PREDICTING FAIR USE:

AN EMPIRICAL STUDY OF COPYRIGHT'S FAIR USE DOCTRINE

Matthew Sag[*]

ABSTRACT

Fair use is often criticized as unpredictable and doctrinally incoherent—a conclusion which necessarily implies that the copyright system is fundamentally broken. This article confronts that critique by systematically assessing the predictability of fair use outcomes in litigation. Concentrating on characteristics of the contested use that would be apparent to litigants pre-trial, this study tests a number of doctrinal assumptions, claims and intuitions that have not, until now, been subject to empirical scrutiny. This article establishes that, based on the available evidence, the fair use doctrine is more rational and consistent than is commonly assumed. Fair use is more than simply the right to hire a lawyer and take one's chances in court.

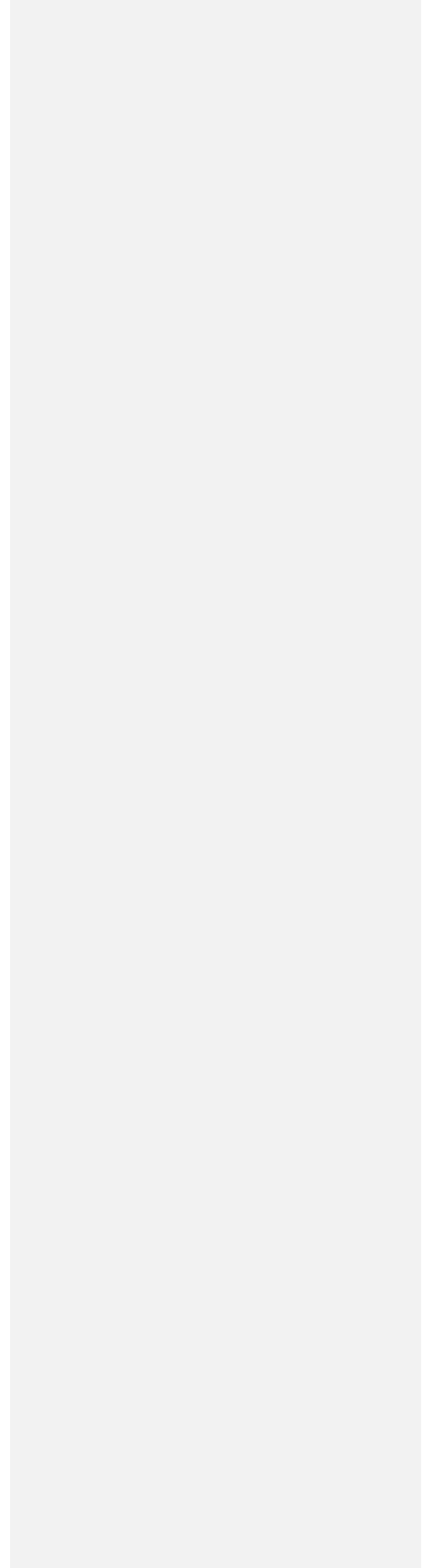[*] Associate Professor, Loyola University Chicago School of Law. © Matthew Sag 2011.

TABLE OF CONTENTS

INTRODUCTION

The concept of fair use is central to modern copyright law in the United States, however almost every aspect of the fair use doctrine is highly contested. Moreover, the application of fair use is conventionally presumed to be uncertain. This assumed incoherence and unpredictability has led many to question the value of fair use.[1] The critique that fair use is merely "a lottery argument" and amounts to no more than "the right to hire a lawyer" is deeply important.[2] In the digital age, innovation and freedom of expression increasingly require the use, reinterpretation and re-mixing of copyrighted content; the fair use doctrine is often the only aspect of copyright law that makes these activities possible. It is not simply end-users who rely on fair use: the doctrine is an essential part of the legal architecture of Internet search, Web 2.0 enterprises and social networking technologies. If fair use is arbitrary and uncertain, our copyright system is fundamentally broken. This article demonstrates that the uncertainty critique is somewhat overblown: an empirical analysis of the case law shows that, while there are many shades of gray in fair use litigation, there are also consistent patterns that can assist individuals, businesses and lawyers in assessing the merits of particular claims to fair use protection.

This article addresses the predictability of fair use by investigating the impact of identifiable characteristics on fair use outcomes. It takes specific doctrinal intuitions about fair use and evaluates the extent to which is they are supported by the data. Each one of these building blocks is interesting in isolation; the aggregation of the discrete findings addresses a question of fundamental importance—the predictability of fair use.

Fair use allows the use of copyrighted material without permission; in so doing, it sets limits on the broad rights of copyright owners to control the reproduction, distribution, performance and display of their works.[3] The importance of fair use can be seen in recent

---

[1] LAWRENCE LESSIG, FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY 187 (2004). Michael J. Madison, *A Pattern-Oriented Approach to Fair Use*, 45 WM AND MARY L. REV. 1525, 1666 (2004) ("Fair use is an ex post determination, a lottery argument offered by accused infringers forced to gamble, after the fact, that they did not need permission before.") *See also* Thomas F. Cotter, *Fair Use and Copyright Overenforcement*, 93 IOWA L. REV. 1271, 1284 (2008) (fair use remains fairly unpredictable and uncertain in many settings); Edward Lee, *Warming Up To User-Generated Content*, 2008 U. Ill. L. Rev. 1459, 1468 (expert predictions of fair use are no better than those of the person on the street, and possibly even worse); James Gibson, *Once And Future Copyright*, 81 NOTRE DAME L. REV. 167, 192 (2005) (fair use is too indeterminate a doctrine to provide a reliable touchstone for future conduct); and David Nimmer, *'Fairest of Them All' and Other Fair Use Fairy Tales*, 66 LAW & CONTEMP. PROBS. 263, 280 (2003) ("Basically, had Congress legislated a dartboard rather than the particular four fair use factors embodied in the Copyright Act, it appears that the upshot would be the same.").

[2] Madison, *supra* note 1; Lessig, *supra* note 1.

[3] Section 106 and 106A of the Copyright Act of 1976 enumerate the exclusive rights of copyright owners, and a limited set of 'moral rights' pertaining to the authors of works of visual art, however these rights are expressly subject to sections 107 through 122 of the Act. Section 107 provides that "[n]otwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, … is not an infringement of

marquee cases determining the legality of new consumer technologies,[4] user-generated content,[5] the limits of fan fiction[6] and critical reinterpretation of literature.[7] The fair use doctrine affects film production, news reporting, political communication, teaching and academic research.[8] Indeed, reliance on fair use is embedded in the core of the Internet, as it is a vital component of the search engine business model, Web 2.0 enterprises and social media.[9] Almost no one doubts that the fair use doctrine either is, or should be, very important; but judges, academics and lawyers express sharp disagreement over different conceptions of fair use and its correct application.[10]

---

copyright.) See 17 U.S.C. §§ 106, 106A and 107.

[4] *See e.g.*, Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984) (holding that the manufacturer of a video-cassette recorder was not liable for copyright infringement, in part because consumer time-shifting of broadcast television for later viewing was fair use and not copyright infringement); Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., 180 F.3d 1072 (9th Cir. 1999) (strongly suggesting that transferring music from to compact disk to mp3 for personal use would be fair use and that the manufacturers of mp3 players would not be liable for copyright infringement); and A & M Records, Inc. v. Napster, Inc., 114 F. Supp. 2d 896, 913 (N.D. Cal. 2000) (holding that peer-to-peer file-sharing of copyrighted music without authorization is not fair use).

[5] Warner Bros. Entm't Inc. v. RDR Books, 575 F. Supp. 2d 513 (S.D.N.Y. 2008) (finding that of an unauthorized Harry Potter Lexicon which contained substantial user-generated content initially derived from a fan website was, nonetheless, a substantial copy of the original J.K. Rowling novels and was not fair use).

[6] *Id.*

[7] Salinger v. Colting, 641 F. Supp. 2d 250 (S.D.N.Y., 2009) (holding that a fictional novel recounting a meeting of *Catcher in the Rye*'s Holden Caulfield at the age of 76 with a the author of that same book, J.D. Salinger, was a substantial copy of the original novel and was unlikely to constitute fair use); Salinger v. Colting, 607 F.3d 68 (2d Cir. 2010) (vacating injunction but agreeing defendants will not likely be able to make out such a defense).

[8] *See* CTR. FOR SOC. MEDIA, AM. UNIV. SCH. OF COMMC'N, DOCUMENTARY FILMMAKERS' STATEMENT OF BEST PRACTICES IN FAIR USE (2005), http://www.centerforsocialmedia.org/files/pdf/fair use final.pdf (reflecting the views of a number of filmmakers as to what constitutes reasonable application of the fair use doctrine); Kenneth D. Crews, *Fair Use and Higher Education: Are Guidelines the Answer?*, ACADEME, Nov./Dec. 1997, at 38, 38-40 (criticizing the usefulness of the "Classroom Guidelines" inadequate guidance on the application of fair use in an educational environment); *American Geophysical Union v. Texaco, Inc.,* 60 F.3d 913 (2d Cir. 1994) (finding that a researcher's unauthorized photocopying of articles from an academic periodical was not a fair use).

[9] Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 816 (9th Cir. 2003). *See generally*, Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U. L. REV. 1607 (2009) (discussing the application of the fair use doctrine to automatic copying, data-processing and other non-expressive uses).

[10] In his classic article on the fair use doctrine, Judge Pierre Leval concedes that his "own decisions had not adhered to a consistent theory, and, more importantly, that throughout the development of the fair use doctrine, courts had failed to fashion a set of governing principles or values." Pierre N. Leval, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105 (1990). *See also*, Basic Books, Inc. v. Kinko's Graphics Corp., 758 F. Supp. 1522, 1530 (S.D.N.Y. 1991) (noting that "courts and commentators disagree on the interpretation and application of the four [fair use] factors.")

Fair use is often criticized as doctrinally incoherent and unpredictable in application,[11] so much so that it has become fashionable to question the value of fair use as a guide to decision making.[12] Scholars have subjected the fair use doctrine to broad ranging theoretical scrutiny and have undertaken precise analysis of leading cases.[13] However, almost all of the literature is normative, anecdotal or confined to particular sub-topics.

There are two primary exceptions. Barton Beebe's pioneering empirical study of the contents of fair use opinions examines the language judges employ, the tests they use and judicial characterizations their own application of the four statutory fair use factors.[14] Pamela Samuelson's systematic distillation of modern fair use jurisprudence into common patterns, or "policy-relevant clusters," makes a convincing case that fair use may not be as doctrinally incoherent as many have suggested.[15] Each of these studies highlights consistent patterns in judicial analysis within the fair use decisions.

In contrast, rather than relying on backward-looking judicial reasons and explanations, this new study assess the predictability of fair use in terms of case facts which exist prior to any judicial determination. The study focuses directly on the ex ante predictability of fair use and empirically tests the significance of the characteristics of disputed uses that would be apparent to potential litigants before their cases go to trial. This approach makes it possible to test a number of doctrinal claims and intuitions that have not, until now, been subject to empirical scrutiny. No one has previously attempted to systematically assess the predictability of fair use from the perspective that matters most—the point of view of prospective litigants. This article fills that void, using statistical analysis to develop a clearer picture of the entire landscape of fair use cases.

