. This buffer to competition allows the author to charge higher prices than she otherwise would, which in turn has two immediate effects. First, some consumers remain willing to purchase the work at a higher price and consequently pay more. Assuming we value the welfare of both consumers and authors equally, this is simply a wealth transfer and is welfare-neutral. Second, those who are unwilling to pay the higher price are forced to go without the work in question.²¹ 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.²²

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”²³ has long been understood. What is sometimes less clearly grasped is that where this balance should be struck depends not just on the relative needs of authors and consumers, but also on how effectively we expect those parties to cooperate and compromise.²⁴ In many situations, authors can license their creations with relative ease

²⁰ See generally, See Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, in THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 614-15 (Nat'l Bureau of Econ. Research ed., 1962); Richard R. Nelson, *The Simple Economics of Basic Scientific Research*, 67 J. POL. ECON. 297, 297-306 (1959).

²¹ This assumes, realistically, the absence of perfect price discrimination. See, Kathleen Carroll & Dennis Coates, *Teaching Price Discrimination: Some Clarification*, 66 S. ECON. J. 466, 471-78 (1999) (noting that the assumption that price discrimination is efficient is often implausible).

²² See e.g., STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW __ (2004) (describing the effect of exclusion on resource allocation).

²³ Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984).

²⁴ Matthew Sag, *Beyond Abstraction: The Law and Economics of Copyright Scope and Doctrinal Efficiency*, 81 TUL. L. REV. 187 (2006) (discussion the relationship between copyright scope and the effectiveness of private ordering). See also, Peter Menell & Suzanne Scotchmer, *Intellectual Property*, in

and the theoretical loss of exclusion is minimal.²⁵ 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”,²⁶ other times it is more like a cudgel. For example, Stephen Joyce, who controls the literary estate of his grandfather, James Joyce, has been accused of attempting to control access to unpublished material in order to influence historical and literary conceptions of his famous grandfather.²⁷ 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.²⁸ Biographers of Howard Hughes have faced similar difficulties.²⁹

Copyright law addresses potential market malfunctions in a number of ways. Doctrines such as the idea-expression distinction protect the expressive elements of the author’s work while guaranteeing subsequent authors the necessary breathing space to make their own contributions by adding to, re-using, or re-interpreting, the facts and ideas embodied in the original work.³⁰ Statutory exemptions and compulsory licenses – such as the special reproduction rights of libraries and archives,³¹ and the compulsory license for making and distributing phonorecords³² – also provide some breathing space, however their scope tends to be limited. The primary way in which copyright law adjusts to potential market malfunctions is through the evolution of the mercurial doctrine of fair use.

Fair use is a flexible standard which limits the scope of copyright protection and renders certain actions relating to copyrighted works non-infringing.³³ 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.³⁴ Judges and legal scholars frequently attest to the importance of the fair use

HANDBOOK OF LAW AND ECONOMICS (A. Mitchell Polinsky & Steven Shavell eds., forthcoming) (manuscript at 3-5, available at http://Socrates.berkeley.edu/~diff>scotch/ms_06_10_2005.pdf).

²⁵ See Goldstein, *supra* note 000 at 5 (discussing product differentiation through versioning in the book publishing and motion picture industries)

²⁶ See Goldstein, *supra* note 000 at 5.

²⁷ *Shloss v. Sweeney*, 2007 U.S. Dist. LEXIS 41847 (D. Cal. 2007); R. Anthony Reese, *Public but Private: Copyright's New Unpublished Public Domain*, 85 TEX. L. REV. 585, 618 (2007). See also, D.T. Max, *The Injustice Collector*, NEW YORKER, June 19, 2006, at 34-43 (an account of Stephen Joyce’s various threats of copyright litigation).

²⁸ *Shloss v. Sweeney*, 2007 U.S. Dist. LEXIS 41847 (D. Cal. 2007).

²⁹ *Rosemont Enters., Inc. v. Random House, Inc.*, 366 F.2d 303 (2d Cir. 1966).

³⁰ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984). See also, *Warner Bros., Inc. v. American Broadcasting Cos.*, 720 F.2d 231, 240 (2d Cir. 1983) (describing the idea-expression distinction as “an effort to enable courts to adjust the tension between these competing effects of copyright protection.”)

³¹ 17 U.S.C. § 108(a).

³² 17 U.S.C. § 115.

³³ 17 U.S.C. § 107 (“the fair use of a copyrighted work ... is not an infringement of copyright”).

³⁴ See, *Report of the Register of Copyrights on the General Revision of the U.S. Copyright Law (1961)*.

Whether these activities are infringing also depends on the similarity between the protectable elements of the original work and the alleged copy and the application of any other relevant exceptions to the rights of the copyright owner. See *Computer Assocs. Int'l v. Altai, Inc.*, 982 F.2d 693 (2d Cir. 1992) (abstraction,

doctrine,³⁵ however, the exact nature of fair use remains elusive and resists straight forward definition.³⁶

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.³⁷

The fair use doctrine is particularly important in situations where the costs of obtaining permission outweigh the benefits of the use. The doctrine also plays a mediating role in situations where the copyright owner withholds permission for reasons that we as a society find unacceptable. For example, a copyright owner usually cannot deny permission to copy in order to stifle parody, criticism or social debate.³⁸

We typically think of market transactions as involving the meeting of a willing seller and a willing buyer who agree to a voluntary exchange.³⁹ Conversely, adopting Wendy Gordon's influential market failure framework,⁴⁰ most applications of the fair use doctrine relate to the failure of one or more of these elements. Either there simply is no meeting because the costs of exchange are too high,⁴¹ or the seller or the buyer is unwilling for a variety of reasons.⁴² A court will find fair use where the unwillingness of the seller is deemed illegitimate.⁴³ Similarly, a court will also find fair use where the

filtration comparison test for substantial similarity); *see also* 17 U.S.C. § 108 – 122 (limitations on exclusive rights in addition to fair use).

³⁵ *See e.g.* Ty, Inc. v. Publ'Ns Int'l, 292 F.3d 512, 518 (7th Cir. 2002) (Posner) (the fair use doctrine plays an essential role in copyright law)

³⁶ *See* Dellar v. Samuel Goldwyn, Inc., 104 F.2d 661 (2d Cir. 1939) (describing fair use as one of the most troublesome doctrines in the whole law of copyright).

³⁷ *See*, Sag, Beyond Abstraction, *supra* note 000 at 250 (criticizing doctrinal recommendations which aim to optimize copyright scope in the abstract but do not account for the effect uncertainty or strategic behavior); Gideon Parchomovsky and Kevin A. Goldman, *Fair Use Harbors*, 93 VA. L. REV. 1483 (2007) (noting that when transaction costs are high or in the presence of strategic holdouts, users will copy less protected expression than they are legally entitled to or refrain from using copyrighted expression altogether).

³⁸ *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994); *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001) (“The Copyright Clause and the First Amendment, while intuitively in conflict, were drafted to work together to prevent censorship.”)

³⁹ [cite to basic definitions of market transaction]

⁴⁰ 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).

⁴¹ Applications of Gordon's market failure framework have largely concentrated on the role of transaction costs in justifying fair use. *See e.g.*, William M. Landes & Richard A. Posner, *An Economic Analysis of Copyright Law*, 18 J. LEGAL STUD. 325, 359-60 (1989).

⁴² Gordon herself does not limit market failure to transaction costs. Her initial formulation also addressed externalities information asymmetries and non-economic motivations as potential causes of market failure. *See* Gordon, *supra* note 000 at 1607-15.

⁴³ *See e.g.* *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596, 607 (9th Cir. 2000) (holding that the plaintiff's desire to control the market for gaming platforms was “understandable” but that “copyright law does not confer such a monopoly”); *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257

buyer's unwillingness to pay is deemed legitimate.⁴⁴ In all three scenarios, some assessment of the underlying market conditions is required.

One of the purposes of this Article is to demonstrate that the technological and social changes of the Internet era are also changes to the underlying structure of copyright markets. Viewed as such, it is not surprising that changed circumstances require re-examination of existing doctrines that are themselves contingent on market conditions. In short, to the extent that copyright markets function differently in the context of new technology, they may also require very different applications of copyright doctrines to accommodate them.

This is not to suggest that copyright has no application online, or that the substantial body of copyright law that has developed over the past two centuries should be discarded. On the contrary, many of the principles and distinctions derived from pre-internet cases are equally applicable online. In many cases, the mere fact that copying took place online is of little or no relevance. For example, the legality of the 383,000 self-described parody videos hosted on YouTube⁴⁵ will largely depend on the amount of copyrighted material taken by the parodist,⁴⁶ and on whether the work is reasonably perceived as a genuine parody or critique of the copyright owner's work.⁴⁷ These questions are the same now as they were in 1994 when the Supreme Court last addressed the issue.⁴⁸

However, this apparent continuity should not blind us to significant underlying changes. The advent of discussion boards, blogs, social networking sites, photo sharing sites and other user-generated content has made the fair use doctrine more important to more people than ever before. The fair use doctrine has become increasingly significant to the general public because the digital technology and the Internet have enabled new forums and new ways to interact with copyrighted material which involve copying.⁴⁹ For example, whereas posting the contents of newspaper article on an Internet discussion board or blog raises the specter of copyright infringement; cutting out the same article and sending it a friend through the U.S. postal service does not. Another significant reason is that sharing user-generated content online makes it much easier for copyright owners to detect any given infringement.⁵⁰ The scope of fair use in relation to YouTube

(11th Cir. 2001) ("The Copyright Clause and the First Amendment, while intuitively in conflict, were drafted to work together to prevent censorship.")

⁴⁴ See, e.g., *Shell v. City of Radford*, 351 F. Supp. 2d 510 (W.D. Va. 2005) (finding that a city police department was entitled, without authorization, to display a plaintiff's copyrighted photographs in the course of its criminal investigation of said plaintiff).

⁴⁵ YouTube search, search query=parody, performed on February 20, 2008 at 2:56 EST.

⁴⁶ *Berlin v. E. C. Publications, Inc.* 329 F.2d 541 (2d Cir. 1964).

⁴⁷ *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994); *Dr. Seuss Enters., L.P. v. Penguin Books USA, Inc.*, 109 F.3d 1394, 1399 (9th Cir. 1997); *SunTrust Bank v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001).

⁴⁸ *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994).

⁴⁹ [Cite to early examples of newsgroup v. newspaper litigation such as *L.A. Times v. Free Republic*, 2000 U.S. Dist. LEXIS 5669 (D. Cal. 2000)]

⁵⁰ See, Molly Shaffer Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. 1535, 1539 (2005) ("The new tools of digital distribution give even amateur artists - without much money, without investors, and without plans to use copyright to make a profit from their work - enough communicative

videos, blogs and other forms of user-generated content is an important issue, but it is not the focus of this Article. Instead, this Article explores a different set of issues that augers a more fundamental change in the way we think about copyright and fair use. Specifically, this Article addresses the operation of the fair use doctrine in relation to copy-reliant technologies such as Internet search engines, electronic archives, plagiarism detection software and other applications which rely on copying expressive works for non-expressive ends.

C. Copy-Reliant Technologies

Much of the discussion that follows concentrates on various forms of search technology, for two reasons. First, the operation of search technology is clearly a significant public policy issue.⁵¹ The Internet has become an integral part of modern existence. For many, it is the dominant medium of communication, research, entertainment, social interaction and political participation. Arguably, the Internet today is as significant as traditional broadcast media ever was. Second, search technology generally raises three significant issues that apply to copy-reliant technologies more generally: the copying of expressive works for non-expressive uses; the potential for high transaction costs associated with copy-reliant technologies; and the role of opt-out mechanisms in addressing these transaction costs problems.

Search technology drives the Internet.⁵² Without reliable search technology the world's 1.2 billion Internet users⁵³ would have very little hope of finding what they were looking for among the hundreds of billions of individual web pages that currently comprise the World Wide Web.⁵⁴ Internet search engines allow users to sift through massive amount of information to find the specific information that is of particular interest to them. Indeed, Web content that is not indexed ceases to exist for all practical purposes.⁵⁵ Without

potential that they need worry about copyright's costs when they build upon copyrighted works. Their activities are suddenly on copyright's radar screen.”)

⁵¹ Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000). [Additional references]

⁵² Although the network of interlinked web pages and resources that comprises the World Wide Web is commonly referred to as “the Internet”, the Internet is also comprised of many other key elements, such as electronic mail, online chat services, and various file transfer networks. Technically, the Internet is a physical network, comprised of millions of household, local, academic, business, and government networks, all linked together by copper wires, fiber-optic cables and wireless connections. Whereas the Internet is defined by its physical features, the Web (or World Wide Web) is defined by its content: the Web is a collection of interconnected documents and other resources, linked by hyperlinks and URLs. Following common usage, references herein to the Internet encompass both the physical layer and the content layer. [cite to WC3 definitions, others]

⁵³ Internet World Stats, *World Internet Usage And Population Statistics*, available at <http://www.Internetworldstats.com/stats.htm>.

⁵⁴ See *infra* note 113 and accompanying text.

⁵⁵ Lucas Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169 (2000) (available at <http://www.nyu.edu/projects/nissenbaum/papers/searchengines.pdf>) (“Without much exaggeration one could say that to exist is to be indexed by a search engine.”)

search engines, most content on the Internet would simply never be found and thus, in most cases, probably never created or posted in the first place.⁵⁶

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 comprehend or enjoy copyrighted works in the way that humans do – they simply process them as raw materials that feed various algorithms and indices. The raw data underpinning modern search engines is gathered initially by automated software agents that continuously “crawl” across the Internet copying web pages that are later analyzed and cataloged. As part of this process, search engines both index and copy each web page they find, and store the HTML code from those pages in a temporary repository called a cache.⁵⁷

Search engine users are directed to particular websites based on the relationship of their search term to the index of pages maintained by the search engine provider.⁵⁸ Like most search engines, Google displays its search results in a menu which features both the title of the relevant webpage and a short “snippet” or extract from the targeted web page. The snippet is followed by both a hyperlink to the actual Web page and another link to the cached version of the page stored on Google’s servers.⁵⁹ 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.

