

# 21:4 BERKELEY TECHNOLOGY LAW JOURNAL

Pages

1017

to

1214

Fall  
2006

**Production:** Produced by members of the *Berkeley Technology Law Journal* on PC computers. All editing and layout is done using Microsoft Word.

**Printer:** Joe Christensen, Inc., Lincoln, Nebraska.  
Printed in the U.S.A.  
The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Library Materials, ANSI Z39.48—1984.

**Copyright © 2006 Regents of the University of California.**  
All Rights Reserved.

*Berkeley Technology Law Journal*  
University of California, Berkeley  
Boalt Hall School of Law  
587 Simon Hall  
Berkeley, California 94720-7200  
(510) 643-6454 (Phone)  
(510) 643-6816 (Fax)  
btlj@law.berkeley.edu  
www.btlj.boalt.org

# BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 21

NUMBER 4

FALL 2006

## TABLE OF CONTENTS

### ARTICLES

- TERRITORIALITY AND INCENTIVES UNDER THE PATENT LAWS: OVERREACHING  
HARMS U.S. ECONOMIC AND TECHNOLOGICAL INTERESTS ..... 1215  
By James R. Farrand
- LAW AND THE SCIENCE OF NETWORKS: AN OVERVIEW AND AN APPLICATION TO THE  
“PATENT EXPLOSION” ..... 1293  
By Katherine J. Strandburg, Gábor Csárdi, Jan Tobochnik, Péter Érdi, & László  
Zalányi
- THE SECRET LIFE OF LEGAL DOCTRINE: THE DIVERGENT EVOLUTION OF SECONDARY  
LIABILITY IN TRADEMARK AND COPYRIGHT LAW ..... 1363  
By Mark Bartholomew & John Tehranian
- DANGEROUS LIAISONS—SOFTWARE COMBINATIONS AS DERIVATIVE WORKS?  
DISTRIBUTION, INSTALLATION, AND EXECUTION OF LINKED PROGRAMS UNDER  
COPYRIGHT LAW, COMMERCIAL LICENSES, AND THE GPL ..... 1421  
By Lothar Determann
- STRIKING OUT “COMPETITIVE BALANCE” IN SPORTS, ANTITRUST, AND  
INTELLECTUAL PROPERTY ..... 1499  
By Salil K. Mehra and T. Joel Zuercher

### NOTES & COMMENTS COMPETITION WINNER

- REVITALIZING THE DOCTRINE OF TRADEMARK MISUSE ..... 1547  
By William E. Ridgway

### CUMULATIVE INDEX

- AUTHOR INDEX ..... 1589
- VOLUME INDEX ..... --
- SUBJECT INDEX ..... --

# SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN 1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited and published four times each year (Spring, Summer, Fall, and Annual Review of Law and Technology) by the students of University of California School of Law, Berkeley (Boalt Hall). Application to Mail at Periodicals Postage Rate is Pending at Berkeley, California, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications Coordinator, 313 Boalt Hall, Boalt Hall School of Law, University of California, Berkeley, CA 94720-7200.

**Correspondence.** Address all correspondence regarding subscriptions, address changes, claims for nonreceipt, single copies, advertising, and permission to reprint to Journal Publications Coordinator, 313 Boalt Hall, Boalt Hall School of Law, Berkeley, CA 94720-7200; (510) 643-6600; journalpublications@law.berkeley.edu. Authors: see section entitled Information for Authors.

**Subscriptions.** Annual subscriptions are \$65.00 for individuals, and \$85.00 for organizations. Single issues are \$27.00. Please allow two months for receipt of the first issue. Payment may be made by check, international money order, or credit card (MasterCard/Visa). Domestic claims for nonreceipt of issues should be made within 90 days of the month of publication; overseas claims should be made within 180 days. Thereafter, the regular back issue rate (\$27.00) will be charged for replacement. Overseas delivery is not guaranteed.