This article proceeds as follows. Part I begins with a detailed explanation of the construction of the dataset, the key variables and some important methodological issues, including the role of selection effects. Part II reviews the extensive case law and commentary devoted to fair use and explains how abstract doctrinal questions, such as transformative use, can be reduced to empirically testable hypotheses. Part III then

---

[11] Leval, *supra* note 10 at 1106; Paul Goldstein, *Fair Use in Context* 31 COLUM. J.L. & ARTS 433 (2008) (describing fair use as "the great white whale of American copyright law, enthralling and enigmatic.); Nimmer, *supra* note 1 at 280 (noting the malleability of fair use factors); Michael W. Carroll, *Fixing Fair Use*, 85 N.C. L. REV. 1087 (2007) (demonstrating the inconsistency of the various judicial interpretations of the fair use defense).

[12] *See supra* note 1.

[13] In the last 10 years alone, at least 237 law review articles have been devoted to the issue of fair use. Based on a Lexis search of U.S. law reviews with a title containing "fair use," but not containing "trademark." In short, the relevant literature is vast. (Search date: February 23, 2011).

[14] Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549 (2008).

[15] Pamela Samuelson, *Unbundling Fair Uses*, 77 FORDHAM L. REV. 2537 (2009). Similarly, Madison, *supra* note 1, at 1623 argues that existing case law can be reframed in terms of "whether the defendants' uses of copyrighted material fell appropriately within or beyond recognized social or cultural patterns."

presents the results of the empirical tests and explores their significance. The conclusion summarizes the principal findings of this study and their implications for fair use and copyright in general.

I.   DATA & METHODOLOGY

A.   Data

This study is based on a unique dataset of more than 220 fair use cases, decided in U.S. Federal District Courts between January 1, 1978 and December 31, 2006. The dataset combines publicly available information from written opinions and court records, as well as data from other sources, such as company databases and the Martindale-Hubbell directory of attorneys and law firms.[16]

B.   Dependant Variable

This article explores the effect of a diverse range of objective factors in fair use litigation and determines whether the characteristics stressed in various judicial and academic accounts are actually important in practice. How should importance be judged? This study treats a case characteristic as important if it affects the outcome of the fair use issue in an individual case. For clarity, this variable is referred to as the successful assertion of fair use or more generally as the defendant's "win rate." This study focuses on the main battle ground of fair use litigation, United States Federal District Courts. The average defendant win rate on the issue of fair use at the district court level is 38.28%.[17] I focus on the outcome of district court litigation because successful appeals on the issue of fair use are rare.[18]

Defendant win rate is a reliable and valid measure of the significance of case characteristics.[19] It is reliable in the sense that different researchers looking at the same data will agree whether a case was won or lost. At the risk of stating the obvious, the win rate is a valid measure of importance because, across a large number of cases, factors which are important should affect the outcomes of cases.

The empirical approach presented here differs from the more impressionistic review that legal scholars typically pursue. For example, at least some authority suggests that factors such as whether a work was creative as opposed to informational, or whether the

---

[16] The data is available for replication upon request and will be permanently available on the author's website upon final publication of this article.

[17] Over the same period of time the average fair use win rate in Federal Courts of Appeal was 42.7%.

[18] Only 15 cases in this data set were subject to an appeal that produced a different result on the issue of fair use.

[19] Lee Epstein & Gary King, *The Rules of Inference*, 69 U. CHI. L. REV. 1, 82-97 (2002) (providing a detailed explanation of validity and reliability in empirical measures).

defendant was engaged in a commercial use, are important in deciding fair use cases. This study indicates that these factors do not, in fact, have any significant influence on case outcomes.[20] Because these factors do not appear to affect case outcomes in aggregate, the conclusion is drawn that they are unimportant. I do not attempt to make a subjective assessment of the importance of the factor based on judicial discussions in individual cases. This could be viewed as a limitation, because it is always possible that in individual cases, certain factors will be important but will nonetheless fail to correlate with the case outcome. However, this lack of subjective analysis is exactly what makes this study worthwhile: there is already an abundance of law review articles filled with subjective accounts of what judges and academics say is important. Rather than adding yet another subjective account, the aim of this study is to test those assertions and intuitions objectively.

C.   Selection Effects

Any study that focuses on litigated cases naturally raises the issue of potential selection bias. As Priest and Klein famously observed, it is very difficult to make inferences about the broader universe of all disputes by observing the very low proportion of cases that actually go to trial.[21] The time-consuming and costly process of litigation does not generate a random sample of all potential disputes; rather, litigation acts as a filter, selecting only those cases where uncertainty about the law, asymmetric stakes, divergent expectations, or other quirks of human behavior have prevented the parties from settling their dispute.[22] All of the results reported in this article must be understood in this context. It cannot be said, for example, that cases in which the defendant takes the plaintiff's creative work and uses it to create an informational work (or vise-versa) will always favor the defendant; what can be said is that this pattern holds within the observable data.[23]

The potential for selection bias notwithstanding, the fact remains that litigated cases are important and are constantly subject to ad hoc empirical assessments. Disputes that culminate in written decisions are the primary source of information for lawyers and judges attempting to discover the content of the law. Written opinions are particularly important because they provide analogies and reasoning that can be extended to future

---

[20] See *infra* Part II-A for an explanation of these doctrinal features and Part III-A for the discussion of the results.

[21] George L. Priest and Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J LEGAL STUD 1, 2 (1984); See also Kevin M. Clermont & Theodore Eisenberg, *Do Case Outcomes Really Reveal Anything About the Legal System? Win Rates and Removal Jurisdiction*, 83 CORNELL L. REV. 581, 587-92 (1998) (discussing how the case-selection effect complicates inferences made from plaintiff win rate data).

[22] *See generally*, Robert H. Gertner, *Asymmetric Information, Uncertainty, and Selection Bias In Litigation* 1993 U. CHI. L. SCH. ROUNDTABLE 75 (summarizing the vast literature following on from the Priest-Klein model.) A related complicating factor in any longitudinal study such as ours is that the selection effect may not even be constant as the results of prior cases necessarily inform the expectations of future litigants.

[23] *See* Part II-A-1 *infra* on transformative use and 'Creativity Shift'.

cases. Lawyers and academics constantly call upon their assessment of 'what really happens' in fair use cases to inform their understanding of the law. It is easy to overlook the fact that these explanations are prone to the very same selection bias that qualifies the findings presented in this study.

The unrepresentative nature of written opinions does not imply that scholars should abandon the field of empirical legal studies, but it does call for some caution in interpreting results. Empirical tools are extremely useful for patterns within the observed data. More importantly, they also allow for assessments of whether those patterns are statistically significant, according to conventional standards, in the sense that they reflect a pattern beyond mere coincidence.[24] Others may decide for themselves whether to draw inferences about future cases or disputes that do not result in written opinions.

II.   THEORY, DOCTRINE & TESTABLE HYPOTHESES

This Part reviews the extensive academic and judicial discussion of various aspects of fair use and demonstrates how abstract doctrinal propositions can be reduced to empirically testable hypotheses. Part II-A focuses on the statutory expression of the fair use doctrine in Section 107 of the Copyright Act. This part establishes the linkage between the four broad statutory factors set out section 107 and the seven separate hypotheses which are then tested in Part III. The remainder of Part II considers important non-statutory theories about fair use. Part II-B assesses the characterization of fair use as a social subsidy from copyright owners to those in need of access to copyrighted works. It explains how the subsidy view of fair use implies that fair use litigation should favor the underdog, and the various ways in which that intuition can be tested. Part II-C addresses how industry differentiation affects fair use litigation. Parts II-B and II-C generate an additional five hypotheses which are also tested in Part III.

A.   The Statutory 'Fair Use Factors'

The fair use doctrine originated as a judicially crafted exception to the rights of copyright owners.[25] It remained one of the key non-statutory aspects of copyright law until 1976, when Congress expressly incorporated it into the Copyright Act.[26] To describe fair use as being 'codified' in the 1976 Act—as people often do—is something of an overstatement.

---

[24] Although the significance of some variables in noted at the .10 level, this study relies primarily on results that meet the usual .05 standard. If a variable is significant at the .05 level, either the null hypothesis that the variable is not significant is wrong, or the patterns observed in the data are the product of the sort of random chance that one would expect to see in only 5% of similar regressions. MICHAEL O. FINKELSTEIN, BASIC CONCEPTS OF PROBABILITY AND STATISTICS IN THE LAW, 54 (2009). Importantly, failing to reject the null hypothesis at a given level of significance does not automatically mean that the null is correct. *See* ROBERT M. LIEBERT & LYNN LANGENBACH LIEBERT, SCIENCE AND BEHAVIOR 92 (4th ed. 1995).

[25] *See generally*, Matthew Sag, *The Pre-History of Fair Use*, BROOKLYN LAW REVIEW (forthcoming).

[26] 17 U.S.C. § 107.

The Copyright Act does not define fair use, nor does it contain any definitive examples of particular fair uses.[27] All that the Act offers by way of guidance is: (i) an open list of examples "*such as* criticism, comment, news reporting, …" which may or may not be fair use;[28] (ii) four general factors to be considered in applying the doctrine, and (iii), since 1992, a minor clarification relating to the status of unpublished works.[29]

The four express factors in section 107 have had a considerable effect on the how issues of fair use are analyzed. Over the past thirty years, judges have increasingly relied on the four factors to frame their analysis in fair use cases. Consequently, lawyers also craft their briefs and their advice to clients around these same factors. This is not surprising, after all the language of section 107, "the factors to be considered shall include," indicates that consideration of the statutory factors is mandatory in determining whether the use of a work is fair.[30] However, these four factors were not intended to be exclusive, nor were they intended to be so specific as to freeze judicial development of the doctrine.[31]

The statutory 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.

There is more to the statutory factors than meets the eye. For example, within the broad scope of "the purpose and character of the use" in the first factor, courts take into account a variety of considerations including whether the defendant's use was commercial and whether that use was 'transformative'—a term of art in fair use jurisprudence discussed in detail below.[32] Thus, in effect the first factor is really two separate factors. Under the

---

[27] *Id.*

[28] Harper & Row, Publrs. v. Nation Enters., 471 U.S. 539, 561 (1985) (the examples enumerated in § 107 should not be regarded as exhaustive, nor do they single out any particular use as presumptively fair).

[29] Act of Oct. 24, 1992, Pub. L. No. 102-492, 106 Stat. 3145 (adding the sentence "The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors" to Section 107). See also Kenneth D. Crews, *Fair Use of Unpublished Works: Burdens of Proof and the Integrity of Copyright*, 31 ARIZ. ST. L.J. 1, 49-66 (1999) (discussing the legislative history of the 1992 amendment).

[30] *See* Universal City Studios, Inc. v. Sony Corp. of Am., 480 F. Supp. 429, 448 (C.D. Cal. 1979) ("The factors are illustrative, not definitive."); Bond v. Blum, 317 F.3d 385, 394 (4th Cir. 2003) (the factors represent "only general guidance about the sorts of copying that courts and Congress most commonly have found to be fair uses.") (citations and internal quotation marks omitted).

[31] H.R. Rep. No. 94-1476, at 66 (1976).