Search engines interact with copyrighted works routinely, automatically and indiscriminately. As such they are vulnerable to claims of copyright infringement at every key stage of their operation. Search engines routinely copy web pages as part of the indexing process. They also display fragments of copyrighted works as search results. Furthermore, search engines such as Google also store those pages so that they can be retrieved by Internet users at a later date.

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

⁵⁶ See generally, JOHN BATTELLE, *THE SEARCH: HOW GOOGLE AND ITS RIVALS REWROTE THE RULES OF BUSINESS AND TRANSFORMED OUR CULTURE* (2005) (an account of the history and significance of Internet search).

⁵⁷ The three most popular search engines are currently Google, Yahoo!, and MSN. Each of these displays “cached” links with their search results. See, Enid Burns, *U.S. Search Engine Rankings, December 2007*, Search Engine Watch, Feb 5, 2008. (estimating Google’s market share at 58.4 percent, Yahoo’s at 22.9 percent and Microsoft’s at 9.8 percent) Available, <http://searchenginewatch.com/showPage.html?page=3628341>.

⁵⁸ The Google search engine’s particular method of determining this relationship is considered to be a vital aspect of its popularity and market dominance. Google’s “page-rank” technology is the subject of U.S. Patent 6,285,999.

⁵⁹ Yahoo and MSN are similar in this respect.

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.⁶⁰ 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.⁶¹ Like all other major search engines, the Google search engine uses automated software agents to continuously “crawl” across the Internet, to locate, analyze and catalog available Web pages.⁶² As part of this process, search engines both index and copy each web page they find, and store the HTML code from those pages in a temporary repository called a cache.⁶³

Search engines allow users to retrieve items from the cache for two main reasons. First, cached links enable Internet users to detect changes that have been made to a particular web page over time.⁶⁴ The differences such comparisons reveal can have important political, educational, and legal ramifications.⁶⁵ Second, the availability of cached links enables users to understand why a seemingly irrelevant web page was indicated as responsive to their original query.⁶⁶

Although these functions relate to the copyrighted expression contained in the original website, they do not replicate the expressive function of the original. By definition, the use of a cached version of a web-page to detect changes is a use that could not be served by the original copyrighted work alone.⁶⁷ Likewise, referring to the cache to better understand the relationship between a particular page and a particular search term is also a use that could not be served by the original copyrighted work alone.

Although it was not expressed in this terminology, the essence of the court’s finding in *Field* was that the non-expressive use of the works in the cache did not interfere with the rights accorded to Field as an author. This conclusion may have been underscored by the court’s conclusion that Field had created the works in question by solely for the purposes of his lawsuit.⁶⁸ The court’s conclusion was also based in part on its assessment that

⁶⁰ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).

⁶¹ *Id.*

⁶² Some search engines operate through directories which depend human review and/or human input in the form of directory descriptions. The Google search engine uses an automated program called the “Googlebot” which is entirely automated. In contrast, Microsoft’s search engine combines human-powered listings from LookSmart with crawler-based results provided by Inktomi. Yahoo! also uses a hybrid model. See Danny Sullivan, *How Search Engines Work*, Search Engine Watch, Mar 14, 2007 (<http://searchenginewatch.com/showPage.html?page=2168031>).

⁶³ The three most popular search engines are currently Google, Yahoo!, and MSN. Each of these displays “cached” links with their search results. See, Enid Burns, *U.S. Search Engine Rankings, December 2007*, Search Engine Watch, Feb 5, 2008. (estimating Google’s market share at 58.4 percent, Yahoo’s at 22.9 percent and Microsoft’s at 9.8 percent) Available, <http://searchenginewatch.com/showPage.html?page=3628341>.

⁶⁴ [Add footnote discussion of the Internet Archive and related litigation]

⁶⁵ *Field v. Google Inc.*, 412 F. Supp. 2d 1106, 1118 (D. Nev. 2006).

⁶⁶ *Id.*

⁶⁷ *Id.* at 1119.

⁶⁸ See *Id.* at 1114 (“Field created the 51 works at issue in this lawsuit Over a three-day period in January 2004 specifically for the purposes of filing a lawsuit”)

Field could have easily opted out of inclusion in the Google search engine if he had chosen to do so.⁶⁹ The significance of both of these rationales will be addressed below.

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

Google’s search engine technology was also at the core of another recent case, *Perfect 10 v. Amazon*.⁷⁰ Perfect 10 is in the business of producing and selling copyrighted images of nude models. It does so primarily through a subscription website which sits behind a password protected paywall.⁷¹ The material on Perfect 10’s own website is neither indexed nor cached by the Google search engine; however, the search engine has no mechanism to exclude images republished by third parties without P10’s authorization.⁷²

To understand 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. The basic operation of image search is similar to text search: the engine employs automated software agents which crawl the Internet and copy the contents of HTML web pages. The image search engine then indexes and analyses these pages in response to search queries from its users. However, image search differs from text search in that the search engine software does not actually recognize and index the images themselves; instead the search engine identifies text in its database associated with other objects identified as images. If text associated with an image file is responsive to a user’s query, the search engine will display a small lower resolution “thumbnail” of the image in the search results. If an Internet user selects that thumbnail, the user’s browser will be directed to retrieve the full-scale image from its original location.⁷³

One of Perfect 10’s several theories of liability was that by producing small thumbnail versions of its copyrighted pictures in the image search results, Google was copying Perfect 10’s work without its authorization.⁷⁴ The district court initially agreed with Perfect 10 on this particular theory of infringement. However, the Court of Appeals for the Ninth Circuit held that the use of thumbnail images in an image based search engine did not constitute copyright infringement because Google’s use of thumbnails was highly transformative in that the images “served a different function” unrelated to “artistic expression.”⁷⁵ As the court explained: “[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information.”⁷⁶

⁶⁹ Field v. Google Inc., 412 F. Supp. 2d 1106, 1119 (D. Nev. 2006).

⁷⁰ Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007).

⁷¹ *Id.*

⁷² *Id.* at 711.

⁷³ *Id.*

⁷⁴ *Id.*

⁷⁵ *Id.* at 721 (citation and quotation omitted). *See also*, Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003).

⁷⁶ Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 721 (9th Cir. 2007).

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

The third case-study relates to a different kind of search engine, one that is still under construction and whose very existence will be determined by legal arguments such as those addressed in this article. Google’s self appointed mission “to organize the world’s information and make it universally accessible and useful” is not limited to that which is already in digital form.⁷⁷ Likened to the Library of Alexandria,⁷⁸ 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 libraries⁷⁹ – searchable by anyone with an Internet connection.⁸⁰ To create this search engine, Google is currently in the process of digitizing vast collections of books, one page at a time.⁸¹

Google Book is designed to allow users to search inside the text of captured books and to generate a list of books relevant to the user’s search terms.⁸² 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.⁸³ In the default scenario, a user who clicks on a book title is presented with bibliographic data about the target book and a small extract or “snippet” of the relevant page containing her search terms.⁸⁴ Users are also presented with additional information about the books targeted by their search term, including links to online bookstores and links to nearby libraries where the book can be obtained.⁸⁵

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.”⁸⁶ Entering the search term generates a menu of books containing the term. Selecting any one book leads to a second set of information about the book including, snippets illustrating the relevance of the search term to the contents of the book, bibliographic information, links

⁷⁷ *Google Corporate Information: Company Overview*, at <http://www.google.com/corporate/>.

⁷⁸ 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.

⁷⁹ Bob Thompson, *Search Me?; Google Wants to Digitize Every Book. Publishers Say Read the Fine Print First*. The Washington Post, August 13, 2006. The University of Michigan Library, Harvard University Library, Stanford University Library, the University of California Library and the New York City Public Library, Oxford University Library– the later two libraries will only allow copying of works known to be in the public domain.

⁸⁰ [Collect citations on law review literature discussing Google Book]

⁸¹ For other descriptions of the Google Book project see [collect citations]

⁸² [Cite to court documents in the Authors Guild litigation]

⁸³ Google has several agreements with publishers to do just that. [Add Details] [discuss Amazon’s Search Inside feature]

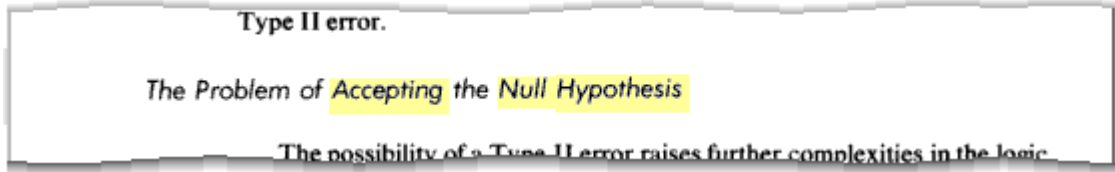
⁸⁴ Google, *What you’ll see when you search on Google Book Search*, at <http://print.google.com/googleprint/screenshots.html>, see also [Cite to court documents in the Authors Guild litigation]

⁸⁵ *Id.*

⁸⁶ This specific example is based on a search conducted by the author on February 21, 2008 at 1.57 EST using the Google Book search engine located at the URL: books.google.com.

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.com, or to find the book in a lending library. A second click generates a list of libraries ranked in order geographic proximity. A third click actually retrieves the call number from the University of Virginia Law Library. 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 149.⁸⁷

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.⁸⁸ It might be hyperbolic to suggest that “all the books in the world [will] become a single liquid fabric of interconnected words and ideas”,⁸⁹ but perhaps great advances in human knowledge deserve a little hyperbole.

Not everyone is so enamored with Google Book. Two significant lawsuits have been filed against Google in relation to the Google Book. The first is by the American Association of University Presses.⁹⁰ The second is a class action representing published authors and The Authors Guild.⁹¹ Both suits seek declaratory and injunctive relief and money damages.

The copyright challenge to Google Book is primarily focused on the way Google is building its search engine, rather than the output of the search engine *per se*.⁹² The information contained in the search results of any one Google Book search is not by itself likely not infringe the copyright of any author for two reasons. First, most of the information Google Book generates falls into the category of uncopyrightable facts about

⁸⁷ JOHN M NEALE & ROBERT M. LIEBERT, *SCIENCE AND BEHAVIOR, AN INTRODUCTION TO METHODS OF RESEARCH* (2d) 149 (1980). Google Book, search term “accepting the null hypothesis”. Search conducted on February 21, 2008 at 1.57 EST. (<http://books.google.com>).

⁸⁸ See, Edward Wyatt, *Google Adds Library Texts to Search Database*, N.Y. TIMES, Nov. 3, 2005, at C11.

⁸⁹ Kevin Kelly, *Scan This Book!*, N.Y. TIMES, May 14, 2006, §6 (Magazine), at 42.

⁹⁰ 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. [add docket citation]

⁹¹ The Authors Guild filed a lawsuit in relation to Google’s scanning and digitizing of library books on September 20, 2005. *See*, *Authors Guild v. Google*, No. 05 CV 8136 (S.D.N.Y. Sept. 20, 2005).

⁹² [cite to complaint in *Authors Guild v. Google*, No. 05 CV 8136]

books.⁹³ Second, even the snippets of material that Google directly copies from the print version a book will not amount to copyright infringement because the amounts taken are too fragmented and insignificant to meet the test of substantial similarity.⁹⁴

However, the manner in which Google is building its formidable database presents more serious copyright issues. In the same way that Internet search engines routinely, automatically and indiscriminately copy html pages as part of the indexing process, the Google Book project requires the routine, automatic and indiscriminate copying of printed library books. 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.

4. TurnItIn.com – Plagiarism Detection Software

The final case-study of copy-reliant technology is plagiarism detection software. A broad range of educational institutions have turned to technological solutions to combat the threat of plagiarism.⁹⁵ Harvard University,⁹⁶ the International Baccalaureate program⁹⁷ and thousands of high schools across the United States⁹⁸ use plagiarism detection software to detect and deter cheating by their students. Plagiarism detection services, such as Turnitin.com, detect improper and unaccredited copying in student papers by comparing new papers to an archive of material available on the Internet and to proprietary databases of previously submitted papers.⁹⁹

This technology has obvious benefits for educators and for students. However, like other copy-reliant technologies, anti-plagiarism software also has its share of critics. The Conference on College Composition and Communication, for example, warns that “such services compromise academic integrity and effective teaching” and “undermine students’ authority over the uses of their own writing.”¹⁰⁰ Ironically, although the CCCC

⁹³ See *infra* Part II-B.

⁹⁴ See *infra* Part II-B.

⁹⁵ Darby Dickerson, *Facilitated Plagiarism: The Saga Of Term-Paper Mills And The Failure Of Legislation And Litigation To Control Them*, 52 VILL. L. REV. 21 (citing various studies of academic integrity that show including a 1999 survey finding 50 percent of students admitted to Internet plagiarism). See, also Jason Johnson, *Cut-and-Paste Is a Skill, Too*, WASH. POST, March 25, 2007, pB1, (arguing that the ease of plagiarism on the Internet means that the term paper is dead). Plagiarism is “the intentional appropriation of the creative output or scholarship of another without attribution.” See, Laurie Stearns, *Copy Wrong: Plagiarism, Process, Property, and the Law*, 80 CALIF. L. REV. 513 (1992).

⁹⁶ See, *Nation In Brief*, WASH. POST, November 3, 2006.

⁹⁷ See, S. Mitra Kalita, *Schools Turn to Software to Help Stop Plagiarism*, WASH. POST, April 15, 2004, p. T4.

⁹⁸ See, Andy Dehnart, *The Web’s Plagism Police*, SALON.COM, June 14, 1999, <http://www.salon.com/tech/feature/1999/06/14/plagiarism> (reviewing several different services); Brock Read, “Anti-Cheating Crusader Vexes Some Professors”, 54 CHRON. HIGHER ED. Issue 25, page A1, February 29, 2008. Available at <http://chronicle.com/free/v54/i25/25a00101.htm>; Maria Glod, *McLean Students Sue Anti-Cheating Service; Plaintiffs Say Company’s Database of Term Papers, Essays Violates Copyright Laws*, WASH. POST, March 29, 2007, p B5.