**Form.** The text and citations in the *Journal* conform generally to the UNITED STATES GOVERNMENT PRINTING OFFICE STYLE MANUAL (29th ed. 2000) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 18th ed. 2005). Please cite this issue of the *Berkeley Technology Law Journal* as 21 BERKELEY TECH. L.J. \_\_\_\_ (2006).

## BTLJ ONLINE

The full text and abstracts of many *Berkeley Technology Law Journal* and *High Technology Law Journal* articles published in previous issues can be found at <http://www.btlj.boalt.org>. Our site also contains a cumulative index, general information about the *Journal*, selected materials related to technology law, and links to other related pages. Author, volume, and subject indexes may also be found in Volume 20, Number 4 (2005) of the *Journal*.

# INFORMATION FOR AUTHORS

The Editorial Board of the *Berkeley Technology Law Journal* invites the submission of unsolicited manuscripts. Submissions may include previously unpublished articles, essays, book reviews, case notes, or comments concerning any aspect of the relationship between technology and the law. If any portion of a manuscript has been previously published, the author should so indicate.

**Format.** Authors may submit manuscripts in electronic or hardcopy form, though electronic submissions are strongly encouraged. Electronic submissions should be sent as attachments in Microsoft Word format to [btlj@law.berkeley.edu](mailto:btlj@law.berkeley.edu). Authors should submit double-spaced, single-sided manuscripts with generous margins. We regret that submissions cannot be returned. Authors should retain an exact copy of any material submitted.

**Citations.** All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 18th ed. 2005). In addition, the author should include his or her credentials, including full name, degrees earned, academic or professional affiliations, and citations to all previously published legal articles.

**Copyrighted Material.** If a manuscript contains any copyrighted table, chart, graph, illustration, photograph, or more than eight lines of text, the author must obtain written permission from the copyright holder for use of the material. A photocopy of such written permission should accompany the submission.

**Mailing Address.** BTLJ highly prefers electronic submissions sent as Microsoft Word attachments to [btlj@law.berkeley.edu](mailto:btlj@law.berkeley.edu), but also accepts hardcopy manuscripts sent to:

Submissions Editor  
*Berkeley Technology Law Journal*  
University of California, Berkeley  
Boalt Hall School of Law  
587 Simon Hall  
Berkeley, California 94720  
(510) 643-6454 (Phone)

# DONORS

The *Berkeley Technology Law Journal* acknowledges the following generous donors to Boalt Hall's Law and Technology Program:

## Benefactors

COOLEY GODWARD LLP

DLA PIPER RUDNICK GRAY CARY

FARELLA BRAUN + MARTEL LLP

FENWICK & WEST LLP

LATHAM & WATKINS LLP

NIXON PEABODY LLP

ORRICK, HERRINGTON  
& SUTCLIFF LLP

SKADDEN, ARPS, SLATE, MEAGHER  
& FLOM LLP

WEIL, GOTSHAL & MANGES LLP

WILSON SONSINI GOODRICH  
& ROSATI PC

## Members

AKIN GUMP STRAUSS HAUER & FELD LLP	HOWREY LLP
ALSCHULER GROSSMAN STEIN & KAHAN LLP	JONES DAY KIRKLAND & ELLIS LLP
BAKER BOTTS LLP	KNOBBE, MARTENS, OLSON & BEAR LLP
BINGHAM MCCUTCHEM LLP	MAYER, BROWN, ROWE & MAW LLP
COVINGTON & BURLING	MCDERMOTT, WILL & EMERY
DAVIS POLK & WARDWELL	MORGAN, LEWIS & BOCKIUS LLP
DAY CASEBEER MADRID & BATCHELDER LLP	MORRISON & FOERSTER LLP
FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP	O'MELVENY & MYERS LLP
FISH & RICHARDSON P.C.	ROPES & GRAY LLP
GUNDERSON DETTMER STOUGH VILLENEUVE FRANKLIN & HACHIGIAN LLP	TOWNSEND AND TOWNSEND AND CREW LLP
HELLER EHRMAN LLP	WHITE & CASE LLP