[32] A work is considered to be transformative work if it imbues the original "with a further purpose or different character, altering the first with new expression, meaning, or message." Campbell v. Acuff-Rose

umbrella of the second factor, "the nature of the copyrighted work," courts explore two quite separate considerations: the creativity of the plaintiff's work and whether it was published.[33] The third factor, "the amount and substantiality of the portion used in relation to the copyrighted work as a whole," is refreshingly self-descriptive.[34] Under this factor, courts essentially evaluate how much the defendant is alleged to have taken and how important that taking was in the context of the plaintiff's work. The fourth factor focuses on the market effect of the defendant's conduct: it might be the most important factor, or it might be no factor at all.[35]

This section explores each of the factors in detail and proposes specific testable hypotheses in relation to each.

1.   The 'Purpose and Character of the Use' and Transformative Use

The phrase 'transformative use' has loomed large in fair use jurisprudence ever since the Supreme Court embraced transformativeness as the heart of fair use in its 1994 *Campbell* decision.[36] The key issue in *Campbell* was the extent of permissible copying in the context of musical parody.[37] According to the Court, parody is a form of transformative use and "provide[s] social benefit, by shedding light on an earlier work, and, in the process, creat[ing] a new one."[38] The Court argued that the creation of transformative works, although not necessary in every case, lies "at the heart of the fair use doctrine's guarantee of breathing space within the confines of copyright."[39] According to *Campbell*, transformative uses not only occupy the core of the fair use doctrine, transformativeness is a factor which reduces the importance of all other factors such that "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."[40]

To begin to understand transformative use, one must understand how a transformative work differs from a 'derivative work'—another important term of art in copyright law. Section 106(2) of the Copyright Act gives copyright owners an exclusive right to prepare

---

Music, Inc., 510 U.S. 569, 579 (1994) (internal citations omitted). *See infra* notes 44 to 46 and accompanying text addressing the distinction between transformative uses and derivative works. *See infra* Part II-A-1 in relation to transformative use and Part II-A-2 in relation to commercial use.

[33] *See infra* Part II-A-3.

[34] *See infra* Part II-A-4.

[35] *See infra* Part II-A-5.

[36] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994), citing Leval, *supra* note 10, at 1111 (other citations and internal quotation marks omitted).

[37] *Id.*

[38] *Id.* at 579.

[39] *Id.*

[40] *Id.*

derivative works based on the copyright owner's original work.[41] As defined in the statute, a derivative work takes a preexisting work and "*recasts, transforms, or adapts*" that work.[42] The kind of transformations referred to here are not necessarily 'transformative' as that term was intended by the Supreme Court in the context of fair use.[43] A transformative work, in the fair use sense, is one which imbues the original "with a further purpose or different character, altering the first with new expression, meaning, or message."[44] Thus the assessment of transformativeness is not merely a question of the degree of difference between two works; rather it requires a judgment of the motivation and meaning of those differences. The difference between a non-infringing transformative use and an infringing derivative work can be illustrated as follows: if *Pride and Prejudice* were still subject to copyright protection, the novel *Pride and Prejudice and Zombies*, which combines Austen's original work with scenes involving zombies, cannibalism, and ninjas, would be considered a transformative parody of the original, and thus fair use rather than infringement.[45] In contrast, a more traditional sequel would merely be an infringing derivative work.[46]

In the wake of *Campbell*, lower courts have arguably muddied the definition of transformative use by applying the label to a wide range of uses beyond parody. Substantial copying of a novel in the service of criticism was regarded as transformative in *Suntrust Bank v. Houghton Mifflin Co.*;[47] as was copying to create a new work of art with no obvious critical element in *Blanch v. Koons*.[48] Copying without modification was regarded as transformative in *Savage v. Council on American-Islamic Relations, Inc.*, where the Islamic organization copied and distributed anti-Islamic statements made by Michael Savage as part of a fund-raising exercise.[49] Recontextualization without modification from one expressive context to another was seen as transformative *Bill Graham Archives v. Dorling Kindersley Ltd.*[50] In addition to these cases, courts have also

---

[41] 17 U.S.C. § 106 (2).

[42] 17 U.S.C. § 101 (defining derivative work).

[43] A different term expressing the same concept would have been more helpful.

[44] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (internal citations omitted).

[45] The example is analogous to *Suntrust Bank v. Houghton Mifflin Co.* 268 F.3d 1257 (11th Cir. 2001).

[46] *See e.g.,* LINDA BERDOLL, MR. DARCY TAKES A WIFE: PRIDE AND PREJUDICE CONTINUES (PRIDE & PREJUDICE CONTINUES) (2004).

[47] 268 F.3d 1257 (11th Cir. 2001).

[48] 467 F.3d 244 (2d Cir. 2006).

[49] 2008 U.S. Dist. LEXIS 60545 (N.D. Cal. July 25, 2008). *See also* Hustler Magazine, Inc. v. Moral Majority, Inc., 796 F.2d 1148 (9th Cir. 1986) (holding that "an individual in rebutting a copyrighted work containing derogatory information about himself may copy such parts of the work as are necessary to permit understandable comment.").

[50] 448 F.3d 605, 609-10 (2d Cir. 2006) (use of promotional posters in a rock biography was "a purpose separate and distinct from the original artistic and promotional purpose for which the images were created"). *See also* Mattel, Inc. v. Walking Mountain Prods., 353 F.3d 792, 796-98, 800-06 (9th Cir. 2003)

found a number of non-expressive uses to be transformative.[51] In particular, several cases have held that automated processing and display of copyrighted photos as part of a visual search engine is a transformative and thus a fair use.[52] This varied application makes a precise definition of transformative difficult.

In fact, several experts have questioned whether applying the designation of 'transformative use' to contexts where the original copyrighted work has not actually been altered renders the term meaningless, or at least hopelessly imprecise. David Nimmer has suggested that in the hands of some judges, transformative use has no content at all and that it is simply synonymous with a finding of fair use.[53] According to Samuelson, a better approach would be to distinguish transformative critiques, such as parodies, from productive uses for critical commentary. Samuelson also suggests that courts should not label orthogonal uses—uses wholly unrelated to the use made or envisaged by the original author—as transformative uses.[54] In a similar vein, I have argued in a previous article that the term transformative use should be confined to expressive uses of copyrighted works and that non-expressive use should be recognized as a distinct category of preferred use.[55] Despite these sensible criticisms and suggestions, it is likely that lower courts will continue to apply the label of transformative use to any use they think ultimately fair, as long as a finding of transformativeness is perceived to be necessary to avoid the presumption of market harm attaching to commercial uses.[56]

If transformative use has indeed become synonymous with fair use, there can be no empirical measure of transformative use. However, taking the concept of transformativeness at its most literal level makes it possible to test at least one of the core meanings of the term. To the extent that transformative use means making a new work

---

(concluding that photos parodying Barbie by depicting "nude Barbie dolls juxtaposed with vintage kitchen appliances" was a fair use).

[51] For a full discussion of the application of the fair use doctrine to automated and non-expressive uses, *see* Sag, *supra* note 9.

[52] Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007); Kelly v. Arriba Soft Corp., 336 F.3d 811, 816 (9th Cir. 2003). See also, A.V. v. iParadigms, LLC, 562 F.3d 630 (4th Cir. 2009) (finding the automated processing of the plaintiff students' work in defendant's plagiarism detection software was fair use).

[53] DAVID NIMMER, NIMMER ON COPYRIGHT § 13.05[A][1] ("Those Second Circuit cases appear to label a use 'not transformative' as a shorthand for 'not fair,' and correlatively 'transformative' for 'fair.' Such a strategy empties the term of meaning - for the 'transformative' moniker to guide, rather than follow, the fair use analysis, it must amount to more than a conclusory label.")

[54] Samuelson, *supra* note 15 at 2557 ("Although *Campbell* defines 'transformative' in a way that encompasses uses for different purposes, copyright law will be more comprehensible and coherent if iterative copying for orthogonal purposes is distinguished from truly transformative uses of prior works.")

[55] Sag, *supra* note 9 at 1647.

[56] This doctrinal morass could have been avoided if the Supreme Court had more clearly rejected the presumption against commercial fair use in *Campbell*.

out of an old one, then it stands to reason that stark differences between the work allegedly copied and the defendant's work should be indicative of transformation. This general concept is operationalized by the variable *Creativity Shift. Creativity Shift* is set to 1 in cases where the plaintiff's work is creative and the defendant's is informational, or vise-versa.[57] In such cases, the defendant has not just created a new work, she has also created a work in a different category. The testable implication arising from transformative use is then:

> H1   *Creativity Shift makes a finding of fair use more likely.*

This hypothesis does not completely encapsulate every sphere of meaning ascribed to transformative use, but it clearly captures at least one substantial aspect of the concept.

2.   The 'Purpose and Character of the Use' and Commercial Use

The status of commercial fair use leaves many law students, lawyers and judges perplexed. The first statutory factor expressly invites consideration of whether the defendant's use "is of a commercial nature or is for nonprofit educational purposes."[58] In *Sony Corporation of America v. Universal City Studios, Inc.,* the Supreme Court went so far as to declare that "every commercial use of copyrighted material is presumptively an unfair exploitation."[59] This much would seem to suggest a clear rule: commercial use is not fair use. Likewise, it suggests a clear hypothesis:

> H2   *Commercial use by the defendant makes a finding of fair use less likely.*

However, *Sony* is not the final word on commercial use. As the Supreme Court later acknowledged in *Campbell v. Acuff-Rose Music, Inc.,* the *Sony* presumption against commercial uses "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."[60] In *Campbell* the Supreme Court rejected the notion that commerciality by itself had any "hard presumptive significance."[61] Instead, the Court adopted a sliding scale in which the transformativeness of the defendant's use dominates other factors, such as

---

[57] *Creativity Shift* is set to 0 in every other case.

[58] 17 U.S.C. § 107(1). This supposed distinction itself is bewildering as educational uses have often been denied fair use protection and there are many non-commercial uses that have nothing to do with education.

[59] Sony Corp. of America v. Universal City Studios, Inc., 464 U.S. 417, 451 (1984). Similarly, the majority in *Harper & Row* declared that "the fact that a publication was commercial as opposed to nonprofit is a separate factor that tends to weigh against a finding of fair use." Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 562 (1985).

[60] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584 (1994).

[61] *Id.* at 579.

commercialism.[62] As a result, the relevance of commercial use is contingent on the transformative nature of the defendant's use. The implication here is that as an alternative to H$_2$:

    H3   Untransformative commercial uses are less likely to be fair uses.

This still leaves the question of how to measure commerciality in particular cases. Part of the confusion about commercial fair use stems from a failure to distinguish between commercial uses and uses by commercial actors. Lawyers with limited copyright experience often reflexively suggest that the status of the defendant as a for-profit entity makes fair use unavailable as a defense to copyright infringement. However, a presumption against commercial actors has no foundation in the statute, which focuses on the character of the "use," not the identity of the user.[63] Discrimination against commercial actors is also inconsistent with the basic economic logic of copyright—copyright creates private incentives to encourage the creation of expressive works;[64] for those incentives to be effective necessarily presupposes commercial exploitation.[65] Put simply, if copyright assumes that "no man but a blockhead ever wrote, except for money"[66] then it makes very little sense to deny the non-blockheads the protection of the fair use doctrine.