⁹⁹ See, Turnitin website, *Proprietary Matching Technology*, <http://turnitin.com/static/plagiarism.html>.

¹⁰⁰ Conference on College Composition and Communication, *CCCC-IP Caucus Recommendations Regarding Academic Integrity and the Use of Plagiarism Detection Services*, available

embraces an author-centric view of “students’ authority over the uses of their own writing’ it also warns that plagiarism detection software fosters “an artificial view of originality and the role of imitation and borrowing in writing.”¹⁰¹

In 2006, students at McLean High School in Virginia objected when the school mandated the compulsory use of anti-plagiarism software.¹⁰² 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.¹⁰³ Two McLean High School students followed up their protest with a copyright infringement lawsuit against iParadigms, the company that provides the Turnitin.com service to their school.¹⁰⁴ 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.¹⁰⁵

Some other plagiarism detection services avoid similar disputes by allowing students to opt-out of inclusion in their reference databases. Nonetheless, like iParadigms, these services are still vulnerable to claims of copyright infringement in relation to the web-based material they incorporate into their services.¹⁰⁶

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

D. Recurring Issues in Copy-Reliant Technologies

Field, *Perfect 10*, *Google Book* and the *Turnitin.com* case each illustrate, to one degree or another, three key phenomenon pertaining to copy-reliant technologies more generally. First, copy-reliant technologies rely on copying expressive works, but they typically do so for non-expressive reasons. Second, copy-reliant technologies face the potential for

<http://ccccip.org/files/CCCC-IP-PDS-Statement-final.pdf>. See also Maria Glod, *Score One for McLean High Students; Administration Amends Anti-Cheating Policy After Protests*, WASH. POST, October 4, 2006, p. B1 (discussing the Conference on College Composition and Communication resolution).

¹⁰¹ *Id.*

¹⁰² *Id.*

¹⁰³ Glod, *McLean Students Sue Anti-Cheating Service*, *supra* note 000.

¹⁰⁴ *Id.* Two students from Arizona also joined the suit.

¹⁰⁵ *Id.* [Add cite to court filing]

¹⁰⁶ The McLean students’ copyright claim is weakened by the fact that the students themselves load their papers into the database after apparently accepting a click-wrap agreement. [see Complaint at ___] In addition to the iParadigms’ contractual argument, [see Summary Judgment motion at ___] it seems that it is the students who are actually performing the copying, not iParadigms [discuss supporting case law from the 9th circuit__]. Arguably, the students real complaint should be addressed to the ___ County School Board, which compels their use of the plagiarism detection service. The merits of such a complaint are beyond the scope of this paper.]

high transaction costs because they typically copy very large numbers of copyrighted works from a diversity of sources. Third, copy-reliant technologies often rely on technologically enabled opt-out mechanisms to address these transaction costs problems. This section explores these issues in relation to the case-studies of copy-reliant technology discussed above in order to lay the groundwork for Parts II and III which follow.

1. *Non-Expressive Uses*

The first element of note in relation to the cases profiled above is that the use being made of the copyright work is quite different to what we ordinarily think of with respect to copyright.

Copyrighted works are typically used to enable the expression contained in those works to be enjoyed, appreciated or at least comprehended by some human actor. The enjoyment of watching a film, listening to music or reading a book is derived from the creative expression contained within those objects. We chose some films, songs and books over others because of the quality of their expression. It is convenient to think of these uses as “expressive” in that they relate to, and are motivated by, the expression embedded within a copyrighted work.¹⁰⁷ Generally when a work is copied, it is copied to communicate at least some part of the work’s original expression: books are copied to be read, not to prop up uneven tables; compact discs are copied to be played, not as substitutes for Frisbees.

This observation, that expressive works are usually copied in contemplation of expressive uses, extends to partial copies as well. Because meaning is derived from context,¹⁰⁸ sampling a segment of music might change what that music expresses, but the end product is expressive in the general sense nonetheless.¹⁰⁹ Indeed, a key issue in copyright cases where part of an expressive work is incorporated into a later work is how “transformative” that appropriation is. The critical question is usually whether the defendant’s use has merely “supplanted” the original or whether it “adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message.”¹¹⁰

The copying at issue in *Field*, *Perfect 10*, *Google Book* and *Turnitin.com* is manifestly different to the usual copyright scenario. In all three cases the entirety of the copyrighted work has been copied, but the purpose of that copying not is to convey the work’s expressive qualities to the public, but rather to enable banks of microprocessors to index the content of those works and to generate meta-data about the works. The critical question these and similar cases raise is: Should a non-expressive use that nonetheless

¹⁰⁷ To the extent that this definition of the “expressive use” of a copyright work departs from a conventional understanding, the reader should understand that it is employed herein as a term of art.

¹⁰⁸ [cite to linguistics textbook, and to literature on statutory interpretation]; Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 Tex. L. Rev. 507 (2008) (criticizing the cognitive theory of trademark dilution).

¹⁰⁹ [add citations on music sampling]

¹¹⁰ *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 578-579 (1994) (citing Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1111 (1990)).

requires copying the entirety of a copyrighted work be found to infringe the exclusive rights of the copyright owner? The doctrinal implications of the non-expressive use of copyrighted works are explored in detail in Part II of this Article.

2. *Transaction costs*

The second observation to be made with respect to these cases is that the transaction costs problems they present are quite different from copyright issues in other contexts. The potential transaction costs faced by search engines and similar copy-reliant technologies are a function of (i) the sheer magnitude of the transactions required, (ii) the decentralization of the relevant actors, and (iii) the diverse nature of those actors.

The sheer scale of the Internet is truly daunting. No technology since the printing press has given rise to a proliferation of copyrighted works equivalent to the explosion of Internet content witnessed since the mid-1990's. A simple comparison helps illustrate this point. The U.S. Library of Congress is the world's largest library, with more than 134 million books, photographs, maps, works of music, manuscripts and other printed materials.¹¹¹ The volume of material available on the Internet has dwarfed this number in a very short period. There are now an estimated 1.2 billion Internet users world-wide.¹¹² 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.¹¹³ The number of copyrighted works in the Internet Archive alone exceeds the entire collection of the Library of Congress by a ratio of more than 600 to 1.

The volume of material on the Internet presents a significant transaction costs problem for Internet search technology in particular because the value of any search engine grows exponentially with its coverage. The Google Book project and plagiarism detection software confront a similar network effect.

The Internet has not only expanded information production, it has radically decentralized it as well. The Gartner consulting firm estimates that around 100 million writers actively maintained a personal website or blog in 2007.¹¹⁴ Even as the mainstream press continues to consolidate into fewer and fewer media empires, the Internet has decentralized news

¹¹¹ See The Library of Congress, *About the Library*, <http://www.loc.gov/about/facts.html>. Note that only 32 million of these items are books.

¹¹² Internet World Stats, *World Internet Usage And Population Statistics*, available at <http://www.Internetworldstats.com/stats.htm>.

¹¹³ The Internet Archive is a non-profit organization that was founded to build an Internet library, with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format. See, The Internet Archive, *About the Internet Archive*, at <http://www.archive.org/about/about.php>.

¹¹⁴ Antony Savvas, *Gartner's top 10 forecasts for 2007 and beyond*, COMPUTERWEEKLY.COM, December 15, 2006, available at <http://www.computerweekly.com/Articles/2006/12/15/220726/gartners-top-10-forecasts-for-2007-and-beyond.htm>.

production and increased both its volume and its diversity.¹¹⁵ Not only are these actors decentralized, they are also diverse.¹¹⁶ Mass-media is characterized by three factors: (i) a clear division between consumers and producers, (ii) a one-way flow of information from producers to consumers, and (iii) high start-up costs. In contrast, the hallmarks of the so-called “new media” and distributed production are: (i) the blurring of the lines between producers and consumers exemplified by blogs and social networking sites (ii) the significant degree of interaction between participants who are both producers and consumers and (iii) low monetary costs, at least for the majority of participants.¹¹⁷

The Internet has complicated the economics of copyright by expanding the range of viable information production strategies. The proliferation of content producers and their heterogeneity is no doubt beneficial, but it presents copy-reliant technologies with a difficult set of transaction costs problems if they are to clear rights before unleashing their automated processes. The doctrinal implications of the potentially high transaction costs associated non-expressive use of copyrighted works by copy-reliant technologies are explored in detail in Part III, in conjunction with the trend toward technologically enabled opt-out mechanisms discussed below.

3. *Private Ordering Through Opt-Outs*

The third feature of note with respect to the four case-studies surveyed above is the role played by default rules and opt-out mechanisms in reducing transaction costs. In other intellectual property contexts, transaction costs problems are often addressed through collective rights management. Collective management is the exercise of copyright and related rights by organizations acting in the interest and on behalf of the owners of rights.¹¹⁸ In cases where the individual management of rights is impossible or impractical, copyright owners rely on collecting societies such as ASCAP to monitor the use of their works and negotiate licenses.¹¹⁹ However, collective rights management is unlikely to be of considerable assistance in relation to search and other copying-reliant technologies because of the scale, decentralization and heterogeneity of Internet content. Nonetheless, Internet entrepreneurs have found other ways to address transaction costs. The primary mechanism for reducing transaction costs on the Internet has been the combination of well understood default rules and opt-out mechanisms.

The Internet is an open system which allows any end-point to communicate with any other end-point through a set of standard protocols.¹²⁰ The architecture of the Internet thus embeds a default rule of unrestricted access. This default requires anyone who does

¹¹⁵ 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)

¹¹⁶ *Id.* (describing information production in a networked information economy and its inclusion of a broader range of participants).

¹¹⁷ *Id.* [add other cites on the effect of new media, web 2.0 and social networking]

¹¹⁸ 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.

¹¹⁹ Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 Cal. L. Rev. 1293, 1293-94 (1996).

¹²⁰ JACK GOLDSMITH & TIM WU, *WHO CONTROLS THE INTERNET?* 23 (2006).

not wish their material to be available to affirmatively opt-out of the open Internet. Website owners remain free to restrict access by blocking unwanted access by specific IP addresses, or by requiring a user account and/or password. Website owners can also control how search engines interact with their copyrighted material by employing a technological device known as the Robots Exclusion Protocol.¹²¹ 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. The Internet's open access 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.¹²² The open, minimalist and neutral design of TCP/IP has enabled an unparalleled diversity of social and technological innovations.¹²³ Open systems and end-to-end architecture were fundamental early Internet technical standards, thus the default of open access continues, in part, simply as a result of path dependence. Second, the norm of openness also continues to flourish because it reflects the preferences of the majority of Internet users.¹²⁴ Most people want their websites to be seen and their emails to be received. Third, those with minority preferences generally have no quarrel with the default of open access; they simply opt-out of the default as it suits them.

This third point requires some elaboration. Every major Internet search engine relies on a technical device called 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.¹²⁵ Adding a robots.txt file a web site is fairly trivial and the feature is automatically generated in the majority of off-the-shelf HTML authoring and web development applications, such as Dreamweaver, Microsoft FrontPage and its successor programs Microsoft Expression Web and Sharepoint Designer.¹²⁶

¹²¹ See infra notes 000 and accompanying text.

¹²² See Vint Cerf & Robert Kahn, *A Protocol for Packet Network Intercommunication*, IEEE TRANSACTIONS ON COMMUNICATIONS, Vol. Com-22, No. 5, May 1974, pp. 637-648 (1974). (<http://www.cs.princeton.edu/courses/archive/fall06/cos561/papers/cerf74.pdf>) (the original specification of the "Transmission Control Protocol" by Vint Cerf and Bob Kahn.); Christos J. P. Moschovitis, *History of the Internet: A Chronology, 1843 to the Present* 80 (1999); Goldsmith & Wu, *supra* note 000 at 23.

¹²³ See, Goldsmith & Wu, *supra* note 000 at 22-24.

¹²⁴ [cite]

¹²⁵ See, *The Web Robots Pages, About/robots.txt*, available at <http://www.robotstxt.org/robotstxt.html>. The original Robots Exclusion Protocol was set out in 1994 by Martijn Koster. See Martijn Koster, *A Standard for Robot Exclusion* (1994), available at <http://www.robotstxt.org/orig.html>. See also, Martijn Koster, *A Method for Web Robots Control* (1996), available at <http://www.robotstxt.org/norobots-rfc.txt> (a draft internet specification of the Robots Exclusion Protocol); WC3 Recommendation, HTML 4.01 Specification, Appendix B.4.1, available at <http://www.w3.org/TR/html4/appendix/notes.html#h-B.4.1.1>.

¹²⁶ For example, to disallow all robots from a website simply requires two lines of code – "User-Agent: *" and "Disallow: /" – placed at the root level, i.e. <http://www.example.com/robots.txt>. To block a particular directory, change the second line to "Disallow: /nameofdirectory/". For further details consult Google's Webmaster Help Center under the heading, "How do I create a robots.txt file?" available at <http://www.google.com/support/webmasters/bin/answer.py?answer=40362>.

Simply including a robots.txt file does guarantee that it will be respected, however mainstream search engines such as Google, Yahoo and Microsoft all comply with the instructions contained in robots.txt files. The Google search engine, for example, is designed to allow site owners to prevent individual pages, sections of a web-site or an entire web-site from being indexed. In the event that content has already been indexed – i.e. if the site controller has failed to activate the robots exclusion standard – and the web site owner changes its preferences, the Google search engine will remove this content from the cache once the robots exclusion standard is activated.¹²⁷ Google's implementation of the Robots Exclusion Protocol is also highly customizable: among other things, site owners can also remove the snippets and/or images that appear below their page's title in Google search results.¹²⁸

The default rules and opt-out mechanisms which govern the Internet reflect a general consensus on the norm of open access. One of the reasons that consensus endures is that those with minority preferences can easily opt-out of the default rules at very low cost. *Field, Perfect 10* and *Google Book* are interesting in part because the copyright owners in these cases chose to object to the default rule instead of simply opting out. Part III of this Article studies the doctrinal implications of high transaction costs in relation to copy-reliant technologies and the use of opt-out mechanisms to resolve those transaction costs.