## Patrons

BAKER & MCKENZIE	KENYON & KENYON
GREENBERG TRAUIG LLP	MARGER JOHNSON & MCCOLLOM PC
HICKMAN PALERMO TRUONG & BECKER LLP	TRELLIS IP LAW GROUP
KEKER & VAN NEST LLP	VAN PELT, YI & JAMES LLP

The *Berkeley Technology Law Journal* is a nonprofit organization and welcomes donations. Donors are recognized appropriately for their contributions. For more information, contact the *Berkeley Technology Law Journal*, 587 Simon Hall, Boalt Hall School of Law, University of California, Berkeley, California 94720, (510) 643-6454, or e-mail [btlj@law.berkeley.edu](mailto:btlj@law.berkeley.edu).

# ADVISORY BOARD

ROBERT BARR

*Executive Director of the Berkeley Center  
for Law & Technology*  
Boalt Hall School of Law  
Berkeley, California

ROBERT C. BERRING, JR.

*Walter Perry Johnson Professor of Law*  
Boalt Hall School of Law  
Berkeley, California

JESSE H. CHOPER

*Earl Warren Professor of Public Law*  
Boalt Hall School of Law  
Berkeley, California

PETER S. MENELL

*Professor of Law and  
Director of the Berkeley Center for  
Law & Technology*  
Boalt Hall School of Law  
Berkeley, California

ROBERT P. MERGES

*Wilson Sonsini Goodrich & Rosati  
Professor of Law and Director of the  
Berkeley Center for Law & Technology*  
Boalt Hall School of Law  
Berkeley, California

REGIS MCKENNA

*Chairman and CEO*  
Regis McKenna, Inc.  
Palo Alto, California

DEIRDRE K. MULLIGAN

*Director of the Samuelson Law,  
Technology & Public Policy Clinic and  
Clinical Professor of Law*  
Boalt Hall School of Law  
Berkeley, California

JAMES POOLEY

*Pooley & Oliver  
Palo Alto, California*

MATTHEW D. POWERS

Weil, Gotshal & Manges LLP  
Redwood Shores, California

PAMELA SAMUELSON

*Professor of Law & Information  
Management and Director of the Berkeley  
Center for Law & Technology*  
Boalt Hall School of Law  
Berkeley, California

LIONEL S. SOBEL

*Professor of Law and Director of the  
International Entertainment & Media Law  
Summer Program in London, England*  
Southwestern University School of Law  
Los Angeles, California

LARRY W. SONSINI

Wilson Sonsini Goodrich & Rosati  
Palo Alto, California

MICHAEL STERN

Cooley Godward LLP  
PALO ALTO, CALIFORNIA

MICHAEL TRAYNOR

Cooley Godward LLP  
San Francisco, California

THOMAS F. VILLENEUVE

Gunderson, Dettmer, Stough, Villeneuve,  
Franklin & Hachigian, LLP  
Menlo Park, California