Recent cases have suggested that a better understanding of commercial use focuses not on the identity of the user, but on the substitution effect of the unauthorized copying. In *A&M Records, Inc. v. Napster, Inc.,* for example, the district court determined that users who downloaded copyrighted music via a free internet file sharing service were engaged in a commercial use because, by distributing music files, "Napster users get for free something they would ordinarily have to buy."[67] Defining commerciality in this way is consistent with the essential purpose of copyright law: creating market incentives by protecting creators of original expression from certain types of substitution, primarily

---

[62] *Id.*

[63] 17 U.S.C. § 107(1).

[64] Eldred v. Ashcroft, 537 U.S. 186, 219 (2003) (copyright promotes the creation and dissemination of expression "by establishing a marketable right to the use of one's expression.") (citing Harper & Row, 471 U.S. at 558).

[65] *See* Sag, *supra* note 9 *at* 1684.

[66] *See* Campbell v. Acuff-Rose Music, 510 U.S. 569, 584 (1994) (quoting 3 JAMES BOSWELL, LIFE OF JOHNSON 19 (G. Hill ed., The Clarendon Press 1934) (1787)).

[67] Napster, 114 F. Supp. 2d at 912. *See also* Harper & Row, 471 U.S. at 562 ("The crux of the profit/nonprofit distinction is not whether the sole motive of the use is monetary gain but whether the user stands to profit from exploitation of the copyrighted material without paying the customary price."); American Geophysical Union v. Texaco, Inc., 60 F.3d 913, 922 (2d Cir. 1994) (researchers at for-profit laboratory gained indirect economic advantage by photocopying copyrighted scholarly articles).

expressive substitution.[68] However, this construction of commerciality is almost exactly equivalent to the concept of market effect under the fourth factor.[69]

If commercial use is an independent factor that predicts the outcomes of fair use cases, it must mean something analytically distinct from market effect. In his recent analysis of the contents of fair use opinions, Beebe found that although 84% of opinions address the issue of commerciality, the sub-factor appears to have no significant influence on case outcomes.[70] Beebe relies on *ex post* judicial assessments of whether a use is commercial. In contrast, this study sets aside such after the fact categorizations because they tell us nothing about the *ex ante* predictability of fair use. This is an important difference because the designation of "commercial use" in fair use cases sometimes confounds common sense understandings of what is, and is not, commercial. Most dictionaries define 'commercial' as "occupied with or engaged in commerce or work intended for commerce,"[71] with 'commerce' defined, for example, as "the exchange or buying and selling of commodities on a large scale involving transportation from place to place."[72] On the other hand, case law tells us that the free file sharing of copyrighted music is commercial,[73] and that the free dissemination of a religious book by a dissident church "unquestionably profits" that church and is thus tantamount to commercial.[74]

This study distinguishes between non-commercial and commercial uses as those terms would be likely to be understood by potential litigants.[75] The category of non-commercial use includes: (i) personal uses, such as time-shifting broadcast television; (ii) educational uses, where the defendant's use of the plaintiffs work was related to some kind of formal education at a recognized educational institution; (iii) research uses, where the defendant's use was for the purpose of general research and was not part of a specific product development process; and (iv) activities such as peer-to-peer file sharing which communicate the plaintiff's work (or some derivative thereof) to the public, but without direct or indirect commercial benefit. The defendant's use is categorized as "commercial"

---

[68] Sag, *supra* note 9 (arguing that a number of copyright doctrines converge around the idea that a subsequent author should be free to compete using her own expression of facts, concepts, and ideas originating with prior authors, but not by offering an original author's own expression to the public).

[69] *See infra*, Part II-E.

[70] Beebe, *supra* note 14 at 556, 598.

[71] *See* e.g., THE MERRIAM-WEBSTER DICTIONARY 99 (2005); 3 THE OXFORD ENGLISH DICTIONARY 552 (2nd ed. 1989); and BLACK'S LAW DICTIONARY 304 (9th ed. 2009).

[72] *Id.*

[73] A & M Records, Inc. v. Napster, Inc., 114 F. Supp. 2d 896, 912 (N.D. Cal. 2000); A&M Records v. Napster, Inc., 239 F.3d 1004, 1015 (9th Cir. 2001).

[74] Worldwide Church of God v. Philadelphia Church of God, Inc., 227 F.3d 1110, 1118 (9th Cir. 2000).

[75] This is not discarding the relevance of market substitution, merely relegating it to its proper place under the fourth factor.

if the defendant used the plaintiff's work as part of a commercial product or service or as an intermediate step to creating a commercial product or service.

To take account of the potential interaction of commercial use with transformative use, I coded a separate variable to distinguish between direct and indirect commercial exploitation. Any use of the plaintiff's copyrighted work in a product or service sold to the public was classified as direct commercial exploitation unless: (i) the work was only used or copied as part of an intermediate process; or (ii) the defendant had taken an extra step, applying its own labor or creativity to somehow change the original copyrighted work. However, the addition of new material that left the original copyrighted work basically the same was not treated as a involving an extra step. Thus, for the purposes of this study, H_3 can be restated as:

> H3   *Direct Commercial Use is less likely to be found to be fair use.*

3.   The Nature of the Work

The next fair use factor instructs courts to consider "the nature of the copyrighted work."[76] Over the past thirty years, courts have typically distilled the amorphous 'nature' of the work into two more tractable considerations: whether the plaintiff's work is creative as opposed to informational;[77] and whether the work is unpublished, as opposed to published.[78]

Several prominent fair use cases from 1984 to 1992 indicated that the fair use doctrine has no application where the plaintiff's work was unpublished. In 1984 a majority of the Supreme Court declared in *Harper & Row, Publishers, Inc. v. Nation Enterprises* that "under ordinary circumstances, the author's right to control the first public appearance of his undisseminated expression will outweigh a claim of fair use."[79] This ruling was followed and expanded in a series of cases in the Second Circuit.[80] However, in 1992, the force of this sub-factor was blunted by an amendment to section 107, adding the following to the statutory description of fair use: "The fact that a work is unpublished

---

[76] 17 U.S.C. § 107(2).

[77] *E.g.,* Leadsinger, Inc. v. BMG Music Publ'g, 512 F.3d 522, 531 (9th Cir. 2007) (describing song lyrics as "a work of creative expression, *as opposed to an informational work*, which is precisely the sort of expression that the copyright law aims to protect.") (emphasis added).

[78] Harper & Row, Publishers, Inc. v. Nation Enterprises 471 U.S. 579, 564 (1985).

[79] *Id.*

[80] Salinger v. Random House 811 F.2d 90, 97 (2d Cir. 1987) (holding that "unpublished works … normally enjoy complete protection against copying any protected expression"); New Era Publ'ns Int'l v. Henry Holt & Co., Inc., 873 F.2d 576, 583 (2d Cir. 1989) ("Where use is made of materials of an 'unpublished nature,' the second fair use factor has yet to be applied in favor of an infringer, and we do not do so here."); Wright v. Warner Books, Inc. 953 F.2d 731, 737 (2d Cir. 1991) ("unpublished works are the favorite sons of factor two" and that "our precedents ... leave little room for discussion of this factor once it has been determined that the copyrighted work is unpublished.").

shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors."[81] Nevertheless, it is appropriate to test the hypothesis:

> H4   *A finding of fair use is less likely where the plaintiff's work is unpublished.*

A second aspect of the nature of the work is that, in principle, the more creative the original work is, the more justification is required to establish a claim of fair use.[82] As the *Campbell* Court explained, "some works are closer to the core of intended copyright protection than others, with the consequence that fair use is more difficult to establish when the former works are copied."[83] Accordingly, the second factor presents another hypothesis:

> H5   *A finding of fair use is less likely where the plaintiff's work is creative.*

4.   Amount Copied

The next statutory fair use factor is "the amount and substantiality of the portion used in relation to the copyrighted work as a whole."[84] The application of this factor is inherently bound up in the first and fourth factors: the extent of permissible copying varies according to the purpose of the copying and the effect of the copying on the rightsholder.[85] Accordingly, the third factor is usually said to be both qualitative and quantitative.[86]

The contextual element in the third factor makes it challenging to reduce to a simple empirical question. There is, however, a reasonably objective distinction that illuminates the role played by the this factor. In most cases it is easy to distinguish between a defendant accused of making a complete copy of the plaintiff's work and one who is accused of making only a partial copy. Thus the relevant hypothesis in relation to the third factor is:

> H6   *A finding of fair use is more likely where the defendant uses only part of the plaintiff's work.*

---

[81] Pub. L. No. 102-492, 106 Stat. 3145 (1992) (codified as amended at 17 U.S.C. § 107 (2000 & Supp. IV 2004))

[82] Campbell v. Acuff-Rose Music, 510 U.S. 569, 586 (1994).

[83] *Id.* at 586.

[84] 17 U.S.C. § 107(3).

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

[86] Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985).

5.   *Market Effect*

The final statutory factor is "the effect of the use upon the potential market for or value of the copyrighted work."[87] Although the Supreme Court once commented that this factor is "undoubtedly the single most important element of fair use," subsequent cases have retreated from that position.[88] In its most recent fair use decision, *Campbell v Acuff Rose, Inc.,* the Supreme Court notably failed to stress the market effect factor; it insisted instead that "all [statutory factors] are to be explored, and the results weighed together, in light of the purposes of copyright."[89]

In his investigation, Beebe noted that of the 141 judgments that found that the market effect factor disfavored the defendant, all but one also ruled against the defendant on the ultimate issue of fair use. Finding a 99% correlation in an empirical study is a bit like finding that 99% of Iraqis voted for Saddam Hussein—it is a statistic so impressive that it engenders disbelief.[90] Although market effect has the ring of an objective factual investigation, it is in fact a highly subjective assessment. In most cases, the judgment as to whether the defendant's conduct falls foul of the fourth factor is somewhat elastic because to ascertain the extent of potential market effect, courts must first determine whether a given use is within the potential market. The proper scope of the plaintiff's current and future copyright markets is thus an input, throughput and output of this analytical process.[91] In light of this imprecision, the apparent symmetry between the way judges describe the fourth factor and the outcomes of fair use cases raises the suspicion that market effect is no factor at all.[92]

The question then becomes: Is there any empirical content to the fourth statutory factor, or is it merely a legal conclusion? The simplest way to understand the market effect factor is as an inquiry into competitive injury or lost sales. Conceived as such, the empirical question is just a matter of finding that injury or those lost sales. This inquiry is easy to conceive but harder to implement because the extent of past losses and probable extent of future losses are keenly contested elements in most fair use trials.

---

[87] 17 U.S.C. § 107 (4).

[88] Harper & Row, Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985).

[89] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 (1994).

[90] John F. Burns, *"Threats and Responses: Support For Hussein; A Show of Loyalty (Just Say Yes) In Iraq Vote for the One and Only"* N.Y. TIMES, Oct. 16 2002.

[91] *See* Matthew Sag, *God in the Machine: A New Structural Analysis of Copyright's Fair Use Doctrine*, 11 MICH. TELECOMM. TECH. L. REV. 381, 392-395 (2005). Jim Gibson argues that the circularity of the market effect determination combined with risk-averse licensing practices causes the inexorable expansion of copyright and a parallel reduction in the scope of fair use. James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882 (2007).

[92] Beebe, *supra* note 14 at 620-21, comes to the same conclusion.