The purpose of this Part was to introduce the concept of copy-reliant technologies in the context of the general relationship between technology, market structure and copyright law. As the four case-studies illustrate, copy-reliant technologies raise significant and novel questions in relation to the non-expressive use of copyrighted works and the role of transaction costs and private ordering through opt-outs. Part II, which follows, explores the doctrinal implications of the non-expressive use of copyrighted works in more detail. It demonstrates that extending the rights of copyright owners to encompass non-expressive uses of their works by copy-reliant technologies would constitute a significant departure from existing copyright principles. It also suggests that the most appropriate way for courts to apply this principle is via the fair use doctrine. Part III then addresses the doctrinal implications of high transaction costs in relation to copy-reliant technologies and the use of opt-out mechanisms to resolve those transaction costs.

II. THE DOCTRINAL IMPLICATIONS OF NON-EXPRESSIVE USE

This Part addresses the a central copyright issue in relation to copy-reliant technologies, whether copying for a non-expressive purpose that nonetheless requires copying the entirety of a copyright work should be found to infringe the exclusive rights of the copyright owner.

¹²⁷ Google, *How can I prevent content from being indexed or remove content from Google's index?* Available at <http://www.google.com/support/webmasters/bin/answer.py?answer=35301&topic=8459>.

¹²⁸ *Id.*

A. The Non-Expressive Use of Copyrighted Works by Copy-Reliant Technologies

The cases reviewed in Part I, *Field*, *Perfect10*, *Google Book*, and *Turnitin.com* all involve the apparent paradox of a non-expressive use of copyright work that nonetheless requires copying the entirety of the work.

1. *Field v. Google*

Field v. Google highlights the significance of non-expressive uses of copyrighted works in the context of copy-reliant technologies. Field's claim was that Google had infringed his copyright by allowing Internet users to access his copyrighted works by retrieving them from the search engine cache.¹²⁹ Without using this terminology, the court found the non-expressive nature of Google's use of the plaintiff's works significant in a number of respects.

First, the court held that Google was not liable for copyright infringement because it was not copying in the relevant sense. Retrieving html pages from the cache was an automated process initiated by the search engine user, not by Google.¹³⁰ The court reasoned that when a user requests a web page contained in the Google cache by clicking on a cached link, it is the user, not Google, who creates and downloads a copy of the page.¹³¹ This finding would not apply to the initial copying performed by Google's automated software agents because it is Google who initiates the actions of those agents, but surprisingly Field did *not* allege that Google committed infringement in that initial copying.¹³²

The Court found that to the extent that Google itself copied or distributed Field's copyrighted works by allowing access to them through cached links, Google had engaged in a fair use of those copyrighted works. The court's fair use analysis relies heavily on the differences between Google's use of the works and any expressive or artistic value that Field's work might have otherwise had. Although the court does not employ the terminology applied in this Article, it is clear that what made Google's use different in the relevant sense was the fact that it was non-expressive.

The court gave a number of reasons why Google's use of the works in the form of cached links did not serve the same function as the original works. Primarily, the copy-reliant technology noted that cached links enable Internet users to detect changes that have been made to a particular web page over time.¹³³ The differences such comparisons reveal can have important political, educational, and legal ramifications.¹³⁴ As the court notes: "by

¹²⁹ *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006).

¹³⁰ *Id.* at 1115. Citing *Religious Tech. Ctr v. Netcom On-Line Commc'n Servs., Inc.*, 907 F. Supp. 1361, 1369-70 (N.D. Cal. 1995); and *CoStar Group, Inc. v. LoopNet, Inc.*, 373 F.3d 544, 555 (4th Cir. 2004).

¹³¹ *Id.* at 1115. (holding that the automated, non-volitional conduct by Google in response to a user's request does not constitute direct infringement under the Copyright Act).

¹³² *Id.* at 1115.

¹³³ [cite to cases involving the Internet Archive.]

¹³⁴ *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006).

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.”¹³⁵ In addition, the court also noted that the availability of cached links enables users to understand why a seemingly irrelevant web page was indicated as responsive to their original query.¹³⁶

Although these functions relate to the copyrighted expression contained in the original website, they do not replicate the expressive function of the original. By definition, the use of a cached version of a web-page to detect changes is a use that could not be served by the original copyrighted work alone.¹³⁷ Likewise, referring to the cache to better understand the relationship between a particular page and a particular search term is also a use that could not be served by the original copyrighted work alone.

For these reasons the court concluded that “the purpose and character” of Google’s use of the copyrighted html pages in the form of cached links was transformative in the sense that it “added something new” and did not “merely supersede the original work.”¹³⁸ The essence of the court’s finding was that the non-expressive use of the works did not interfere with the rights accorded to Field as an author.

2. *Perfect 10*

The relevant issue in the *Perfect 10* case was whether Google had violated copyright law by producing small thumbnail versions of the plaintiff’s copyrighted pictures in a menu of image search results.¹³⁹ Like the *Field* case, the copying performed by Google’s automated search engine in this case also hinged on the non-expressive nature of the copying performed by Google.

One of Perfect 10’s primary theories of liability was that by directing users to the location of the original (presumably infringing) images, the image search engine directly infringed Perfect 10’s exclusive rights to publicly display and to distribute its copyrighted images.¹⁴⁰ In relation to the full-scale images, both the district court and the Court of Appeals held that merely linking to a full-size image could not constitute a reproduction, distribution or display of that image.¹⁴¹

The reason for the court’s decision is simple – a hyperlink is not a copy, nor is it a display, nor is it a performance. A hyperlink is an entirely functional instruction that directs an Internet browser – operated by an end user – to retrieve specific material from a specific location where it is freely (if not legally) available. The copyright owner’s

¹³⁵ *Id.* at 1119.

¹³⁶ *Id.*

¹³⁷ *Id.*

¹³⁸ *Id.*

¹³⁹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 718 (9th Cir. 2007).

¹⁴⁰ See, 17 U.S.C. § 106 (3) (the right of distribution); 17 U.S.C. § 106 (5) (the right of public display for pictorial works).

¹⁴¹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 717 (9th Cir. 2007).

exclusive rights pertain to the display and distribution of “copies” of the work, not merely instructions as to where such copies might be obtained. As the court said:

The HTML merely gives the address of the image to the user's browser. The browser then interacts with the computer that stores the infringing image. It is this interaction that causes an ... image to appear on the user's computer screen. [Defendant] may facilitate the user's access to infringing images. However, such assistance raises only contributory liability issues and does not constitute direct infringement of the copyright owner's display rights.”¹⁴²

The court also expressly rejected the notion that merely making images “available” violates the copyright owner's distribution right.¹⁴³

Perfect 10's other main theory of liability was that Google's thumbnail-representations of the plaintiff's copyrighted pictures were themselves infringing copies of the plaintiff's art. The Court of Appeals held that the plaintiff had raised a *prima facie* case of infringement because the thumbnail images were actual copies of Perfect 10's copyrighted works and were literally displayed by the search engine.¹⁴⁴ However, the court found that Google's creation of thumbnail-representations did not infringe Perfect 10's rights, in large measure because the copying the plaintiff complained of was directed to a different use.¹⁴⁵

The essence of the Court of Appeals decision with respect to the thumbnails was that the use of the pictures by Google as a pointing device must be distinguished from the use of the pictures to fulfill the demand for the pictures as expressive works of themselves. The court described Google's use of thumbnails as being “highly transformative” in that the images “served a different function” unrelated to “artistic expression.”¹⁴⁶ As the court explained: “[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information.”¹⁴⁷ In other words, although the thumbnail pictures 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. Similar to the *Field* case, the court in *Perfect 10* based its decision in the differences between the plaintiff's expressive uses of the images and the search engines non-expressive use.

¹⁴² Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 717 (9th Cir. 2007).

¹⁴³ *Id.*

¹⁴⁴ *Id.* at 719.

¹⁴⁵ *Id.* at 725. (reversing the district court's ruling that the use of thumbnails was not fair use).

¹⁴⁶ *Id.* at 721 (citation and quotation omitted). *See also*, Kelly v. Arriba Soft Corp., 336 F.3d 811, 819 (9th Cir. 2003).

¹⁴⁷ *Id.* at 721.

3. *Google Book*

The distinction between expressive and non-expressive uses is also likely to play a significant role in the Google Book lawsuits currently before the courts. Although Google is copying expressive works, it does not appear to be using those works in an expressive manner.

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. In the same way that Internet search engines routinely, automatically and indiscriminately copy html pages as part of the indexing process, the Google Book requires the routine, automatic and indiscriminate copying of printed books. The object of all this indiscriminate copying is the production of metadata – thus to understand the Google Book controversy, it is first necessary to appreciate the value of metadata in the information age.

Information is only useful to the extent that it is relevant, discernable and available. There are thousands of volumes of information in even the smallest libraries; however, these dusty tombs are mere ornaments unless a user has some means to locate a particular book, or better yet a particular page, that may be of interest. This is the point where “metadata” becomes valuable. Metadata refers simply to information about information, or data about data.¹⁴⁸ The traditional (and now obsolete) library “card catalog” is an archetypal metadata repository – the card catalog contains information on the author, title and subject matter, but it does not contain the volumes themselves.¹⁴⁹ As the volume of available information increases, so to does the value of metadata used to organize, search, rank and retrieve that information. However, as discussed in more detail below, the undoubted value of certain facts does not alter their fundamental incompatibility with copyright protection.¹⁵⁰

The copyright issues relating to Google Book must be analyzed in two distinct parts: first, the intermediate copying which produces metadata; and second, the copying and displaying of fragments of books to display along with search results. Publicly displaying a copy of a library book so that users could read it would clearly be an expressive use of

¹⁴⁸ The American Library Association’s Task Force on Metadata defines Metadata as “structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities.” American Library Association, Task Force on Metadata *Summary Report*, June 1999 (<http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html>). See also Michael W. Carroll, *Open Access Publishing And The Future Of Legal Scholarship: The Movement For Open Access Law*, 10 LEWIS & CLARK L. REV. 741, 757 (2006) (discussing the significance of metadata for legal scholarship); Ganesan Shankaranarayanan & Adir Evan, *The Metadata Enigma*, 49 COMMS. ACM 88 (2006)

¹⁴⁹ IFLA Study Group on the Functional Requirements for Bibliographic Records. Functional Requirements for Bibliographic Records, *Final Report*. — München: K.G. Saur, 1998. — (UBCIM publications ; new series, vol. 19). — ISBN 3-598-11382-X; Tillett, Barbara. FRBR: A Conceptual Model for the Bibliographic Universe. Library of Congress Cataloging Distribution Service, 2004.

¹⁵⁰ *N.Y. Mercantile Exch., Inc. v. IntercontinentalExchange, Inc.*, 497 F.3d 109, 111 (2d Cir. 2007), *cert. denied* *N.Y. Mercantile Exch., Inc. v. IntercontinentalExchange, Inc.*, 2008 U.S. LEXIS 2443 (U.S., Mar. 17, 2008).

the work. If that work is copyrighted and the author does not consent, such use would also be infringing. However, this is not what Google proposes.

The information Google proposes to make available is meta-level information about expressive works, it is not a substitute for the expressive content of the works. Although producing metadata about books requires copying, this intermediate use is non-expressive in that it simply serves to provide a platform for data analysis and extraction. The data so produced does not contain any of the author's original expression. In addition, the fragments of works displayed as search results are almost certainly non-infringing because they are too fragmented, brief and insubstantial to meet the requirements of the substantial similarity test for copyright infringement.¹⁵¹

4. *Plagiarism Detection Software*

Plagiarism detection services rely on access to entire copies of student term papers and any works from which they might be copied; yet the services do not necessarily cause any of the copyrighted content they process to be displayed to or read by human end-users. Plagiarism detection software works by comparing strings of text in new works to strings of text in existing works.¹⁵² If a match is found, the software indicates as much. By itself, the report that a new work is similar to another work already in the database in no way reproduces or communicates the expressive qualities of either work. In practice, plagiarism detection providers also issue reports identifying the text allegedly copied and the source document, however, the basic matching function is performed with no communication of expression at all. Thus, in its basic function at least, anti-plagiarism software is paradigmatic non-expressive use.

B. The Implications of Copying for a Non-Expressive Use

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

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

¹⁵¹ See *infra*, Part II-B.

¹⁵² [expand description of matching software and digital fingerprinting]

copying the work. However, given the significant role of non-expressive copying in Internet search engines and other copy-reliant technologies, the legality of non-expressive copying is an issue that must now be addressed.

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

The Copyright Clause in the U.S. Constitution is expressly directed to the promotion of "the Progress of Science and useful Arts".¹⁵³ Copyright exists to encourage the creativity of authors and to promote the creation and dissemination of information.¹⁵⁴ 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."¹⁵⁵ Where that balance is struck dictates what the public can copy and what authors can control. Just as importantly, it also mediates relationships between different generations of authors: initial authors and those who build upon the works of others.¹⁵⁶ Thus, while copyright aims to give authors an incentive to create and share their works, it also strives to provide subsequent authors with sufficient "breathing space" to make their own additive contributions.¹⁵⁷ The copyright system is predicated both on the existence of certain rights to protect authors from unfair competition, and on significant gaps in those rights that give other authors freedom to breath.

Viewing copyright in terms of the communication of the expressive elements of the author's work is consistent with both economic and rights-based understandings of copyright. For the economist, copyright creates certain exclusive rights to give authors an incentive to invest in the creation of works that would otherwise be freely copied.

¹⁵³ U.S. CONST. ART. I, § 8, cl. 8.

¹⁵⁴ *Eldred v. Ashcroft*, 537 U.S. 186 (2003).

¹⁵⁵ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429 (1984).

¹⁵⁶ 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).

¹⁵⁷ *MGM Studios Inc. v. Grokster, Ltd.*, 545 U.S. 913, 933 (2005); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 479 (1984) ("The fair use doctrine must strike a balance between the dual risks created by the copyright system: on the one hand, that depriving authors of their monopoly will reduce their incentive to create, and, on the other, that granting authors a complete monopoly will reduce the creative ability of others.")