# BOARD OF EDITORS

# 2006-2007

---

*Editor-in-Chief*  
CORRIN DRAKULICH

*Managing Editor*  
YAS RAOUF

*Senior Article Editors*  
TIMOTHY P. BEST  
ANNA ZICHTERMAN

*Senior Executive Editor*  
TASHICA WILLIAMS

*Senior Annual Review Editors*  
A.H. RAJANI  
ALISON WATKINS

---

*Submissions Editors*  
ALAN GALLOWAY  
JEFF THOMAS

*Production Editor*  
LIYING SUN

*Symposium Editors*  
SAJJAD MATIN  
SARALA NAGALA

*Notes & Comments Editor*  
DANIEL DOBRYGOWSKI

---

*Article Editors*

JAN-MICHAEL COHEN  
MICHAEL DAVIS-WILSON  
JAYNI FOLEY  
SARAH GETTINGS

GALEN HANCOCK  
DAVID JACKSON  
JEFFREY KUHN  
JEREMY MULDER

DEANA SOBEL  
WINSTON SU  
NIKI WOODS  
CHRISTOPHER YEH

---

*Executive Editors*

PAUL FINLAYSON  
BOBBY GLUSHKO  
YASER HERRERA

PUNEET KAKKAR  
SHAKTI NARAYAN

LORI SANTOS  
NATALIE SKORDILIS  
MARGARET THOMSON

# MEMBERSHIP

# 2006-2007

---

## *Associate Editors*

---

DANIEL DOBRYGOWSKI  
TOM FLETCHER  
ALAN GALLOWAY  
ROBERT GLUSHKO  
JING GU  
ERIC HANSON

THOMAS KING  
JEFFREY KUHN  
JENNIE LEVIN  
SHAKTI NARAYAN  
JEFF THOMAS

ANNA TSIFTSOGLU  
JULIA WEINERT  
STEPHANIE WELLS  
NIKI WOODS  
VIRGINIA ZAUNBRECHER

---

## *Members*

---

PARKER ANKRUM  
JEREMY BUXBAUM  
LAURA COHEN  
NATE DAVIS  
PAUL FINLAYSON  
JAYNI FOLEY  
ALICIA GÁMEZ  
SARAH GETTINGS  
BLAKE GREENE  
ANDREA GROVES  
SANDY HABERNY  
ROBERT HAMILTON  
YASER HERRERA  
DANA HOWELLS  
DAVID JACKSON  
CAROL JOHNS  
FAITH JONES

PUNEET KAKKAR  
JOSH KAMZAN  
THOMAS KUA  
JENNIFER LANE  
KEVIN MACK  
KAMILLA MAMEDOVA  
SAJJAD MATIN  
KRISTY MURPHY  
JOSEPHINE MUTUGU  
SARALA NAGALA  
WILLIAM NEWMARK  
JASON RICHARDSON  
EDUARDO ROCHA  
ADAM ROSENBLUM  
LORIANI SANTOS  
MICHAEL SAUNDERS  
JASON SAVICH  
DANIEL SCHACHT

AARON SCHOHN  
BOBAK SIAVOSHY  
SARAH SIMMONS  
NATALIE SKORDILIS  
DEANA SOBEL  
JESSE SOLOMON  
M. LEAH SOMOANO  
WINSTON SU  
LILA SUBRAMANIAN  
DIRAN TASHJIAN  
MARGARET THOMSON  
MORRIS TU  
CECILIA WALSH  
CHRIS WESTBERG  
CHRISTOPHER YEH  
ISSAC ZAUR  
KAI ZHU

# TERRITORIALITY AND INCENTIVES UNDER THE PATENT LAWS: OVERREACHING HARMS U.S. ECONOMIC AND TECHNOLOGICAL INTERESTS<sup>1</sup>

By James R. Farrand<sup>†</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1216
II. ORIGINS AND ENACTMENT OF SECTION 271(F) .....	1219
A. DEEPSOUTH AND TRADITIONAL PRINCIPLES OF TERRITORIALITY IN PATENT LAW .....	1220
B. ENACTMENT OF SECTION 271(F) .....	1225
III. APPLICATION OF SECTION 271(F)—UNCERTAINTIES, JUDICIAL DECISIONS, AND THEIR EFFECTS .....	1232
A. APPLICABILITY TO CHEMICAL COMBINATIONS.....	1232
B. APPLICABILITY TO PATENTED METHODS/PROCESSES.....	1233
C. COMPUTER SOFTWARE—INTANGIBLES AS “COMPONENTS” AND “COPIES” MADE ABROAD AS “SUPPLIED” FROM THE U.S.....	1239
1. <i>Typical Software Fact Pattern</i> .....	1241
2. <i>Did the Software Qualify as a Section 271(f) “Component”?</i> ...	1242
3. <i>Did the Software Copies Made Abroad Qualify as Components             Supplied from the United States?</i> .....	1245
D. COMPONENTS DESIGNED BUT NOT “PRODUCED” IN THE UNITED STATES—U.S. COMPANIES MOVING PRODUCTION OFF-SHORE.....	1250
E. EFFECTS OF SECTION 271(F) EXTENSIONS.....	1252
IV. ASSESSING SECTION 271(F).....	1254
A. DO WE HAVE SUFFICIENT INFORMATION ON WHICH TO BASE AN EVALUATION OF SECTION 271(F)?.....	1255