Although it is hard to verify the likely extent of market interference in any particular case, it seems intuitive that the likelihood of competitive injury increases when the plaintiff and defendant are in the same industry. To analyze this issue, each plaintiff and defendant in the dataset was assigned an industry classification based on the North American Industry Classification System ("NAICS").[93] This data was then used to generate an *Industry Separation* dummy variable (a variable that is either 0 or 1) indicating whether plaintiff and defendant were in the same industry (*Industry Separation* = 0) or not (*Industry Separation* = 1).[94] The testable hypothesis relating to market effect is:

> *H7   Industry Separation makes a finding of fair use more likely.*

This section has established the motivation for several testable propositions in relation to the statutory fair use factors and the various sub-factors that have emerged from the case law over the past 30 years. The remainder of this Part explores additional theories about fair use which are less closely tied to the statutory text.

B.   Does Fair Use Favor the Underdog?

Both judges and academics have characterized the fair use doctrine as a redistributive tool favoring the politically and economically disadvantaged. For example, in *Sony Corp. v. Universal City Studios,* Justice Blackmun characterized the fair use doctrine as "a form of subsidy" at the expense of authors that permits limited use of a work "for the public good."[95] A number of scholars have also emphasized the function of fair use in redistributing value from copyright owners to "preferred classes of users."[96] This characterization is not neutral, it reflects a skeptical or grudging view of fair use. After all, subsidies are treated with great skepticism in a free-market economy, and so the

---

[93] NAICS codes are used by Federal statistical agencies in classifying business establishments for the purpose of collecting, analyzing, and publishing statistical data related to the U.S. business economy. See http://www.census.gov/eos/www/naics/.

[94] The purpose of measuring *Industry Separation*, I used two-digit NAICS codes.

[95] Sony Corp. v. Universal City Studios, 464 U.S. 417, 477-78 (1984) (Blackmun, J., dissenting, citing LATMAN FAIR USE STUDY 31; Wendy Gordon, *Fair Use as Market Failure: A Structural Analysis of the Betamax Case and its Predecessors*, 82 COLUM. L. REV. 1600, 1630 (1982).)

[96] Jane C. Ginsburg, *Authors and Users in Copyright*, 45 J. COPYRIGHT SOC'Y U.S.A. 1, 15 (1997) (emphasizing the "redistribution" of value from copyright owners to preferred classes of users.); Jane C. Ginsburg, Copyright, *Common Law, and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. CIN. L. REV. 151, 169 (1997) (viewing fair use as subsidy from copyright owner in favor of uses with public benefits). *See also* Robert P. Merges, *The End of Friction? Property Rights and Contract in the "Newtonian" World of On-Line Commerce*, 12 BERKELEY TECH. L.J. 115, 134-35 (1997) (characterizing fair use as a tax on copyright owners and a subsidy in favor of certain classes of users); 2 PAUL GOLDSTEIN, COPYRIGHT § 10.2.2.1(a) (2005) (the commercial-noncommercial distinction in fair use can be characterized as "a covert subsidy to worthy nonprofit enterprises such as schools and universities.").

subsidy view of fair use naturally suggests that the doctrine's benefits should be reserved for the truly deserving, if not eliminated altogether.[97]

What is the foundation for this view? There is an intuition that fair use is some kind of social subsidy favoring underdog defendants—defendants who lack status, power and money, particularly compared to their litigation opponents. This underdog view of fair use arises, in part, from the focus of the statutory examples on activities such as criticism, comment, news reporting, teaching, scholarship, and research.[98] It also stems from the distinction between uses made for "nonprofit educational purposes" and those of a "commercial nature" found in the first fair use factor.[99]

Furthermore, the intuition that fair use favors the underdog can also be traced to the predominant law and economics account. In her influential 1982 article, *Fair Use as Market Failure,* Wendy Gordon argued that fair use should be understood as a doctrine justifying unauthorized copying in instances of market failure in general, and high transaction costs in particular.[100] Fair use has commonly been found in situations where the cost of collecting a license fee would exceed the license fee itself.[101] This view is easy to reconcile with the statutory text: the fourth fair use factor considers the "effect of the use upon the potential market for, or value of, the copyrighted work"—there is no market effect where there is no effective market.[102]

The market failure view leads to a preference for the underdog, because an underdog's failure to pay the customary price for a particular copy of a work is less likely to

---

[97] *See e.g.,* LEE WILSON, FAIR USE, FREE USE, AND USE BY PERMISSION: HOW TO HANDLE COPYRIGHTS IN ALL MEDIA ("Fair use is a kind of public policy exception to the usual standard for determining copyright infringement;" fair use requires a "countervailing public interest … typically creates some social, cultural or political benefit"); COPYRIGHT PROTECTION ON THE INTERNET: HEARINGS ON H.R. 2441 BEFORE THE SUBCOMM. ON COURTS AND INTELLECTUAL PROPERTY HOUSE COMM. ON THE JUDICIARY, 104TH CONG., 2d SESS. (1996) (statement of Barbara A. Munder, Senior Vice President of the McGraw-Hill Companies and Chair of the Information Industry Association Board of Directors) (arguing that the burden of subsidizing access "is a broader societal responsibility, not one that should be borne primarily – let alone exclusively – by copyright owners.") Of course, many commentators resist the label of subsidy precisely because once fair use is identified as such it becomes easier to argue for its elimination. *See* Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. 354, 363 (1999) (arguing that the public domain is not a subsidy to users).

[98] 17 U.S.C. § 107.

[99] 17 U.S.C. § 107(1).

[100] 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) (arguing that fair use exception to the exclusive rights of copyright can only be justified by the presence of market barriers such as high transaction costs, externalities, non-monetizable benefits, or anti-dissemination motives). *See also* Lydia Pallas Loren, *Redefining the Market Failure Approach to Fair Use in an Era of Copyright Permissions Systems*, 5 J. INTELL. PROP. L. 1 (1997) (focusing on externalities).

[101] *See* Ty, Inc. v. Publ'Ns Int'l, 292 F.3d 512, 518 (7th Cir. 2002) (Posner J.)

[102] 17 U.S.C. § 107(4).

adversely affect the copyright's broader market. The underdog is also, in general, a less concentrated interest and thus the transaction costs of collecting royalties are likely to be higher, so it is more likely that the costs of collection would exceed potential revenue.

Testing the doctrinal intuition that fair use favors the underdog involves reducing that general proposition down to something more concrete. Although litigants do not come to court with the labels underdog and overdog stamped across their foreheads, there is a well developed literature exploring the divergent experience of the haves and have-nots in litigation.[103] This study borrows extensively from this literature to suggest appropriate measures of underdog status.

1.   Legal Personality as a Measure of Underdog Status

Overdogs and underdogs are conventionally defined in terms of status, power, and resources.[104] In studies of Supreme Court decision making, it is commonplace to classify governments and corporations as overdogs because they are presumed to have more resources and higher status; and, for similar reasons, to classify individuals as underdogs.[105] Although it is possible to measure status and empowerment in many different ways, one of the most common approaches is to reduce the question to legal personality: i.e., whether a party is an organization or an individual natural person. One of the pioneers of this field, Marc Galanter, argues that the well-resourced repeat-players are almost always organizations. Consequently, the superiority of the "haves" over the "have-nots" can be tested simply by comparing the fortunes of organizations, businesses, and government bodies against individual natural persons.[106] There is a large body of evidence to suggest that overdogs are indeed advantaged by their superior financial resources, their greater experience with the legal system, and their ability to strategically settle unfavorable cases.[107]

In general the evidence suggests "that organizations do better than individuals in almost every kind of litigation, at almost every stage, and as both plaintiffs and defendants."[108]

---

[103] Marc Galanter, *Why the 'Haves' Come Out Ahead: Speculations on the Limits of Legal Change*, 9 LAW & SOC'Y REV. 95, 97 (1974) (summarizing the literature).

[104] *Id.*

[105] *See e.g.,* Sidney S. Ulmer, *Selecting Cases for Supreme Court Review: An Underdog Model*, 72 AM. POL. SCI. REV. 902 (1978) (examining litigant status as an independent variable in Supreme Court decision making, finding that liberal justices supported underdogs and conservative justices supported upperdogs).

[106] Marc Galanter, *Planet of the APs: Reflections on the Scale of Law and its Users*, 53 BUFFALO L. REV. 1369, 1389 (2006).

[107] *Id.*

[108] *Id. See also* Terence Dunworth & Joel Rogers, *Corporations in Court: Big Business Litigation in U.S. Federal Courts, 1971-1991*, 21 LAW & SOC. INQUIRY 497, 557 (1996) (Finding a 17 percentage point success gap in favor of Fortune 2000 companies as plaintiffs in federal cases and 29 percentage point success gap as defendants); Theodore Eisenberg & Henry S. Farber, *The Litigious Plaintiff Hypothesis: Case Selection and Resolution*, 28 RAND J. ECON. S92, S99 (1997) (Finding a success gap for corporate

For example, in a study of the success of appellants before selected U.S. Courts of Appeals, Donald Songer and Reginald Sheehan found that "the success rates of appellants consistently increase with each incremental increase in their strength relative to the strength of the respondent."[109] Accordingly, the notion that fair use favors the underdog leads to the testable implication that:

> *H8*   *A finding of fair use is more likely if the defendant is a natural person.*

Unfortunately, at least from an empiricist's perspective, the doctrinal intuition that fair use favors the underdog cuts against the obvious and well established disadvantages of the underdog in litigation. The litigation literature establishes a strong prior that, compared to corporate entities, natural persons will fare badly in litigation. I address this issue more in more detail in Part III, below.

2.   Legal Representation as a Measure of Underdog Status

Based on the intuition that underdogs would not have the same access to legal services as overdogs, the characteristics of the parties' attorneys and law firms may provide an another measure of empowerment. I cross-referenced court records with the widely used Martindale-Hubbell directory to determine experience ratings for the individual attorneys and law firms in each case.[110] The variables used in this analysis relate to the Martindale-Hubbell "AV" ranking for attorneys and law firms. An AV ranking indicates that the lawyer (or firm) in question received an average rating of 4.5 or above out of 5 and is regarded as "preeminent" by his or her peers.[111] Only 40% of plaintiffs and 51% of defendants retained as their principal representative an attorney with the highest possible Martindale-Hubbell rankings.[112] A preference for acclaimed experience was more

---

plaintiffs versus individual plaintiffs of 23 percentage points in federal diversity cases not including personal injury cases).

[109] Donald R. Songer & Reginald S. Sheehan, *Who Wins on Appeal? Upperdogs and Underdogs in the United States Courts of Appeals*, 36 AM. J. POL. SCI. 235, 246 (1992); See also, Donald R. Songer et al., *Do the 'Haves' Come Out Ahead Over Time? Applying Galanter's Framework to Decisions of the U.S. Courts of Appeals, 1925-1988*, 33 LAW & SOC'Y REV. 811 (1999) (Finding that the "haves" win more frequently in published decisions, even after controls are introduced for the ideological makeup of the panel.).

[110] MARTINDALE-HUBBELL INC., THE MARTINDALE-HUBBELL LAW DIRECTORY 1978-2006.

[111] Martindale-Hubble, *Peer Review Ratings*, http://www.martindale.com/xp/legal/About_Martindale/Products_and_Services/Peer_Review_Ratings/ratings.xml.