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.¹⁵⁸ The natural rights argument for copyright is primarily an extension of the Lockean Framework of labor as the basis of property ownership to intangibles.¹⁵⁹ However, this justification for property does little by itself to establish either its form or its limitations.¹⁶⁰ “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.”¹⁶¹ In either case, the guiding principle of copyright is that one should not generally be entitled to offer the author’s copyrighted expression to the public as a substitute for the work of the author.

Copyright consists of a bundle of discrete exclusive rights, such as the reproduction right, the derivative right, and the public performance and display rights.¹⁶² These rights are defined, articulated and limited by a number of initially judge-made doctrines, such as the idea-expression distinction, the threshold of substantial similarity and the fair use doctrine.¹⁶³ As this section explores in more detail below, these doctrines each limit copyright protection to the expressive aspects of original works of authorship in a way that confirms the centrality of communication to the public at the heart of copyright.

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

¹⁵⁸ *Eldred v. Ashcroft*, 537 U.S. 186, 212 (2003). See Sag, *Beyond Abstraction*, *supra* note 000 (reviewing the law and economics of copyright).

¹⁵⁹ See JOHN LOCKE, *TWO TREATISES OF GOVERNMENT* bk. II 27 (Peter Laslett ed., 1988) (As Locke famously argued, “[e]very Man has a Property in his own Person. This no Body has any Right to but himself. The Labour of his Body, and the Work of his Hands, we may say, are properly his. Whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joynted to it something that is his own, and thereby makes it his Property.”)

¹⁶⁰ Wendy J. Gordon, *A Property Right in Self-Expression: Equality and Individualism in the Natural Law of Intellectual Property*, 102 *YALE L.J.* 1533 (1993) (arguing that natural rights theory is necessarily concerned with the rights of the public as well as with the rights of those whose labors create intellectual products).

¹⁶¹ See Justin Hughes, *The Philosophy of Intellectual Property*, 77 *GEO. L.J.* 287, 330 (1988). See generally, Margaret Jane Radin, *Property and Personhood*, 34 *STAN. L. REV.* 957 (1982) (“to achieve proper self-development -- to be a *person* -- an individual needs some control over resources in the external environment.”)

¹⁶² 17 U.S.C. §§ 106(1) – (6).

¹⁶³ The idea-expression distinction and the fair use doctrine are also reflected in the sections 102(b) and 107 of the Copyright Act of 1976, however, they remain essentially common law features of the copyright system.

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

Copyright in an expressive work does not confer any exclusive rights in the facts, ideas, concepts, or discoveries contained in that work, regardless of the form in which they are described, explained or illustrated in such a work.¹⁶⁴ This principle, often simply abbreviated to the “idea-expression distinction,” is longstanding at common law and was expressly incorporated into the Copyright Act in the 1976 revision.¹⁶⁵

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.”¹⁶⁶ The distinction holds even in those unusual cases where the true value of the work lies in the methods, systems, and ideas it discloses, rather than in the way those concepts are expressed.¹⁶⁷ For example, in the *Selden* case itself, the plaintiff had developed a novel and useful method of bookkeeping, the practice of which created value regardless of how the method was communicated.¹⁶⁸ Nonetheless, the plaintiff’s copyright in his instructional material was limited to the expression of his useful methods and did not encompass those methods themselves.¹⁶⁹ Of course, in most cases, protecting the unique expression of an idea is sufficient to ensure that the author will be able to appropriate a return on her investment.

Copyright law also clearly distinguishes between facts and their expression, providing no protection for the former and only limited protection for the latter.¹⁷⁰ In *Feist v. Rural Telephone*, 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. As the Court explained facts — whether they be telephone numbers and addresses or the details of historical occurrences — are not “original” to the author.¹⁷¹ The author’s copyright is limited to her expression of those facts, not the facts themselves.¹⁷² The *Feist* Court further held that the selection and arrangement of that

¹⁶⁴ *Harper & Row, Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 547 (1985) (holding that “no author may copyright facts or ideas”); 17 U.S.C. § 102(b).

¹⁶⁵ 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.”

¹⁶⁶ *Baker v. Selden*, 101 U.S. 99, 102 (1879)

¹⁶⁷ *Id.*

¹⁶⁸ Indeed, although *Selden*’s application for a patent was rejected, the method may well have been patentable under today’s standards. See, *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*, 149 F.3d 1368, 1373 (Fed. Cir. 1998)). But see, *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. 124 (2006) (Breyer dissenting) (noting that although the Federal Circuit’s decision in *State Street* does say that a process is patentable if it produces a “useful, concrete, and tangible result,” the Supreme Court “has never made such a statement and, if taken literally, the statement would cover instances where this Court has held the contrary.”)

¹⁶⁹ *Id.*

¹⁷⁰ *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340. (holding that facts are not copyrightable and that the copyright in a factual compilation is thin).

¹⁷¹ *Id.* at 348 (“Copyright protection may extend only to those components of a work that are original to the author”).

¹⁷² *Harper & Row, Publr. v. Nation Enters.*, 471 U.S. 539, 556 (1985) (“No author may copyright his ideas or the facts he narrates.”)

information was also not copyrightable in that particular case because the organization of the telephone directory was “so mechanical or routine as to require no creativity whatsoever.”¹⁷³

Through the idea-expression distinction, copyright law protects the expressive elements of the author’s work while guaranteeing subsequent authors the necessary breathing space to make their own contributions by adding to, re-using, or re-interpreting the facts and ideas embodied in the original work. Subsequent authors may not compete with the copyright owner by offering her original expression to the public as a substitute for the copyright owner’s work, but they are free to compete with their own expression of the same facts, concepts and ideas.¹⁷⁴ Accordingly, the idea-expression distinction is a central element of the balance between the interests of authors in the exploitation of their writings and society’s competing interest in the free flow of ideas, information, and commerce.¹⁷⁵

This principle, the exclusion of facts and ideas from the ambit of copyright protection, applies with equal force to non-expressive copying in the digital age. In spite of the fact that metadata is increasingly valuable in the information age, it is no more copyrightable than it was 100 years ago. The undisputed value of individual facts, such as the title of book or its location in a library, does not change the copyright status of those facts. As a general rule, metadata is not subject to copyright protection: one can extract and reproduce facts, names and dates from a news paper article, or ideas and processes from an instructional text, without infringing the author’s copyright.¹⁷⁶ Whether congress should, or even could, alter the traditional contours of copyright by extending copyright protection to facts and ideas is a worthy topic of debate – the fact remains that as of now, it has not.¹⁷⁷

The idea-expression distinction is a central element of copyright’s balance between the interests of authors in the exploitation of their writings and society’s competing interest in the free flow of ideas, information, and commerce.¹⁷⁸ The idea-expression distinction

¹⁷³ Feist Publ’ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 380 (holding that the selection, coordination, and arrangement of Rural’s white pages did not satisfy the minimum constitutional standards for copyright protection). See also, Matthew Bender v. West Publishing . (1998) Assessment Technologies v. Wiredata (2003)

¹⁷⁴ [see landes & posner on possible economic rationales for the idea-expression distinction] [add information costs argument] [draw parrallell between the exclusion of ideas form copyright and the exclusion of natural phenomenon from patentable subject matter, see Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 548 U.S. 124 (2006) (Breyer dissenting) (summarizing applicable principles)]

¹⁷⁵ Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984). See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 240 (2d Cir. 1983) (describing the idea-expression distinction as “an effort to enable courts to adjust the tension between these competing effects of copyright protection.”)

¹⁷⁶ Feist Publ’ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991); Baker v. Selden, 101 U.S. 99 (1880).

¹⁷⁷ [Cite literature on database protection and on possible constitutional limitations in particular.]

¹⁷⁸ Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 429 (1984). See also, Warner Bros., Inc. v. American Broadcasting Cos., 720 F.2d 231, 240 (2d Cir. 1983) (describing the idea-expression distinction as “an effort to enable courts to adjust the tension between these competing effects of copyright protection.”)

limits the rights of the copyright owner to the expressive elements of the author's work: in the analog context this is achieved by simply declining to hold that the copying of facts and ideas alone amounts to infringement; but in the digital context it also requires that copying for purely non-expressive purposes, such as the automated extraction of data, should not be regarded as infringing.

2. *The Collective Work Right*

The collective work right also demonstrates that communication to the public is the touchstone of copyright infringement. The Copyright Act gives authors the exclusive right to reproduce their works in copies; however, the Act also confers a special privilege on the owners of collective works, such as magazines and newspapers, which allows them to reproduce and distribute individual contributions as part of the collective work and revisions thereof.¹⁷⁹ The collective work right creates an apparent conflict with the general reproduction right by allowing magazines and newspapers to reproduce the works of individual authors without their consent in certain circumstances. That conflict came to a head in the 2001 case of *New York Times v. Tasini*.¹⁸⁰ In that case, six freelance authors sued the a group of publishers, including the New York Times, for allowing articles written by the authors to be placed in electronic databases without the authors' consent. The publishers relied on their "privilege," contained in section 201(c) of the Copyright Act, to reproduce and distribute the freelance author's contributions as part of a revision to a collective work.¹⁸¹

The Supreme Court's resolution of the conflict between the general reproduction 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 New York Times' broad construction of its collective right, holding that because the articles in question were "presented to, and retrievable by, the user in isolation, clear of the context the original print publication" they did not qualify as part of a revision to the original collective work.¹⁸²

The defendants in *Tasini* had argued that their conversion of printed back-issues to an electronic form amounted to revision of the collective work, and was thus sheltered under § 201(c) of the Act. From the New York Times' perspective, electronic storage was no different to the conversion of newsprint to microfilm – in either case the entire issue was archived exactly as initially printed to facilitate later retrieval of specific articles. The Court, however, rejected the notion that the form of storage mattered at all, given that users did not perceive the articles as stored but only as retrieved by the New York Times'

¹⁷⁹ 17 U.S.C. § 201 (c) Contributions to collective works. Copyright in each separate contribution to a collective work is distinct from copyright in the collective work as a whole, and vests initially in the author of the contribution. In the absence of an express transfer of the copyright or of any rights under it, the owner of copyright in the collective work is presumed to have acquired only the privilege of reproducing and distributing the contribution as part of that particular collective work, any revision of that collective work, and any later collective work in the same series."

¹⁸⁰ *N.Y. Times Co. v. Tasini*, 533 U.S. 483 (2001).

¹⁸¹ *Id.*

¹⁸² *Id.* at 488.

database.¹⁸³ Unlike microfilm files, the database presented the individual articles to the user devoid of their initial context. The Court's view was that only user perception mattered and that the fact that the articles were stored in their initial sequence was irrelevant to both readers and authors alike. The Court thus held that "[i]n determining whether the Articles have been reproduced and distributed as part of a revision of the collective works in issue, we focus on the Articles as presented to, and perceptible by, the user of the Databases."¹⁸⁴

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

3. *Substantial Similarity*

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

The copyright owner's exclusive right to "reproduce the work in copies" extends to both exact and inexact reproductions.¹⁸⁵ In both cases, however, the Copyright Act leaves the threshold of reproduction – the question of how much of the copyrighted work must be copies – undefined. In cases of "non-literal infringement", i.e. where the accused work is not an exact copy of the copyright owner's work, courts assess whether the allegedly infringing work possesses a "substantial similarity" to the copyrighted work, not the use of copyrighted expression *per se*.¹⁸⁶

Determining the threshold of substantial similarity is defined in reference to the perspective of the ordinary observer.¹⁸⁷ Infringement is defined in reference to the perspective of the consuming public, because the copyright owner's "legally protected

¹⁸³ *Id.* at 504 ("The crucial fact is that the Databases, like the hypothetical library, store and retrieve articles separately within a vast domain of diverse texts. Such a storage and retrieval system effectively overrides the Authors' exclusive right to control the individual reproduction and distribution of each Article.") (citations omitted)

¹⁸⁴ *Id.* at 499.

¹⁸⁵ 17 U.S.C. § 106(1); *Nichols v. Universal Pictures Corp.*, 45 F.2d 119, 121 (2d Cir. 1930).

¹⁸⁶ See, *Laureyssens v. Idea Group, Inc.*, 964 F.2d 131, 140 (2d Cir. 1992); (To prove infringement, a "plaintiff must first show that his or her work was actually copied. . . . If actual copying is established, a plaintiff must then show that the copying amounts to an improper appropriation by demonstrating that substantial similarity to protected material exists between the two works.") See also, *Tufenkian Import/Export Ventures, Inc. v. Einstein Moomjy, Inc.*, 338 F.3d 127, 131 (2d Cir. 2003).

¹⁸⁷ See, *Shine v. Childs*, 382 F. Supp. 2d 602, 614 (S.D.N.Y. 2005) (summarizing authorities).

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.”¹⁸⁸ 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.¹⁸⁹ In *Newton v. Diamond*, for example, the plaintiff alleged that the Beastie Boys had infringed his copyright in a musical composition by including a six second sample of a licensed sound recording in their own musical creation, “Pass the Mic”.¹⁹⁰ The Beastie Boys had obtained a license with respect to the sound recording, but had not thought it necessary to seek a license from the composer given the parsimonious nature of the composition.¹⁹¹

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.¹⁹² In cases of fragmented literal similarity, courts determine whether the copying amounts to infringement “by considering the qualitative and quantitative significance of the copied portion in relation to the plaintiff’s work as a whole.”¹⁹³ 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.”¹⁹⁴

¹⁸⁸ *Arnstein v. Porter*, 154 F.2d 464, 473 (2d Cir. 1946) (footnotes omitted); See also, *Warner Bros., Inc. v. American Broadcasting Cos.*, 720 F.2d 231, 240 (2d Cir. 1983) (“By assuring the author of an original work the exclusive benefits of whatever commercial success his or her work enjoys, the law obviously promotes creativity. At the same time, it can deter the creation of new works if authors are fearful that their creations will too readily be found to be substantially similar to preexisting works.”)

¹⁸⁹ 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 other wise, however, that decision is almost certainly in error on this point. See, *Bridgeport Music, Inc. v. Dimension Films*. 410 F.3d 792 (6th Cir. 2005); 4-13 Nimmer on Copyright § 13.03.