---

1. By mistake and without the author’s knowledge or permission, the Journal of the Patent & Trademark Office Society published an incomplete version of the July 2006 draft of this article in its September 2006 issue, 88 JPTOS 761-795. The Patent & Trademark Office Society has apologized for this mistake and is publishing that apology in the November 2006 issue of the JPTOS.

© 2007 James R. Farrand

† Mr. Farrand is Of Counsel in the Los Angeles office of Arnold & Porter LLP, and teaches Patent Litigation at Loyola Law School, Los Angeles. The views expressed here are those of Mr. Farrand and not of others at Arnold & Porter LLP or of the firm itself.

B.	BENEFITS—DOES SECTION 271(F) HELP INCENTIVIZE INNOVATION?.....	1259
C.	RISKS AND BURDENS FOR U.S. PRODUCERS.....	1266
	1. <i>Strengthening and Proliferation of U.S. Patents</i> .....	1267
	2. <i>Uncertainty as to Patent Coverage and Validity</i> .....	1270
	a) Complexity of Patent Claims and Defenses .....	1271
	b) Unknown Patents.....	1272
	c) Poor Patent “Quality” .....	1273
	3. <i>Technological and Industrial Trends—International Sourcing of “Components” and Product Complexity</i> .....	1274
	4. <i>Likely Injunctions</i> .....	1275
D.	OTHER POSSIBLE ARGUMENTS IN FAVOR OF SECTION 271(F) .....	1278
	1. <i>Knowledge Elements and Unfairness</i> .....	1278
	2. <i>Efficiency in Patent Enforcement</i> .....	1281
E.	BOTTOM LINE ASSESSMENT .....	1283
V.	<b>FUTURE APPLICATIONS OF SECTION 271(F) AND SPECIAL ASPECTS OF THAT PROVISION IN CONTEMPORARY PATENT ENFORCEMENT</b> .....	1284
	A. HIGH-VALUE COMPONENTS AND INFORMATION COMPONENTS .....	1285
	B. GENOMICS AND STEM CELLS.....	1285
	C. PATENT TROLLS.....	1286
VI.	<b>CONCLUSION</b> .....	1290

## I. INTRODUCTION

The U.S. patent laws are all about incentives. They create incentives for invention and for detailed public disclosure of inventions. Equally important, they incentivize investment to commercialize the resulting products and processes. We expect these incentives to promote U.S. prosperity.

Surprisingly, a 1984 addition to the U.S. Patent Act creates substantial perverse incentives. Sections 271(f)(1) and (f)(2) of the Act (collectively, “Section 271(f)”) reach beyond the traditional territorial limits of patent law. More important, from a practical standpoint, they do so in ways that harm U.S. producers and the U.S. economy while failing to serve any U.S. interests.

Section 271(f) establishes a new type of “infringement” to protect holders of U.S. patents against a broadly defined form of competition from U.S. producers in *foreign* markets. Section 271(f) allows the patent holder to forbid everyone else from exporting *unpatented* components for inclusion abroad in combination products of the type claimed in the U.S. patent. Under Section 271(f), such export generally constitutes infringement

even though the combination product will not only be made, but also sold and used, strictly abroad and will never be brought into the U.S.