[112] According to Martindale-Hubbell, their Peer Review Ratings evaluate lawyers in the United States and Canada based on the anonymous opinions of members of the Bar and the Judiciary. Attorneys with less than three years experience are unlikely to have a Martindale-Hubbell rating because the company usually conducts its first review three years after Bar admission. Attorneys with an average score of 4.5 to 5 out of 5 receive an 'AV' or 'Preeminent' rating, which, according to Martindale-Hubbell, is a "testament to the fact that a lawyer's peers rank him or her at the highest level of professional excellence." *Id.*

apparent in the selection of law firms: 81% of plaintiffs and 77% of defendants chose law firms with the highest possible Martindale-Hubbell rating.

Comparing the representation of both sides in each case, I generated dummy variables to indicate those cases where the plaintiff availed itself of AV-rated lawyers (or law firms) and the defendant did not.[113] The hypothesis here is that, if fair use favors the underdog defendant, those defendants who retain less experienced attorneys and law firms will actually be more likely to be advantaged by the application of the doctrine. This is actually two related hypotheses:

> H9   *If the defendant is an underdog in terms of attorney representation,*
> *a finding of fair use is more likely; and*

> H10   *If the defendant is an underdog in terms of law firm representation,*
> *a finding of fair use is more likely.*

As before, these hypothesis contradict the general expectation that more highly rated lawyers and more prestigious law firms should ordinarily obtain superior results for their clients.[114] Although the intuition that better lawyers will produce more favorable outcomes is both obvious and appealing, there is surprisingly little evidence to that effect. It is not clear, for example, whether prestige translates to lawyer quality. In a recent study of public defenders randomly assigned to felony cases in Nevada, David Abrams and Albert Yoon conclude that more experienced lawyers produced better outcomes, but the authors did not find that any significant differences in sentencing based on the attorney's legal educational background.[115] More generally, Haire, Lindquist, and Hartley find only mixed support for the proposition that attorney expertise influences litigant success.[116]

---

[113] The plaintiff/defendant correlations for attorney ratings and firm ratings were quite low, .13 and .06 respectively.

[114] Another issue to consider is that the selection of lawyers and law firms in the data set is not the product of random assignment. The parties decide how much to invest in legal services based on their anticipation of the risks and rewards of litigation and their own perceptions of the incremental advantage that can be obtained by retaining more experienced counsel. There is no obvious way to control for the endogeneity of these factors. *See* Stanton Wheeler, Bliss Cartwright, Robert A. Kagan, & Lawrence M. Friedman, *Do the 'Haves' Come Out Ahead? Winning and Losing in State Supreme Courts, 1870-1970*, 21 LAW & SOCIETY REV. 403 (1987) (Comparing solo practitioners to lawyers practicing in firms and partnerships in state supreme court cases, finding that stronger parties tended to retain attorneys affiliated with a firm and that clients represented by firms fared better.)

[115] David S. Abrams & Albert H. Yoon, *The Luck of the Draw: Using Random Case Assignment to Investigate Attorney Ability*, 74 U. CHI. L. REV. 1145 (2007) (A defendant assigned to an attorney in the top 10 percent of experience had average incarceration rate 14 percentage points lower than those with attorneys in the bottom 10 percent. A similar comparison of expected sentence length yields an expected sentence reduction of 1.2 months, simply due to the randomly assigned attorney.).

[116] Susan Brodie Haire, Stefanie A. Lindquist, & Roger Hartley, *Attorney Expertise, Litigant Success, and Judicial Decisionmaking in the U.S. Courts of Appeals,* 33 LAW & SOC'Y REV. 667 (1999). (Reviewing products liability cases decided from 1982 and 1993).

Thus, although one might expect attorney and firm ratings to show a positive correlation with litigation outcomes, the extant literature suggests this can only be a weakly held prior.

C.   Industry Effects

Dan Burk and Mark Lemley have made a strong case for the importance of industry variation in patent law.[117] It stands to reason that a similar phenomena might apply to copyright. The same NAICS data used to generate the industry separation variable referred to above can also be used to chart industry differences in fair use litigation.[118]

As seen in Figures 3 and 4 (below), for both plaintiffs and defendants, the 'Information' industries dominate fair use litigation (41% and 52%, respectively). In the NAICS coding scheme, the 'Information' category, includes publishing, sound recording, motion pictures, telecommunications, radio, satellite, television, data processing and libraries and archives.[119] For defendants, no other category accounts for more than 10% of the population; whereas, not surprisingly, the 'Arts, Entertainment & Recreation' group accounts for 20% of plaintiffs.[120]

---

[117] Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1578 (2003) (finding that patent doctrine varies significantly by industry).

[118] See *supra* note 93 and accompanying text.

[119] NAICS 511110 to 519190.

[120] The 'Arts, Entertainment & Recreation' category, NAICS 711110 to 7113990, includes theatre companies, dance companies, music groups, sports teams, promoters of performing arts, sports, and similar events, independent artists, writers, and performers as well as museums, zoos, amusement parks, casinos, golf courses and other amusement or recreational facilities.

Figure 1: Plaintiff Industries (Major Categories)
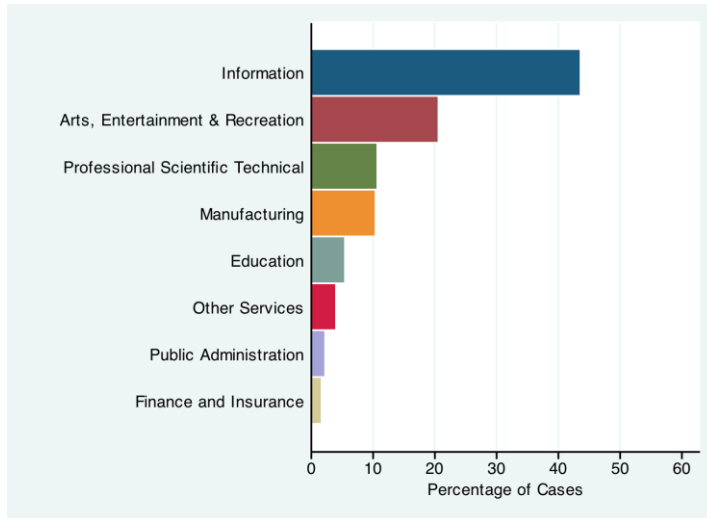
Figure 2: Defendant Industries (Major Categories)
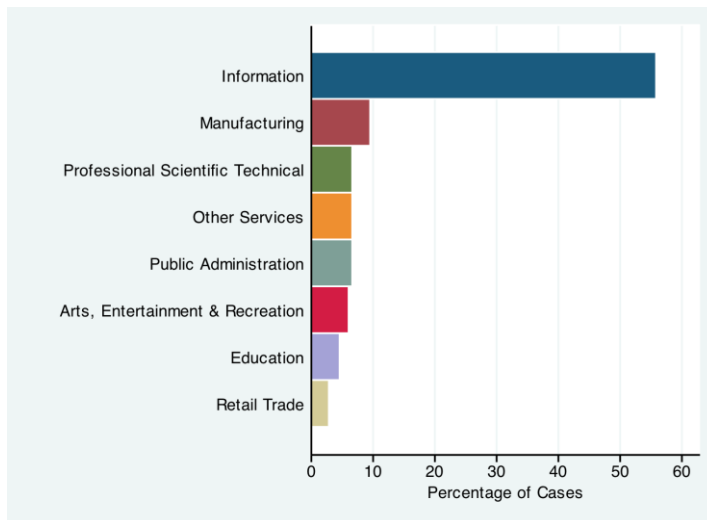
The population in this study is too small to test the significance of individual industries on fair use outcomes, however it is possible to test the more general hypothesis that the *Industry Group* variable will be a significant factor in fair use outcomes. Accordingly, the final two hypotheses are:

H11   *Plaintiff Industry Group is a significant factor in fair use outcomes; and*

H12   *Defendant Industry Group is a significant factor in fair use outcomes.*

Even if *Industry Group* does not turn out to be significant, it is important to include some measure of industry effects in the regression model as a control variable in order to reduce the chances that effects observed in relation to other variables are not simply a product of unobserved characteristics associated with either plaintiff or defendant. Including *Industry Group* in the analysis does not preclude this possibility, but it goes some way towards minimizing it.

III.   RESULTS

The previous Part demonstrated how substantial doctrinal questions relating to fair use can be translated into three clusters of testable hypotheses. The first group consists of implications arising directly from the four statutory fair use factors:

H1   *Creativity Shift makes a finding of fair use more likely;*

H2   *Commercial use by the defendant makes a finding of fair use less likely;*

H3   *Direct commercial use makes a finding of fair use less likely;*

H4   *A finding of fair use is less likely where the plaintiff's work is unpublished;*

H5   *A finding of fair use is less likely where the plaintiff's work is creative;*

H6   *A finding of fair use is more likely where the defendant uses only part of the plaintiff's work; and*

H7   *Industry Separation makes a finding of fair use more likely.*

The second group of propositions relates less directly to the statutory text of section 107. Instead, these hypotheses flow from judicial and academic commentary in relation to the fair use doctrine. These additional hypotheses are:

H8   *A finding of fair use is more likely if the defendant is a natural person;*

H9   *If the defendant is an underdog in terms of attorney representation, a finding of fair use is more likely; and*

H10   *If the defendant is an underdog in terms of law firm representation, a finding of fair use is more likely.*

The third and final set of propositions relate to the effect of party characteristics tied to broader industry classifications. These are:

> H11   Plaintiff Industry Group is a significant factor in fair use outcomes; and

> H12   Defendant Industry Group is a significant factor in fair use outcomes.

The table below reports the results of four separate multivariate logistic (or logit) regressions analyzing the relationship between a finding in favor of fair use (the dependant variable) and the various factors discussed in Part II (the independent variables).[121] I use logit because the dependant variable of interest in this study is dichotomous—either the defendant's claim of fair use was successful, or it was not. Linear regression fit the values of the outcome variable along straight line, however, given that the values here are all either 0 or 1, the usual way to model dichotomous dependant outcomes is using logistic regression to predict the probability that the outcome = 1.

Model 1 shows the effect of the statutory factors on fair use. Specifically it explores the relationship between a finding in favor of fair use and the following seven variables: *Creative Shift, Commercial Use, Direct Commercial Use, Creative Work, Unpublished Work, Partial Copy* and *Industry Separation*. The remaining models reported in the table below are variations of the first model. Model 2 adds variables relating to underdog status to the analysis of the core statutory factors: these variables are *Plaintiff Natural Person, Defendant Natural Person, Defendant Attorney Underdog* and *Defendant Law Firm Underdog*. Model 3 combines the statutory factors with the plaintiff and defendant industry variables, *Plaintiff Industry* and *Defendant Industry*. Model 4 contains all three groups of variables and should be considered as the main finding of this study.

---

[121] Logit (or logistic regression) is a regression model designed for estimation of binary outcomes. I report logit results, but I have also verified the results using probit (another regression model designed for estimation of binary outcomes) and Ordinary Least Squares (OLS) (which is more suitable for a linear regression model) and the results are substantively the same in all relevant respects. See Appendix for OLS results.