¹⁹⁰ *Newton v. Diamond*, 388 F.3d 1189 (9th Cir. 2004).

¹⁹¹ 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.” See, *Newton v. Diamond*, 388 F.3d 1189, 1191 (9th Cir. 2004). Note that sound recordings and their underlying compositions are separate and distinct copyrighted works. 17 U.S.C. § 102(a)(2), (7)

¹⁹² 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.”)

¹⁹³ *Newton v. Diamond*, 388 F.3d 1189, 1195 (9th Cir. 2004) (citing *Worth v. Selchow & Righter Co.*, 827 F.2d 569, 570 n. 1 (9th Cir. 1987); *Jarvis v. A&M Records*, 827 F. Supp. 282, 289-90 (D.N.J. 1993); 4 Nimmer § 13.03[A][2], at 13-47 to 48).

¹⁹⁴ *Newton v. Diamond*, 388 F.3d 1189, 1195 (9th Cir. 2004).

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.¹⁹⁵ In other words, trivial copying of expressive elements is not copyright infringement because it does not interfere with the copyright owner's exclusive right to communicate her work to the public.

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

4. *Allegations of Intermediate Copying in Hollywood*

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

Meritless claims of copyright infringement are a recognized cost of doing business in Hollywood.¹⁹⁶ Some of these claims are merely opportunistic, others appear to be motivated by the plaintiff's genuine belief that all his or her own ideas are special and that there are no coincidences.¹⁹⁷ *Madrid v. Chronicle Books* is representative of the phenomenon. In that case, the author of a one-page poem about a land of monsters who are afraid of human children alleged that the Pixar film *Monsters, Inc.*, infringed her copyright.¹⁹⁸ The court, in contrast, held that the inverted plot of monsters afraid of children was generic.¹⁹⁹ Some of these cases involve similar themes,²⁰⁰ others involve

¹⁹⁵ *Newton v. Diamond*, 388 F.3d 1189, 1193 (9th Cir. 2004) ("The principle that trivial copying does not constitute actionable infringement has long been a part of copyright law."); *Id.* at 1195 ("the dispositive question is whether the copying goes to trivial or substantial elements.")

¹⁹⁶ PETER BISKIND, *DOWN AND DIRTY PICTURES: MIRAMAX, SUNDANCE, AND THE RISE OF INDEPENDENT FILM* [page cite] (2004).

¹⁹⁷ The Latin expression, *Post hoc ergo propter hoc*, (after this, therefore because of this) nicely captures the logical fallacy.

¹⁹⁸ *Madrid v. Chronicle Books*, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002).

¹⁹⁹ *Id.* Two antecedents spring immediately to mind. First, E.T. hiding in the cupboard from Elliot, *E.T.: The Extra-Terrestrial*, Amblin Entertainment (1982). Second, Max's dominion over the fearful monsters in *Where The Wild Things Are*. See, MAURICE SENDAK, *WHERE THE WILD THINGS ARE* (1963).

similar descriptive titles applied to the same general subject,²⁰¹ and several others involve a similarity discernable only to the plaintiffs themselves.²⁰²

Confronted with motions for summary judgment, plaintiffs often urge the courts to allow them to scrutinize every single draft of the defendant's screen play in the hope that some earlier version of the work will disclose a greater resemblance to their own copyrighted work than the finished film did.²⁰³ These requests are invariably denied.²⁰⁴

Courts refuse to entertain discovery with respect to drafts of a non-infringing final work precisely because infringement requires at least some potential interference with the copyright owner's expectation of exclusivity. As noted in *Davis v. United Artists*, "the ultimate test of infringement must be the film as produced and broadcast, we do not consider the preliminary scripts."²⁰⁵ Courts do not refuse to examine interim drafts

²⁰⁰ In *Litchfield v. Spielberg*, in which the author of a musical play about two aliens stranded at the north pole accused the producers of the motion picture, *E.T., – The Extra Terrestrial*. *Litchfield v. Spielberg*, 736 F.2d 1352 (9th Cir. 1984) (finding no substantial similarity between the sequences of events, mood, dialogue and characters of the two works). In *Madrid v. Chronicle Books* the author of a one-page poem about a land of monsters who are afraid of human children alleged that the Pixar film *Monsters, Inc.*, infringed her copyright, *Madrid v. Chronicle Books*, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002). See also, *Warner Bros., Inc. v. American Broadcasting Cos.*, 720 F.2d 231, 235 (2d Cir. 1983) (finding that the protagonist in the television series, *The Greatest American Hero*, was not sufficiently similar to the D.C. Comic's *Superman* character to warrant consideration of the plaintiff's copyright infringement claim by a jury.) *Id.* at 243 ("In the genre of superheroes, Hinkley follows Superman as, in the genre of detectives, Inspector Clouseau follows Sherlock Holmes.")

²⁰¹ In *Davis v. United Artists, Inc.*, the author of the 1972 Vietnam novel entitled "Coming Home" claimed copyright infringement in relation to the 1978 Vietnam film, also titled "Coming Home". *Davis v. United Artists, Inc.*, 547 F. Supp. 722 (S.D.N.Y. 1982) (finding no similarity between the two works). In *Walker v. Time Life Films, Inc.*, the author of the autobiographical policeman's tale, *Fort Apache* alleged that the Time Life film, *Fort Apache, The Bronx* amounted to copyright infringement. *Walker v. Time Life Films, Inc.*, 615 F. Supp. 430, 435 (S.D.N.Y. 1985) (holding that no reasonable observer could find substantial similarity and that any similarity that may exist is either trivial, abstract or non-protectible as a matter of law).

²⁰² For example, in *Stromback v. New Line Cinema*, the author of the dark and disturbing screen-play outline about a callous reporter who brings down a corrupt governor entitled, *The Keeper*, accused the writers of the film, *Little Nicky*, of copyright infringement. *Little Nicky* is a comedy about the Devil and three sons, one of whom has a speech impediment and is played by Adam Sandler. *Stromback v. New Line Cinema*, 384 F.3d 283 (6th Cir. 2004) (finding no similarity between the works other than at the most superficial level). Equally incomprehensible is the claim in *Flaherty v. Filardi*, in which the producers of *Bringing Down the House*, an odd couple film about a lonely tax attorney who meets a woman on the Internet who unknown to him happens to be in prison, were alleged to have infringed the copyright the screen play screenplay *Amoral Dilemma*, the rather grim story of a disaffected young Manhattan insurance attorney who knowingly corresponds with a death row prisoner. 2007 U.S. Dist. LEXIS 69202, 8-9 (S.D.N.Y. 2007).

²⁰³ *Walker v. Time Life Films, Inc.*, 615 F. Supp. 430, 434 (S.D.N.Y. 1985) (request to discover drafts denied). *Stromback v. New Line Cinema*, 384 F.3d 283 (6th Cir. 2004). *Flaherty v. Filardi*, 2007 U.S. Dist. LEXIS 69202, 8-9 (D.N.Y. 2007) (copyright claim to interim drafts of a published non-infringing final work dismissed as a matter of law).

²⁰⁴ See, e.g., *Walker v. Time Life Films, Inc.*, 615 F. Supp. 430, 434 (S.D.N.Y. 1985) (noting that courts routinely reject requests to consider earlier drafts of screenplays).

²⁰⁵ *Davis v. United Artists, Inc.*, 547 F. Supp. 722, 724 n.9 (S.D.N.Y. 1982) (citing *Fuld v. National Broadcasting Co., Inc.*, 390 F. Supp. 877, 882 n.4 (S.D.N.Y. 1975). See also, *Walker v. Time Life Films, Inc.*, 615 F. Supp. 430, 435 (S.D.N.Y. 1985) ("The Court considers the works as they were presented to the

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.”²⁰⁶

The refusal of courts to entertain copyright infringement allegations in relation to unpublished drafts and preliminary scripts demonstrates the practical importance of the principle of expressive substitution. Because the copyright owner’s rights are generally limited to the communication of their original expression to the public, a film maker is perfectly entitled to start with Jane Austen’s *Emma* and rework the plot over and over again until she comes out with *Clueless*.²⁰⁷ Intermediate scripts that never see the light of day do not communicate the author’s original expression to the public and thus cannot constitute copyright infringement.

This section has demonstrated the centrality of expressive substitution to a variety of doctrines and applications, the idea-expression distinction, substantial similarity and the collective work right, and finally, the refusal of courts to entertain infringement actions solely based on unpublished screen-play drafts. This brief review demonstrates that communication to the public is the touchstone of copyright infringement and that no copyright liability should be found without such expressive communication.

C. Doctrinal Incorporation

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

Explicit recognition of this principle of non-expressive use would significantly clarify the legal status of copy-reliant technologies such as Internet search engines, plagiarism detection software and the Google Book project. However, acknowledging the principle of non-expressive use raises the subsidiary question of how the principle should be implemented. Consistent with the traditional contours of copyright, it is certainly open to

public.”); *Madrid v. Chronicle Books, Pixar*, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002) (“Since a court considers the works as they were presented to the public, discovery in this case relating to the steps of creation involved in *Monsters, Inc.* would be pointless.”); *Stromback v. New Line Cinema*, 384 F.3d 283, 299 (6th Cir. 2004). (“In deciding infringement claims, courts have held that only the version of the alleged infringing work presented to the public should be considered.”)

²⁰⁶ *Warner Bros., Inc. v. American Broadcasting Cos.*, 720 F.2d 231, 241 (2d Cir. 1983)(citing 3 Nimmer § 13.03[B] at 13-38.1 to -38.2, *Eden Toys, Inc. v. Marshall Field & Co.*, 675 F.2d at 501; *Durham Industries, Inc. v. Tomy Corp.*, 630 F.2d at 913 & n.11.)

²⁰⁷ *Clueless* (Paramount 1995).

a court to rule that the use of a copyrighted work that does not communicate its expressive content to the public is *per se* non-infringing. However, there are also reasons why courts might hesitate to adopt such a categorical rule.

The first reason is that adopting a categorical rule that non-expressive copying is non-infringing may simply shift the focus of argument to questions of category definition and away from substantive questions. Although the principle of non-expressive use articulated in this Article is important, it is not free from ambiguity.

The archetypal non-expressive use of a copyrighted work would be using a novel as a paper-weight, or a compact-disc as a drink coaster. More significantly, the extraction of factual information – such as names dates and places – is also a non-expressive use, in that it does not relate to the expression of these facts, but to the facts themselves.²⁰⁸ Similarly, generating factual information about a work should also be categorized as a non-expressive use of the underlying work. For example, publishing the fact that the novel *Moby Dick* was written by Herman Melville in 1851 and contains the word “whale” 783 times would not infringe any copyright in the book because this information about the work is independent of the expressive value of the work.²⁰⁹ In relation to copy-reliant technologies, plagiarism detection software illustrates one of the clearest applications of the non-expressive use principle. Anti-plagiarism services copy expressive works in order to run a series of calculations and comparisons and ultimately generate metadata about the works. That data itself communicates none of the expressive content of the works.

Nonetheless, many other significant copy-reliant technologies present some degree of ambiguity as to whether they should be regarded as expressive or non-expressive. In both *Field* and *Perfect 10*, the courts found that the primary purpose of the copying at issue was non-expressive. In the *Field* case, the court held that although allowing html pages to be retrieved from the search engine cache also allowed them to be read, the primary use of the cache was non-expressive and thus non-infringing.²¹⁰ In *Perfect 10*, the thumbnail representations were clearly visible to the public, but the court found that the thumbnails did not fulfill a demand for the originals as expressive works.²¹¹

In both these cases there was at least the technical, if unlikely, possibility that the search engine copying could function as an expressive substitute for the copyright owners’ original works.²¹² Nonetheless, in both cases the courts found that the copying at issue did not fulfill a demand for the originals as expressive works.²¹³ 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

²⁰⁸ 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.”)

²⁰⁹ *Moby Dick* is in the public domain in the United States and is available at Project Gutenberg at <http://www.gutenberg.org/etext/2701>.

²¹⁰ *Field v. Google Inc.*, 412 F. Supp. 2d 1106 (D. Nev. 2006)

²¹¹ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701 (9th Cir. 2007)

²¹² *Field* at __; *Perfect 10* at __.

²¹³ *Field* at __; *Perfect 10* at __.

expressive use of copyright works given that its predominant uses – verifying the integrity of search results, checking the authenticity of documents and confirming the date that they were actually posted – were unrelated to the expressive function of the original works.²¹⁴ In *Perfect 10*, the court acknowledged the possibility that some users might see the thumbnail representations as substitutes for the originals, but, despite this possibility, the court dismissed the plaintiff’s claim of expressive substitution as speculative and unlikely.²¹⁵ In the court’s opinion, because the thumbnail-representations were used by the image search engine to show users which websites contained images relevant to their search terms, they were not substitutes for the originals.²¹⁶

Field and *Perfect 10* illustrate the context sensitivity of a finding of non-expressive use in cases where a version of the copied work is made available to the public. Google Book raises a different sort of ambiguity in relation to non-expressive use. The construction of the Google Book database involves the actual copying of millions of expressive works for an intermediate purpose that is itself entirely non-expressive. In this regard it is exactly analogous to plagiarism detection software. However, while the process of data-generation itself is not an expressive use, the search engine linked to that data does provide expressive snippets of copyrighted books to end-users in response to their search requests.

Does this mean that the intermediate copying performed by Google should be deemed to be expressive in nature? Probably not. First, most of the information produced by the Google Book search engine is metadata about the books which is not amenable to copyright protection because of the idea-expression distinction.²¹⁷ Second, even to the limited extent that the search engine displays expressive snippets of books to end-users, those snippets are too fragmented and insubstantial to amount to infringing copies of the books themselves.²¹⁸ The better view is that a later expressive use that is non-infringing does not detract from the non-expressiveness of an intermediate use. However, if the snippets displayed by Google were more extensive, the answer might be different. This context sensitivity indicates that a categorical approach to the issue of non-expressive use may prove to be unwieldy.