A key point here is that U.S. laws cannot reach competition by *foreign* producers in *foreign* markets. Accordingly, Section 271(f) restricts only U.S. suppliers of the relevant components. Effectively discriminating against U.S. producers, it leaves all off-shore producers—including U.S. companies that choose to produce the components off-shore or to contract for their off-shore production—free of any restraints.

This discriminatory effect escaped criticism when Congress enacted Section 271(f), and it would not have seemed very serious in the 1970s and '80s where that provision originated. Commercial and technological developments since that time, however, have made this effect serious and unmistakable. Recent court decisions have extended the application of Section 271(f) and magnified its negative consequences. Newly published studies of the U.S. patent system further illuminate the shortcomings of that provision.

As detailed below, Section 271(f) has an extraordinarily bad cost/benefit profile: it fails to enhance the incentives for innovation or investment while creating serious risks for companies that produce, or consider producing, technologically advanced components in the U.S. for world markets. In the last two years, that provision has produced huge, highly publicized awards of damages—based on worldwide product sales—and similarly broad injunctive orders. During the same period, another prominent decision emphasized the safety of producing the same products anywhere in the world *except* in the U.S. It is hard to imagine a worse combination of patent-based incentives from the standpoint of U.S. employment, U.S. technological and productive capacity, and the U.S. economy.

The potential beneficiaries of Section 271(f) have also shifted since its enactment. That provision gives added leverage and increased potential damages awards to holders of certain types of U.S. patents. When Congress enacted Section 271(f), U.S. entities held the great majority of outstanding U.S. patents and were receiving the great majority of newly issued U.S. patents. Now, though, non-U.S. entities receive and hold roughly half of all U.S. patents being issued. The reverse is true, however, as to the *targets* of Section 271(f) claims. Section 271(f) burdens only U.S. producers, i.e., entities that make products, employ persons, and pay taxes in the U.S.—those that help maintain our technological strength. Accordingly, horribly ironic results seem likely in the near future under Section 271(f). For example, a Japanese company owning a U.S. patent could use Section 271(f) to bar U.S. producers of non-patented components from

markets around the world even though producers of the same components in Japan, China, and other countries would remain free to compete in world markets.<sup>2</sup>

Some of the “patent reform” proposals circulated on Capitol Hill in the last two years would have overruled recent judicial extensions of Section 271(f) or repealed the provision entirely. Likewise, a case now pending before the U.S. Supreme Court could narrow the application of that provision in some respects.<sup>3</sup> All of the substantial patent reform proposals, however, died with the recently completed session of Congress. Likewise, even if the Supreme Court limits the application of Section 271(f) as much as possible in the case now pending before it, a broad range of present and prospective U.S. producers will remain subject to the discriminatory burdens and perverse incentives that provision creates.

Section 271(f) discourages investment in U.S. productive facilities, production of technologically advanced components in the U.S., and the employment of workers in U.S. technology sector jobs. This tends to undermine U.S. technological strength. At the same time, as we shall see, it produces no incremental incentives for innovation. Moreover, better, non-discriminatory alternatives are available to provide patent protection in foreign markets. No other country seems to have a law similar to Section 271(f). Congress should recognize the wisdom of all other patent-granting countries and repeal Section 271(f). Until that happens, U.S. courts should cut-back its operation as much as possible. Traditional principles of statutory interpretation, if used in place of the Federal Circuit’s recent avowedly expansionist approach, will support some of the needed cut-backs.

Part II of this Article reviews the background and origins of Section 271(f), including its departure from the traditionally territorial character of patent laws and the unusual circumstances that created an apparent justification for such a provision in the first place. This Part also examines the limited legislative history and stated rationale for the provision and explains why its application has remained unclear in several important respects for more than twenty years.