Table 1: The Effect of Various Objective Factors on a Finding in Favor of Fair Use (Multivariate Logit Regression)

Doctrinal Basis	Variable	Regression Model Finding of Fair Use			
		(1)	(2)	(3)	(4)
Transformative use	Creative Shift	**1.320****	**1.336****	**1.188****	**1.287****
		(.41)	**(.46)**	**(.43)**	**(.46)**
Commercial use	Commercial Use	-.311	-.483	-.308	-.412
		(.47)	(.53)	(.50)	(.53)
	Direct Commercial Use	**-.746***	**-.796***	**-.672#**	**-.760***
		(.34)	**(.35)**	**(.35)**	**(.35)**
Nature of the Work	Creative Work	-.071	-.437	-.180	-.427
		(.34)	(.39)	(.36)	(.39)
	Unpublished Work	.386	.099	.094	.017
		(.42)	(.50)	(.44)	(.50)
Amount Copied	Partial Copy	.452	**.776***	.498	**.747***
		(.32)	**(.36)**	(.32)	**(.36)**
Market Effect	Industry Separation	.100	-.122	-.178	-.207
		(.30)	(.33)	(.31)	(.33)
Underdog	Plaintiff natural person		**1.570****		**1.323****
			(.42)		**(.45)**
	Defendant natural person		.269		.244
			(.48)		(.47)
	Defendant attorney underdog		**-.962***		**-.869#**
			(.48)		**(.48)**
	Defendant law firm underdog		.257		.270
			(.45)		(.45)
Industry	Plaintiff Industry			**.004****	.016
				(.01)	(.013)
	Defendant Industry			.011	.007
				(.01)	(.01)
	Constant	-.429	-.486	-3.025	-1.73
	Pseudo R^2	.081	.166	.123	.172

N = 222. ** Significant at the .01 level, * Significant at the .05 level, # Significant at the .10 level.
Logit coefficients obtained using Robust Standard Errors. Standard errors in parentheses.

1.   The Core Statutory Fair Use Factors

The variables *Creative Shift* and *Direct Commercial Use* were statistically significant in each of the four model specifications. The positive and statistically significant coefficient associated with *Creative Shift* shows that, as predicted in H_1, this proxy measure for transformative use makes a finding of fair use more likely. Likewise, the significant

negative coefficient for *Direct Commercial Use* confirms H$_3$, that direct commercial use makes a finding of fair use less likely.

These results are not merely statistically significant, they are also clearly substantively significant. Based on the regression in Model 4, the average predicted probability of a finding in favor of fair use is .62 if there is a *Creative Shift,* but only .33 otherwise.[122] This means that, holding everything else constant, the chances of a fair use win are almost double, increasing from 33% to 62%, when this kind of transformative use is present. In other words, not knowing anything else about the defendant's use, a plaintiff can expect to win a clear majority of cases where there is no indication of transformative use, but otherwise expect to lose all but 38% of the time.

The substantive effect of *Direct Commercial Use* is similarly impressive. The predicted probability of a finding in favor of fair use is .29 if the defendant is making a direct commercial use, and .47 otherwise. In addition the combination of these to effects is also substantively important. For example, the predicted probability of a successful assertion of fair use if there is *Creative Shift* and the defendant is not engaged in a direct commercial use is .72—i.e., more than seven cases in ten. The evidence emphatically confirms the significance of transformative in fair use litigation.

The other fair use factors do not initially show a significant effect. *Commercial Use*, *Creative Work*, *Unpublished Work*, and *Industry Separation* are insignificant across all four regression models. It must be stressed here that failure to find an effect is not the same thing as finding no effect.[123] It is possible that a different research design, or simply more data could lead to a different result. Nonetheless, for the moment what can be said is there is no evidence that commercial use, the creative or unpublished nature of the copyrighted work or market effect (measured in terms of industry separation) play any role in determining the outcome of fair use cases.

Does this prove that transformative use is the alpha and omega of fair use? Not quite. First, it should be clear that although certain core aspects of transformative use are captured in the *Creative Shift* and *Direct Commercial Use* variables, these variables do not account for every aspect of transformative use. The clearest indication of the shortcomings of these two measures is that neither would indicate that the defendant in *Campbell v. Acuff-Rose*—the first judicial invocation of the term transformative use— was engaged in a transformative use.[124]

Second, it is probably too soon to conclude that 'the amount and substantiality of the portion used' has no effect on fair use outcomes. Although *Partial Copy* was not significant in Models 1 or 3, this variable was significant in Models 2 and 4. Recall that

---

[122] Predicted probabilities range between 0 and 1 and are rounded to two decimal places.

[123] See *infra* note 24.

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

what distinguishes these two models is the inclusion of variables relating to party status and empowerment characteristics. The significance of *Partial Copy* in these models means that, controlling for certain attributes of legal personality and legal representation, *Partial Copy* is a significant predictor of fair use.[125] The average predicted probability of fair use for cases involving partial copying is .44, whereas it is only .34 in those involving total copying. Thus, although its significance depends on the model specification, it seems reasonable to conclude that, consistent with H_6, a finding of fair use is more likely where the defendant uses only part of the plaintiff's work.

As discussed in more detail in the conclusion, the preceding results are informative both in terms of what they show and what they do not show. The evidence that transformative use and partial use increase the odds of a finding of fair use suggests that there is some empirical content to the statutory factors. However, the lack of evidence in relation to commercial use is also important, given the disputed nature of that sub-factor.

2.   Party Status Variables

The party status and empowerment variables in Models 2 and 4 yield some interesting results. The coefficient for *Plaintiff Natural Person* is positive and significant which means that a finding of fair use in favor of the defendant is more likely if the plaintiff is a natural person and less likely if the plaintiff is a corporation. This is consistent with the well developed literature on the success of haves over the have-nots in litigation discussed in Part II.[126] However, it is in direct opposition the notion that fair use favors the underdog and undermines the characterization of fair use as a redistributive tool favoring the politically and economically disadvantaged.[127]

The *Defendant Natural Person* coefficient is not significant in either direction. Arguably, given the well established disadvantages of the underdog in litigation this null result may be a product of the collision of a fair use underdog effect and a more general litigation overdog effect. There are, however, many other possible explanations for the failure to find an effect, including: selection effects, the relatively small size of the study population and the potential for omitted variable bias.

Defendant Attorney Underdog produces a negative and significant coefficient in Models 2 and 4. This suggests that defendants who come to court with less experienced attorneys than their opponents are less likely to benefit from a finding of fair use. Again, this is consistent with the expectation that access to better lawyers should lead to better results, but it is inconsistent with the hypothesis that fair use favors the underdog. It should be

---

[125] Interestingly there are no significant correlations between these factors and the partial copy variable.

[126] *See supra* notes 103 to 109 and accompanying text.

[127] *See supra* Part II-B discussing the origin of this view.

noted that this finding is not particularly strong, however, given that it only holds at the .10 level of significance in the full specification in Model 4.[128]

The results in relation to *Plaintiff Natural Person* and *Defendant Attorney Underdog* are contrary to the underdog theory of fair use. If fair use indeed favored the underdog, one would expect to see natural person plaintiffs subjected to fair use less often and defendant underdogs prevailing on the issue of fair use more often. However, the regression analysis reported here suggests exactly the opposite; it confirms, if anything, the existence of an overdog effect in relation to fair use litigation.

3.   Industry Effects

As the table indicates, the *Plaintiff Industry* variable is significant in Model 3, but not in Model 4, where industry variables are combined with party status variables. *Defendant Industry* is not significant in either model. Further analysis shows that there is a significant negative interaction between *Plaintiff Industry* and *Defendant Natural Person*, *Defendant Attorney Underdog* and *Defendant Law Firm Underdog*.[129] One interpretation of these results is that the appearance of an industry effect in Model 3 is an artifact of other party characteristics captured in Model 4. The industry effect disappears once party status and empowerment characteristics are taken into account. Under this interpretation, there is no evidence to support the industry effects hypotheses, H_{11} and H_{12}. This suggests that patent law scholars may wish to revisit evidence of systemic industry variation in that field and consider whether other factors relating to party status and empowerment are not more relevant than industry.

CONCLUSION

This article has undertaken a comprehensive empirical review of the fair use doctrine in copyright law. This study goes beyond merely synthesizing the content of fair use decisions: reducing the fundamental aspects of fair use doctrine to a set of testable implications allows for an empirical assessment of many of the claims, theories and intuitions that have developed around fair use. In short, it enables an assessment of the degree to which assumptions about fair use are supported by the data.

The evidence from litigated cases analyzed in this article confirms the centrality of transformative use.[130] A transformative work is one which imbues the original "with a

---

[128] The fact that no similar effect is discernable for *Defendant Law Firm Underdog* also suggests some caution before accepting this result as proof of the advantages of retaining more higher status legal representation.

[129] The Plaintiff Industry*Defendant Natural Person interaction term is significant at the .10 level. The Plaintiff Industry*Defendant Attorney Underdog and Plaintiff Industry*Defendant Law Firm Underdog interaction terms are significant at the .05 level.

[130] *See* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994) (holding that the creation of transformative works lies "at the heart of the fair use doctrine's guarantee of breathing space within the

further purpose or different character, altering the first with new expression, meaning, or message."[131] Measured in terms of the variable *Creativity Shift*,[132] it appears that transformative use by the defendant is a robust predictor of a finding of fair use. Likewise, the contrast between the significance of direct commercial use and the insignificance of commercial use overall reinforces the dominance of transformative use over other factors. The empirical evidence presented here contradicts the assessment that transformative use is merely an ex post synonym for fair use.[133]

The results of this study also confirm the significance of the statutory factor which addresses "the amount and substantiality" of the defendant's unauthorized use of the plaintiff's work.[134] Although the effect is not as large as ~~it~~ was ~~for~~ transformative use, there is clear evidence confirming that partial copying weighs in favor of the defendant's assertion of fair use. At a practical level, this reinforces the wisdom of minimizing the extent of unauthorized use when relying on the fair use defense. However, it is also apparent from the data that there are many strong cases of fair use involving copying the entirety of the plaintiff's work. Technologies which rely on digital processing of entire copyrighted works, such as internet search engines and plagiarism detection software, nonetheless present a very strong case for fair use.[135]

The rejection of an erroneous theory is sometimes even more valuable than the confirmation of a valid one. One of the most important findings of this study is that there is no evidence that commercial use reduces the defendant's chance of maintaining a fair use defense. As noted in Part II, the presumption against commercial use expressed in dicta in the Supreme Court's *Sony* decision is difficult to reconcile with private commercial incentives copyright is designed to produce. Ten years later, this presumption was implicitly rejected by the Court in *Campbell*.[136] The evidence contained in this study indicates that the there is no anti-commercial bias in fair use. It follows that the *Sony* presumption should be unequivocally discarded.

Indeed, in addition to rejecting the proposition that commercial use makes fair use less likely, the related view which characterizes the doctrine as a kind of subsidy or a

---

confines of copyright.")

[131] *Id.*

[132] Creative shift is an asymmetry between the works of the plaintiff and the defendant such that one is more creative and the other is more informational.

[133] See e.g., Nimmer, *supra* note 53.

[134] 17 U.S.C. § 107(3).

[135] See e.g., *Perfect 10, Inc. v. Amazon.com, Inc.,* 508 F.3d 1146 (9th Cir. 2007) (finding that the automated processing and display of thumbnails of copyrighted photos as part of a visual search engine was fair use); Kelly v. Arriba Soft Corp., 336 F.3d 811, 816 (9th Cir. 2003) (same); *A.V. v. iParadigms, LLC.,* 562 F.3d 630 (4th Cir. 2009) (finding the automated processing of the plaintiff students' work in defendant's plagiarism detection software was fair use).