As this brief discussion illustrates, Internet search engines are strong candidates for non-expressive use, but the extent to which that label actually fits may depend on a detailed assessment of specific facts. For example, the claim of non-expressive use in relation to an image search engine that reproduced full scale images as opposed to thumbnails would be doubtful. The categorization of intermediate non-expressive uses that are intertwined with expressive uses is also a source of some ambiguity. Given these and other potential ambiguities, a categorical approach to the issue of non-expressive use may prove to be unwieldy.

²¹⁴ *Field, Id.*

²¹⁵ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701 (9th Cir. 2007)

²¹⁶ *Id.*

²¹⁷ See *supra* notes 000 to 000 and accompanying text for further discussion of the idea-expression distinction.

²¹⁸ See *supra* notes 000 to 000 and accompanying text for further discussion of the requirement of substantial similarity and its application to literal but fragmented copying.

The second reason that courts should hesitate before adopting a categorical rule in relation to non-expressive uses of copyrighted works is that the same considerations that would determine whether a use was non-expressive are already a significant part of fair use analysis.

D. Fair Use and Non-Expressive Use

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

The Copyright Act requires courts to consider four factors in making a fair use determination. These factors are (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.²¹⁹ In reality, the Section 107 factors are neither complete,²²⁰ nor are they individually or cumulatively determinative.²²¹ Even the notion that there are four factors is misleading: beneath the statutory factors lies an amalgamation of interconnected meta-factors, sub-factors and presumptions. The implications of non-expressive use in relation to fair use are explored below.

The "Purpose and Character" of Non-Expressive Uses

The defense of non-expressive use is perhaps most clearly relevant under the first fair use factor, "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes."²²² Indeed, recognizing that all of the copyright owner's exclusive rights are implicitly defined and limited in reference to expressive communication to the public, makes sense of both expressive and non-expressive fair uses.

²¹⁹ 17 U.S.C. § 107.

²²⁰ *Bond v. Blum*, 317 F.3d 385, 394 (4th Cir. 2003) ("These factors are not meant to be exclusive, but rather illustrative, representing only general guidance about the sorts of copying that courts and Congress most commonly have found to be fair uses.") (citations omitted); *Universal City Studios, Inc. v. Sony Corp. of America*, 480 F. Supp. 429, 448 (C.D. Cal. 1979) ("The factors are illustrative, not definitive.")

²²¹ *Sag, God in the Machine*, *supra* note 000. See also, *Madison supra* note 000 at 1564 ("[T]he facial emptiness of the statutory language means that alone, it is almost entirely useless analytically, except to the extent that it structures the collection of evidence that a court might think relevant to its decision."). David Nimmer adopts an even more cynical view. He argues that "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. At base, therefore, the four factors fail to drive the analysis, but rather serve as convenient pegs on which to hang antecedent conclusions." David Nimmer, *"Fairest of Them All" And Other Fairytales of Fair Use*, 66 *LAW & CONTEMP. PROBS.* 263 (2003).

²²² 17 U.S.C. § 107(1).

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

whether the new use merely supersedes the objects of the original creation . . . or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message; it asks, in other words, whether and to what extent the new work is “transformative”. . . . Although such transformative use is not absolutely necessary for a finding of fair use, . . . the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works.²²³

Traditionally, the concept of transformative use has been applied to new expressive uses that “provide social benefit, by shedding light on an earlier work, and, in the process, creating a new one.”²²⁴ However, there is no reason why a use must relate to the expressive appeal of a work in order to qualify as transformative. Indeed as the court in *Perfect 10* held, the non-expressive use of a work in an 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.”²²⁵ In this case it was significant that not only did the search engine provide a social benefit by incorporating an original work into a new work, the non-expressive use of the original work as an electronic reference tool made its transformative fair use character all the more apparent.²²⁶ The defendant's fair use argument was strengthened not simply by the fact that it put the images “in a different context so that they are transformed into a new creation”²²⁷ but fundamentally because this non-expressive recontextualization served an entirely different purpose from original photos.

By construction, a non-expressive copying of copyrighted works by copy-reliant technologies does not substitute for the author's original expression. Courts have recognized this principle in application in cases such as *Field* and *Perfect 10*, even if they have not explicitly framed it in these terms. Thus in the *Perfect 10* case, the court concluded that Google's use of thumbnail-representations was “highly transformative” because although “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.”²²⁸

²²³ See *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 579 (1994) (citations omitted). See also, Pierre N. Leval, *Toward a Fair Use Standard*, 103 Harv. L. Rev. 1105, 1111 (1990).

²²⁴ *Id.* at 579.

²²⁵ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 721 (9th Cir. 2007)

²²⁶ The Court of Appeals reiterated its conclusion in *Kelly* 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.” See, *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 721 (9th Cir. 2007); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818-19 (9th Cir. 2003). See also *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 22-23 (1st Cir. 2000).

²²⁷ *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 721 (9th Cir. 2007).

²²⁸ *Id.* 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. See, e.g., *Mattel, Inc. v. Walking*

In sum, just as expressive fair uses are generally so far removed from the author's original expression that they are unlikely to act as a substitute for it, non-expressive uses of should be presumed to be fair uses because, by their very nature, they do not "supersede the objects of the original creation."²²⁹

Non-Expressive Use and Commercial Fair Use

As part of their consideration of the first factor – “the purpose and character of the use” – courts are instructed to consider “whether such use is of a commercial nature or is for nonprofit educational purposes.”²³⁰ 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.²³¹ 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. If a use is non-expressive, its status as commercial or non-commercial is irrelevant because non-expressive uses do not substitute for the author's original expression.

In both *Sony* and *Harper & Row*, the Supreme Court indicated that commercial uses are disfavored under the fair use doctrine. Writing for the majority in *Sony*, Justice Stevens suggested that: “If the Betamax were used to make copies for a commercial or profit-making purpose, such use would presumptively be unfair. The contrary presumption is appropriate here, however, because the District Court's findings plainly establish that time-shifting for private home use must be characterized as a noncommercial, nonprofit activity.”²³² 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.”²³³

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

Mountain Prods., 353 F.3d 792, 796-98, 800-06 (concluding that photos parodying Barbie by depicting “nude Barbie dolls juxtaposed with vintage kitchen appliances” was a fair use); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 609-610 (2d Cir. 2006) (use of promotional posters in a rock-biography was “a purpose separate and distinct from the original artistic and promotional purpose for which the images were created”).

²²⁹ *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 583 (1994).

²³⁰ 17 U.S.C. § 107(1).

²³¹ Indeed, the Ninth Circuit's approach to commerciality in *Napster* defines the concept exclusively in terms of market substitution. *A&M Records v. Napster, Inc.*, 239 F.3d 1004, 1015 (9th Cir. 2001) (holding that “commercial use is demonstrated by a showing that repeated and exploitative unauthorized copies of copyrighted works were made to save the expense of purchasing authorized copies.”)

²³² *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449 (1984).

²³³ *Harper & Row, Publs. v. Nation Enters.*, 471 U.S. 539, 562 (1985).

Court has reaffirmed most recently in *Eldred*, copyright promotes the creation and publication of free expression “by establishing a marketable right to the use of one’s expression.”²³⁴ 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.”²³⁵ The virtues of creative production freed from the shackles of patronage and direct government control apply equally to all forms of private production, regardless whether they rely on the fair use doctrine or not. Thus, the economic and political logic of copyright is inconsistent with placing special burdens on the private sector for no other reason than its pursuit of profit.

Noncommercial uses may have other characteristics, such as a greater degree of spillovers, which justify fair use, but there are no inherent differences between the uses of commercial and non-commercial actors.²³⁶ In a modern free-market economy, most copyright works of interest to the public at large are created by private commercial actors. Newspapers, television broadcasts and Internet search engines are predominantly commercial, and even though schools and universities are often operated by “not for profit” corporations, they are still commercial in the sense that they operate on a fee for service basis. As the Supreme Court recognized in *Campbell*:

If, indeed, commerciality carried presumptive force against a finding of fairness, the presumption would swallow nearly all of the illustrative uses listed in the preamble paragraph of § 107, including news reporting, comment, criticism, teaching, scholarship, and research, since these activities are generally conducted for profit in this country. Congress could not have intended such a rule, which certainly is not inferable from the common-law cases, arising as they did from the world of letters in which Samuel Johnson could pronounce that “no man but a blockhead ever wrote, except for money.”²³⁷

The Court in *Campbell* rejected the notion that commerciality by itself had any “hard presumptive significance.”²³⁸ Instead, the Court adopted a sliding scale approach to commercial use, arguing that because “the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works... the more transformative the new work, the less will be the significance of other factors, like commercialism, that may weigh against a finding of fair use.”²³⁹

²³⁴ *Eldred v. Ashcroft*, 537 U.S. 186, 219 (2003) (citing *Harper & Row* 471 U.S., at 558).

²³⁵ Neil Weinstock Netanel, *Copyright and a Democratic Civil Society*, 106 YALE L.J. 283, 288 (1996).

²³⁶ See generally, Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257, 261 (2007); Lydia Pallas Loren, *Redefining the Market Failure Approach to Fair Use in an Era of Copyright Permission Systems*, 5 J. INTELL. PROP. L. 1, 51-53 (1997)

²³⁷ *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 584 (1994) (citing 3 BOSWELL’S LIFE OF JOHNSON 19 (G. Hill ed. 1934, other citations omitted).

²³⁸ *Id.* at 585.

²³⁹ *Id.* at 579. As Barton Beebe notes, while commentators have assumed that the commerciality presumption was finally discarded in *Campbell*, it remains a tenacious meme pervading some court decisions. See, Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U.

This sliding scale approach to commercial uses makes sense in light of the principle of expressive substitution articulated in this Article. The hallmark of transformative works protected by the fair use doctrine is that they do not substitute for the author's original expression. Rather they "add[] something new, with a further purpose or different character, altering the first with new expression, meaning, or message."²⁴⁰ Arguably, the commerciality of non-expressive uses is irrelevant by definition – non-expressive uses are incapable of substituting for the author's original expression.

There is ready support for this position in the case law to date. In *Kelly v. Arriba Soft Corp.*, an image search case preceding *Perfect 10*, the Ninth Circuit ruled that the replication of copyrighted images in thumbnails would not substitute for the full-sized images.²⁴¹ 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.²⁴² 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."²⁴³

Non-Expressive Use and "Amount and Substantiality"

Non-expressive use is also significant in terms of the third fair use factor, "the amount and substantiality of the portion used in relation to the copyrighted work as a whole."²⁴⁴ The third factor eschews mechanical quantification and recognizes that the amount of tolerable copying varies according to both the purpose of the defendant's use and the effect of that use on the copyright owner. The issue at the heart of the third factor inquiry is not simply what percentage of the copyright owner's original work has been taken, but what proportion of the work's expressive value has been appropriated. The argument made here is that because a non-expressive use does not substitute for the expressive value of the author's original expression, it should be viewed as qualitatively insignificant under the third factor, even if it requires total literal copying.

Pa. L. Rev. 549, 598 (2008). In the Fifth Circuit for example, courts tend to nonetheless give commerciality quasi presumptive force relying on old circuit court opinions that should have been displaced by the Supreme Court's more recent decision. Cases such as *Triangle Publications* which holds that "any commercial use tends to cut against a fair use defense" and *MCA, Inc. v. Wilson* which holds directs courts to "consider whether the alleged infringing use was primarily for public benefit or for private commercial gain," seem entirely out of step with the fair use doctrine in the post-*Campbell* era. See, *Triangle Publ'ns, Inc. v. Knight-Ridder Newspapers, Inc.*, 626 F.2d 1171, 1175 (5th Cir. 1980); *MCA, Inc. v. Wilson*, 677 F.2d 180 (2d Cir. 1981).

²⁴⁰ *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 579 (1994); Leval supra note 000 at 1111.

²⁴¹ *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 815 (9th Cir. 2003).

²⁴² *Perfect 10, Inc. v. Amazon.com, Inc.*, 487 F.3d 701, 724 (9th Cir. 2007).

²⁴³ *Id.* (citing *Campbell v. Acuff-Rose Music*, 510 U.S. at 591).

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

Even in the realm of expressive uses, there no linear relationship between the percentage of a work copied and its propensity to fair use. All other things being equal, the more a defendant copies, the more likely she is to interfere with the copyright owner's right to market her works to the public. Thus Napster users who trade complete copies of copyrighted music over the Internet are treated very differently from collage artists who copy only parts of works and add their own significant creative input.²⁴⁵ 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 music player.²⁴⁶ In cases relating to photography and other visual works, courts have occasionally allowed defendants to reproduce entire images where it was unlikely that any market harm would result and the complete reproduction was considered necessary for the defendant's transformative purpose.²⁴⁷

As the Supreme Court recognized in *Campbell*, "the extent of permissible copying varies with the purpose and character of the use".²⁴⁸ 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.²⁴⁹

Just as the extent of permissible copying varies according to purpose, it also varies according to effect. In *Harper & Row*, the defendant copied only a few hundred words from an entire manuscript of the biography of former President Gerald Ford, yet the Supreme Court held that this constituted a substantial taking under the third factor because The Nation had selected it quotes "precisely because they qualitatively embodied Ford's distinctive expression," taken "the most interesting and moving parts of the entire

²⁴⁵ A&M Records v. Napster, Inc., 239 F.3d 1004, 1015 (9th Cir. 2001); *Blanch v. Koons*, 2006 U.S. App. LEXIS 26786 (2d Cir. N.Y., Oct. 26, 2006).

²⁴⁶ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 449-450 (1984) ("Moreover, when one considers the nature of a televised copyrighted audiovisual work, and that time-shifting merely enables a viewer to see such a work which he had been invited to witness in its entirety free of charge, the fact that the entire work is reproduced does not have its ordinary effect of militating against a finding of fair use.") (citations omitted); *Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys.*, 180 F.3d 1072 (9th Cir. 1999). See also, *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 971 (9th Cir. 1992) ("Consumers are not invited to witness Nintendo's audiovisual displays free of charge, but, once they have paid to do so, the fact that the derivative works created by the Game Genie are comprised almost entirely of Nintendo's copyrighted displays does not militate against a finding of fair use.")

²⁴⁷ See, *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003) (concluding that images used for a search engine database are necessarily copied in their entirety for the purpose of recognition); *Nunez v. Caribbean Int'l News Corp.*, 235 F.3d 18, 24 (1st Cir. 2000) (concluding that to copy any less than the entire image would have made the picture useless to the story); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 613 (2d Cir. 2006) (holding that total copying "does not necessarily weigh against fair use because copying the entirety of a work is sometimes necessary to make a fair use of the image"); *Mattel Inc. v. Walking Mt. Prods.*, 353 F.3d 792, 803 n.8 (9th Cir. 2003) (holding that "entire verbatim reproductions are justifiable where the purpose of the work differs from the original").

²⁴⁸ *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 587 (1994).

²⁴⁹ *Id.*

manuscript,” and structured its article around these quoted excerpts.²⁵⁰ The Court’s finding in relation to the third factor rests on the fact that The Nation had taken essentially the heart of the book’s expressive value.²⁵¹ The third factor does not rely on mechanical quantification of the amount of the original work used; it asks courts to assess how much of the value of the original work is present in the later use.²⁵²

The extent to which a use is non-expressive plays a vital role in the assessment of the third fair use factor. A use that is non-expressive does not substitute for the expressive value of the author’s original expression; accordingly, a non-expressive use should be viewed as qualitatively insignificant under the third statutory factor, even if it requires total literal copying.

Again, existing case law is consistent with this proposition. In the *Bill Graham Archives* case, for example, the court found that the reproduction of reduced scale versions of a sequence of concert posters incorporated into a biographical timeline constituted a fair use.²⁵³ The key question in this case, as in other fair use cases, was not simply what percentage of the copyright owner’s original work was reproduced, but what proportion of the expressive value of the work was taken. The court held that even though the copyrighted images were copied in their entirety, the defendant had not taken a substantial part of the original expressive value, for two reasons. First, the use of the images placed in chronological order on a timeline was “transformatively different from the mere expressive use of images on concert posters or tickets.”²⁵⁴ Second, and related, the visual impact of the artistic expression in the original work was significantly limited in the reproduction because of the reduced size of the images.²⁵⁵

In *Perfect 10*, the court held that although thumbnail-representations were in one sense a literal copy of the entire photograph, their reduced size and image quality was consistent with their non-expressive use. Although the thumbnails copied the original images, their reduced size and use as pointing devices ensured that they did not substitute for the expressive value of the author’s original expression.²⁵⁶ Consistent with its earlier decision in *Kelly*, the court found that the representation of an entire photographic image was reasonable in light of the purpose of an image search engine.²⁵⁷ As the court explained, while a user can identify relevant text by seeing merely a fraction of it, recognizing images necessitates seeing a representation of the complete image.²⁵⁸ In *Perfect 10*, as in

²⁵⁰ Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 U.S. 539, 565 (1985).

²⁵¹ *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.”)

²⁵² Sag, God in the Machine, *supra* note 000 at 391.

²⁵³ Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605 (2d Cir. 2006)

²⁵⁴ *Id.* at 609, 612.

²⁵⁵ *Id.* at 613.

²⁵⁶ Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701, 721-722 (9th Cir. 2007) (“Here, Google uses Perfect 10’s images in a new context to serve a different purpose.”)

²⁵⁷ *Id.*

²⁵⁸ *Id.* (citing Kelly, 336 F.3d at 821.)

Kelly, the court found that the third fair use factor did not weigh in favor of either party.²⁵⁹

Likewise, in *Field*, the court found that making entire web-pages available in the search engine cache served a purpose which could not be effectively accomplished by using only portions of the Web pages. The court found that Google's non-expressive uses of the cached pages – such as verifying the authenticity of live pages and assessing the relevance of search queries – required caching complete reproductions of the plaintiff's html pages. Accordingly, the district court concluded that because “Google uses no more of the works than is necessary in allowing access to them through “Cached” links, the third fair use factor is neutral, despite the fact that Google allowed access to the entirety of Field's works.”²⁶⁰

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

The Market Effect of Non-Expressive Uses

The fourth fair use factor is “the effect of the use upon the potential market for or value of the copyrighted work.”²⁶¹ The *Harper & Row* Court described the fourth fair use factor as “undoubtedly the single most important element of fair use.”²⁶² Barton Beebe, in contrast, concludes that the fourth factor is “no factor at all.”²⁶³ As detailed below, although the fourth factor risks collapsing into circularity because everything is a potential market effect, courts have in fact avoided this outcome by applying certain limiting principles which emphasize that the copyright market is limited to expressive substitution. The logical implication of the exclusion of economic consequences that do not arise from expressive substitution is that non-expressive uses have no *cognizable* market effect under the fourth factor.

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

²⁵⁹ *Id.*; *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 821 (9th Cir. 2003)

²⁶⁰ *Field*.

²⁶¹ 17 U.S.C. § 107(1).

²⁶² *Harper & Row, Publs. v. Nation Enters.*, 471 U.S. 539, 566 (1985)

²⁶³ Beebe, *supra* note 000 at 620–621 (“Ultimately, the paradox of the fourth factor is that it is everything in the fair use test and thus nothing. To assert, as a descriptive matter, that it is the most important factor - or, as a normative matter, that it is too important is meaningless, primarily because it is no factor, no independent variable, at all. Instead, regardless of what we might hope - or fear - it would be, the actual doctrine of the fourth factor consists in practice of a few propositions of law that judges should keep in mind as they synthesize the various factual findings that they have made under the previous factors.”)

circularity.²⁶⁴ 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.”²⁶⁵ The aggregation of any harm which is likely to result from widespread use is reasonable, however the Court offers no particular reason why *all* uses should be presumed to become widespread.²⁶⁶

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.²⁶⁷ Taken at face value it becomes impossible for a defendant to prove that their particular use, if widely replicated, would not displace some potential future market in some derivative of the copyright owner’s work. As the Second Circuit noted in *Texaco* “were a court automatically to conclude in every case that potential licensing revenues were impermissibly impaired simply because the secondary user did not pay a fee for the right to engage in the use, the fourth fair use factor would *always* favor the copyright holder.”²⁶⁸

Courts avoid this potential circularity by adopting a number of limiting principles in relation to the fourth factor. First, the derivative market is limited by likelihood: “The market for potential derivative uses includes only those that creators of original works would in general develop or license others to develop.”²⁶⁹

²⁶⁴ [cite Jim Gibson’s Yale piece and Gordon’s response.] Indeed, this circle 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 effected.

²⁶⁵ *Harper & Row, Publs. v. Nation Enters.*, 471 U.S. 539, 562 (1985)

²⁶⁶ This is arguably either a distortion of the Senate Report which comments that “Isolated instances of minor infringements, *when multiplied many times*, become in the aggregate a major inroad on copyright that must be prevented.” Senate Report, at 65 (emphasis added). Note that in *Campbell* the slippery slope presumption is weakened to a matter for consideration, but still without any analysis of which uses are likely to become widespread and which are not. [cite].

²⁶⁷ 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. The Supreme Court lent credence to these kinds of arguments by remanding the case in *Campbell* to the district court to determine whether the 2 Live Crew parody had dampened the potential demand for non-parody derivatives of the original song in the rap genre, a market hitherto unexplored by the copyright owner.

²⁶⁸ 60 F.3d at 930 n.17 (citations omitted). See also, Leval, *supra* note 000, at 1124 (“[b]y definition every fair use involves some loss of royalty revenue because the secondary user has not paid royalties”); 4 Melville B. Nimmer & David Nimmer, *Nimmer on Copyright* § 13.05[A][4] (2005) (“it is a given in every fair use case that plaintiff suffers a loss of a potential market if that potential is defined as the theoretical market for licensing the very use at bar.”)

²⁶⁹ *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994). See also *Texaco*, 60 F.3d at 930 (“we look at the impact on potential licensing revenues for “traditional, reasonable, or likely to be developed markets.”) In *American Geophysical*, the majority of the Second Circuit found that photocopying academic articles by researchers in a for-profit corporation was not fair use, largely because of the availability of licensing facilitated through the Copyright Clearance Center. The majority found that through this collection rights organization, the publishers had created “a workable market for institutional users to obtain licenses for the right to produce their own copies of individual articles via photocopying.” The majority in *American*

Second, fair use cases often turn on the simple question of whether the particular market claimed by the plaintiff is one that is cognizable under copyright. This is not merely a question for the fourth factor; it permeates consideration of all of the factors. The market harms which courts refuse to recognize illustrate yet again the copyright owner's exclusive rights are limited to the communication of their original expression to the public. This principle is reflected in the seemingly unrelated cases involving parody and the reverse engineering of computer software. In both scenarios, courts exclude consideration of market effects that do not arise from expressive substitution.

As Ben Kaplan put it, a theatre review or a parody “may quite legitimately aim at *garroting the original*, destroying it commercially as well as artistically” with no liability to the copyright owner.²⁷⁰ Thus, in assessing whether a defendant has “harmed” the copyright owner under the fourth fair use factor, courts look to whether “the secondary use usurps or substitutes for the market of the original work,” rather than whether it “suppresses or even destroys the market for the original work or its potential derivatives.”²⁷¹

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

when a lethal parody, like a scathing theater review, kills demand for the original, it does not produce a harm cognizable under the Copyright Act. Because parody may quite legitimately aim at garroting the original, destroying it commercially as well as artistically, the role of the courts is to distinguish between biting criticism that merely suppresses demand and copyright infringement, which usurps it.²⁷²

Geophysical argued that its reliance on potential licensing revenues was not circular because “only an impact on potential licensing revenues for *traditional, reasonable, or likely to be developed markets* should be legally cognizable when evaluating a secondary user's effect upon the potential market for or value of the copyrighted work.” (emphasis added).

²⁷⁰ Benjamin Kaplan, *An Unhurried View of Copyright* 69 (1967) (emphasis added). See also, *Fisher v. Dees*, 794 F.2d 432, 437-438 (9th Cir. 1986) (citing Kaplan); *Castle Rock Entertainment v. Carol Publ'g Group*, 150 F.3d 132, 145 (2d Cir. 1998) (same); *New Era Publications*, 904 F.2d at 160 (same); *Campbell*, 510 U.S. at 591-92 (same).

²⁷¹ *Castle Rock Entertainment v. Carol Publ'g Group*, 150 F.3d 132, 145 (2d Cir. 1998). See also, *Campbell* (“[A] lethal parody, like a scathing theater review, kills demand for the original, [but] does not produce a harm cognizable under the Copyright Act.”); *Id.* (“the role of the courts is to distinguish between biting criticism that merely suppresses demand and copyright infringement, which usurps it.”); *Fisher v. Dees*, 794 F.2d 432, 437-438 (9th Cir. 1986) (“Biting criticism suppresses demand; copyright infringement usurps it.”); *New Era Publications*, 904 F.2d at 160 (“a critical biography serves a different function than does an authorized, favorable biography, and thus injury to the potential market for the favorable biography by the publication of the unfavorable biography does not affect application of factor four.”)

²⁷² *Campbell v. Acuff-Rose Music*, 510 U.S. 569, 591-592 (1994) (quoting in part, Benjamin Kaplan, *An Unhurried View of Copyright* 69 (1967))

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.²⁷³ Federal courts have consistently held that making unauthorized copies of a computer program as a necessary step in reverse engineering is fair use.²⁷⁴ For example, in *Sony v. Connectix*, the Ninth Circuit held that although the defendant's Virtual Game Station console directly competed with Sony in the market for Sony-compatible-gaming-platforms, the Virtual Game Station was a "legitimate competitor" in that market.²⁷⁵ 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."²⁷⁶

Both parody and reverse engineering illustrate the exclusion of market effects that do not arise from expressive substitution. This rationale is 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."²⁷⁷ Consistent with its decision in *Sega*,²⁷⁸ the Ninth Circuit held in *Connectix* that intermediate copying of software could be protected as fair use if the copying was necessary to gain access to the functional elements of the software.²⁷⁹ The court based its ruling firmly in the importance of maintaining the idea expression distinction: "We drew this distinction because the Copyright Act protects expression 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."²⁸⁰ 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.

²⁷³ Computer programs are written in source code, a human readable language, but they are typically distributed in object code which is only readable by computers. The object code distributed on a compact disc or in the memory of a video game console is protected by copyright. Yet the same object code also contains ideas and performs functions that are not entitled to copyright protection. Unlike other copyright protected works, the unprotectable elements of computer programs distributed in object code are hidden from view. With the right tools, experienced programmers can extract the unprotectable elements from object code; however these methods almost invariably require making an unauthorized copy, or multiple unauthorized copies, of the program.

²⁷⁴ [collect citations]

²⁷⁵ *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596, 607 (9th Cir. 2000); see also, *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522-23 (9th Cir. 1993).

²⁷⁶ *Sony*, 203 F.3d at 607; see also, *Sega*, 977 F.2d at 1523-24 (An attempt to monopolize the market by making it impossible for others to compete runs counter to the statutory purpose of promoting creative expression and cannot constitute a strong equitable basis for resisting the invocation of the fair use doctrine).

²⁷⁷ *Id.* at 598.

²⁷⁸ *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1993).

²⁷⁹ *Sony Computer Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596, 607 (9th Cir. 2000)

²⁸⁰ *Id.* at 603.

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

As established earlier in this Part, the exclusive rights of the copyright owner are limited to the communication of original expression to the public. Furthermore, acts of copying which do not communicate the author's original expression to the public should not be held to constitute copyright infringement. As demonstrated above, these principles are not merely compatible with the fair use doctrine; they are necessary to make sense fair use cases dealing with both expressive and non-expressive fair uses.

It is of course unrealistic to attempt to reduce the entirety of fair use jurisprudence into any one coherent principle. Nonetheless, the general proposition that acts of copying that are unlikely to substitute for the copyright owner's original expression are favored by the doctrine explains the majority of cases. Non-expressive uses should be presumed to be fair uses because, by their very nature, they do not substitute for the author's original expression. Accordingly, non-expressive use should be favored under the first, third and fourth factors – such uses are non-substitutive in 'purpose and character', appropriate a qualitatively insignificant proportion of the value of the copyright owner's original expression, and produce no cognizable market effect under the fourth factor.

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

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