[136] Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 584 (1994).

redistributive tool favoring the politically and economically disadvantaged can also be discardedset aside. Regression analysis of the effect of party status variables on the probability of a finding of fair use either failed to find the predicted underdog effect, or found exactly the opposite.

Together, the significance of transformative use, the apparent irrelevance of commercial use, and evidence against the underdog view of fair use, tell us something much about the broader nature of the doctrine. Fair use is not a timid exception to the rights of copyright owners or a compulsory form of copyright charity, it is an integral part of the copyright system. The right to make a fair use of a copyrighted work is equally applicable to Fortune 500 companies as it is to struggling artists. Fair use protects creates breathing space for cultural engagement in the form of reinterpretation and re-mixing of copyrighted content. Just as importantly, fair use also makes it possible for large commercial entities to build tools such as search engines that make the Internet work and to create the platforms such as YouTube and Facebook for sharing individual self-expression.

Is fair use predictable? The evidence presented here that transformative use and partial copying are both strong indicators of fair use makes it difficult to sustain the charge of incoherence and unpredictability. The lack of significance of commercial use and the nature of the work, combined with the subjectivity of market effect, may make fair use appear to be more unpredictable than it really is. The fact that the Copyright Act requires judges to consider these factors even though they have little or no forecasting potential inflates their importance. If we simply put these factors to one side and focus on the significant variables, the fair use doctrine begins to look far more rational.

This point can be illustrated by returning to the predicted probabilities derived from the final regression model in Table 1. The overall predicted probability of a finding in favor of fair use is .38. This number increases progressively as additional favorable facts are added. The existence of a *Creative Shift* increases the predicted probability of fair use to .62. If, in addition, where there was no direct commercial use, the predicted probability mounts rises to .72. Adding the fact of partial copying expands the predicted probability further to .76. Finally, adding the observation that the plaintiff is a natural person increases the predicted probability of a finding in favor of fair use to .88. These predictions are not iron-clad guarantees, they are simply estimations of the approximate size of the effects. There will no doubt continue to be many aspects of fair use litigation that can not be captured in averages, generalizations or statistical approximations. There remains, in other words, a role for legal doctrine, subjectivity and advocacy. AAlthough this study does not indicate that the outcomes of fair use cases are obvious or mechanical, nevertheless it does suggest that fair use is much more than merely the right to hire a lawyer and take one's chances. Properly understood, fair use jurisprudence is fairly useful.

STATISTICAL APPENDIX

Case Selection

The data in this study consists of U.S. Federal District Court cases resulting in a written opinion published in the Lexis database where the issue of fair use was raised and adjudicated.[137] The data includes rulings on preliminary injunctions, motions and cross-motions for summary judgment, bench trials, motions to dismiss and motions for judgment notwithstanding verdict. The data includes all such cases decided between January 1, 1978 (the day the Copyright Act of 1976 came into effect) and December 31, 2006.

Alternative Analysis using Linear Regression

Like Table 1 in the main text, Table 2 (below) reports the results of four separate multivariate regressions analyzing the relationship between a finding in favor of fair use and the various factors discussed in Part II.

Model 1 shows the effect of the statutory factors on fair use. Specifically it explores the relationship between a finding in favor of fair use and the following seven variables: *Creative Shift, Commercial Use, Direct Commercial Use, Creative Work, Unpublished Work, Partial Copy* and *Industry Separation*. The remaining models reported in the table below are variations of the first model. Model 2 adds variables relating to underdog status to the analysis of the core statutory factors: these variables are *Plaintiff Natural Person, Defendant Natural Person, Defendant Attorney Underdog* and *Defendant Law Firm Underdog*. Model 3 combines the statutory factors with the plaintiff and defendant industry variables, *Plaintiff Industry* and *Defendant Industry*. Model 4 contains all three groups of variables and should be considered as the main finding of this study.

The difference between the Table 1 in the main text and Table 2 below is that the former reports logitistic regression specifications whereas the current table reports Ordinary Least Squares (OLS) regressions. The dependant variable of interest in this study is dichotomous—either the defendant's claim of fair use was successful, or it was not. Whereas a linear regression model attempts to predict the values of the outcome variable along straight line, the usual way to model dichotomous dependant outcomes is using logistic regression (or a logit model) to predict the probability that the outcome = 1.

---

[137] A search for the term "fair use" in the date range (01/01/1978) to (12/31/2006) in the Lexis *US District Court Cases* database yields 1001 initial results, however the majority of these are trademark cases and relate to a different fair use doctrine entirely. The additional search term "17 U.S.C. § 107" reduces the population to 319. These cases were reviewed individually to exclude cases where fair use is referred to in passing, but not adjudicated.

Table 2: The Effect of Various Objective Factors on a Finding in Favor of Fair Use (Multivariate OLS Regression)

Doctrinal Basis	Variable	Regression Model Finding of Fair Use			
		(1)	(2)	(3)	(4)
Transformative use	Creative Shift	.301** (.09)	.265** (.09)	.260** (.09)	.255** (.09)
Commercial use	Commercial Use	-.070 (.11)	-.093 (.11)	-.065 (.11)	-.080 (.11)
	Direct Commercial Use	-.158* (.07)	-.152* (.07)	-.137# (.07)	-.145* (.07)
Nature of the Work	Creative Work	-.014 (.07)	-.078 (.07)	-.036 (.07)	-.078 (.07)
	Unpublished Work	.084 (.10)	.025 (.10)	.024 (.09)	.007 (.10)
Amount Copied	Partial Copy	.095 (.07)	.147* (.07)	.099 (.07)	.142* (.07)
Market Effect	Industry Separation	.020 (.06)	-.022 (.06)	.031 (.06)	-.038 (.06)
Underdog	Plaintiff natural person		.391** (.08)		.271** (.09)
	Defendant natural person		.051 (.09)		.042 (.09)
	Defendant attorney underdog		-.185* (.08)		-.168* (.08)
	Defendant law firm underdog		.054 (.09)		.054 (.09)
Industry	Plaintiff Industry			.008** (.00)	.003 (.00)
	Defendant Industry			.002 (.00)	.002 (.00)
	Constant	.402	.402	-.109	.145
	R^2	.105	.205	.155	.213

N = 222. ** Significant at the .01 level, * Significant at the .05 level, # Significant at the .10 level.
OLS coefficients obtained using Robust Standard Errors. Standard errors in parentheses.

Although logisitic regression is usually preferred for evaluating relationship between explanatory variables and dichotomous outcomes, it is helpful to duplicate the analysis using linear regression such as OLS because OLS coefficients describe a constant ratio between the independent and dependant variable, which means that one can compare the substantive impact of two statistically significant variables simply by comparing their size. The same information can be presented graphically to illustrate the relative substantive importance of the various factors. In the figure below, gray shading indicates that the variable was not statistically significant, red shading indicates a significant negative coefficient (i.e., as the value of the variable increases, fair use becomes less likely), and dark blue indicates a significant positive coefficient.

Figure 3 OLS Coefficients Illustrating the Effect on a Finding in Favor of Fair Use
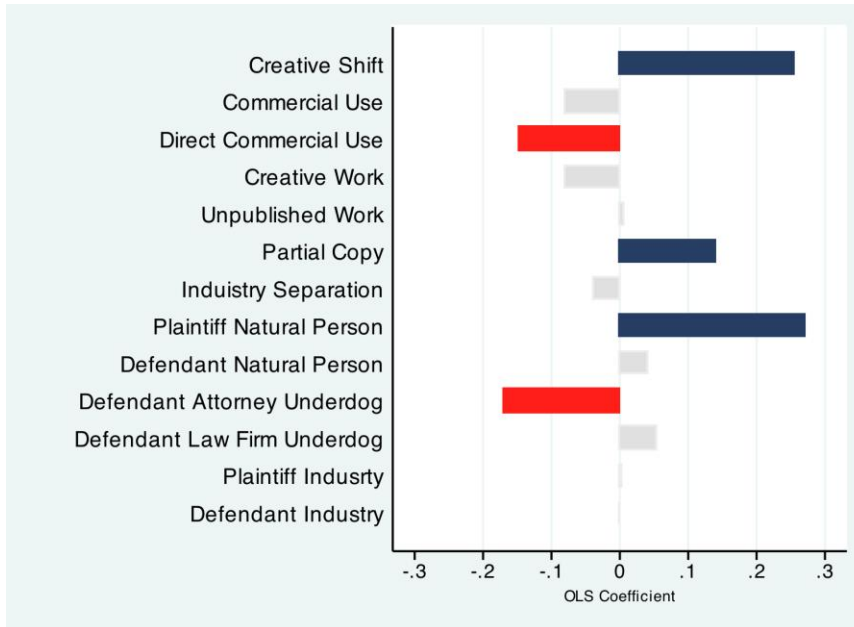
Source: OLS regression 4, Statistical Appendix, Table 2. N = 222. Dark blue coefficients are positive, Red coefficients are negative, Gray coefficients are not statistically significant.

Consistent with the discussion in the main text, it is apparent from Figure 3 that transformative use as indicated by *Creativity Shift* makes a finding of fair use more likely, as does partial copying by the defendant. Comparing the size of these two coefficients, it is clear that the effect for *Creativity Shift* is much stronger. It is also apparent that direct commercial use by the defendant weighs significantly against a finding of fair use, but not by as much as *Creativity Shift* weighs in favor. Note that the coefficient for commercial use in general (i.e., direct and indirect commercial use) is negative as predicted, but is too small to be significant. There is no statistically significant evidence that fair use favors non-commercial actors or disfavors commercial ones. The figure also illustrates that there is no statistically significant evidence that creative or unpublished works are less susceptible to fair use, or that *Industry Separation* (a proxy for market effect) predicts the outcomes of fair use cases.

Beyond the statutory fair use factors, there is no evidence that fair use favors the underdog. In fact, there is statistically significant evidence to the contrary that fair use litigation favors the overdog and it is apparent from Figure 3 that the magnitudes of these party status variables are comparable to the strongest effects seen in the statutory fair use factors. Most notably, the fact that the plaintiff is a natural person is just as important a pro-defendant factor as *Creativity Shift*.

Independent Variables

Table 3 Summary Statistics for Independent Variables

Variable	Mean	Standard Deviation	Range [Min Max]	Mean Finding in Favor of Fair Use if variable =1
Creativity Shift	.189	.392	[0 1]	.619
Commercial Use	.824	.381	[0 1]	.361
Direct Commercial Use	.473	.500	[0 1]	.286
Creative Work (Plaintiff)	.455	.499	[0 1]	.416
Unpublished Work (Plaintiff)	.135	.343	[0 1]	.433
Partial Copy	.419	.494	[0 1]	.441
Industry Separation	.568	.497	[0 1]	.405
Natural Person (Plaintiff)	.288	.454	[0 1]	.609
Natural Person (Defendant)	.167	.374	[0 1]	.459
Defendant Underdog (Attorney Experience)	.167	.374	[0 1]	.243
Defendant Underdog (Law Firm Experience)	.171	.378	[0 1]	.368

N=222.

Statistical Software

The statistical analysis for this study was performed in Stata/SE 11.1.