Part III then examines the major judicial extensions of Section 271(f) in the last two years. These reflect the Federal Circuit’s recent abandonment of significant limitations previously thought to narrow the application of Section 271(f). Part III also recounts the very recent objections of several Federal Circuit judges to some of these extensions, and it describes

---

2. See *infra* text accompanying note 160.

3. *AT&T Corp. v. Microsoft Corp.*, 414 F.3d 1366 (Fed. Cir. 2005), *cert. granted*, 127 S. Ct. 467 (2006).

the case now pending before the Supreme Court which presents one or two of these extensions for high-court review. Finally, Part III describes: (i) the huge economic impact Section 271(f) can have on U.S. producers and (ii) a recent Federal Circuit decision recognizing a particular limitation on the reach of Section 271(f), which, unfortunately, merely increases the discriminatory effects and perverse incentives that provision creates.

Part IV assesses the merits and faults of Section 271(f). It first considers the adequacy, *vel non*, of the available information to support such an assessment. Part IV then reviews the extent to which Section 271(f) furthers the objectives of the patent laws, the objectives of U.S. trade and economic policy, or other widely shared objectives such as efficiency and perceived fairness in patent enforcement. Part IV also considers whether Section 271(f) significantly burdens U.S. production and examines the incentives it creates for U.S. and foreign companies planning investments in productive facilities. Part IV concludes that Section 271(f) fails to further the objectives of the patent laws or other U.S. interests while burdening U.S. producers and creating all the wrong incentives with respect to investment, production, employment, and technological capacity. Finally, Part IV explains why Supreme Court review or Federal Circuit en banc review could modestly reduce, but could not come close to eliminating, the ill effects of Section 271(f).

The concluding Part V notes the potential application of Section 271(f) in some new areas if Congress and the courts fail to repeal it or substantially limit its scope. Part V also briefly addresses the special usefulness of Section 271(f) for a particular category of patent enforcers derisively termed “patent trolls.”

## II. ORIGINS AND ENACTMENT OF SECTION 271(F)

According to its proponents, Congress enacted Section 271(f) to close a “loophole” in the U.S. patent laws that was revealed in the U.S. Supreme Court’s 1972 decision in *Deepsouth Packing Co. v. Laitram Corp.*<sup>4</sup> *Deepsouth* displayed the long-recognized principle of territoriality in patent law operating in extreme and unusual circumstances. The Justices split 5-4 on the result, and the outcome seemed harsh and hyper-technical. That decision, together with a major shift in U.S. patent protection in the early 1980s, led to the enactment of Section 271(f).

---

4. *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518 (1972).

### A. *Deepsouth* and Traditional Principles of Territoriality in Patent Law

From the earliest days of U.S. patent law, territorial principles have strictly confined the offense of infringement. Unauthorized exploitation of a claimed invention could infringe but only if it took place within the United States, including its territories and possessions.<sup>5</sup> The main infringing acts were, and continue to be: (i) making, using, and/or selling a claimed product *in the U.S.*; (ii) practicing a claimed process *in the U.S.*; and (iii) importing claimed items *into the U.S.* from abroad.<sup>6</sup>

Over many years, numerous courts and scholars have emphasized this territorial principle.<sup>7</sup> Similar territorial principles govern the patent laws of

---

5. 35 U.S.C. §§ 154(a)(1), 271(a), (c), (e), (g) (2000 & Supp. III 2003); *see* 35 U.S.C. § 100(c) (2000).

6. 35 U.S.C. §§ 154, 271 (2000 & Supp. III 2003). Likewise, in codifying the related common law doctrines of secondary liability for “inducing” infringement and “contributory infringement,” Congress limited their application with the phrase “within the United States” or a variation thereof. *See* 35 U.S.C. § 271(b), (c). The 1988 Process Patent Amendment Act added Section 271(g) to the Patent Act, making the importation into the U.S. of a product made abroad by a process patented in the U.S. an act of infringement. Subsequently, pursuant to the Trade-Related Aspects of Intellectual Property (“TRIPS”) agreement of the 1994 Uruguay Round General Agreement on Tariffs and Trade (“GATT”), Congress amended 35 U.S.C. §§ 154 and 271 to add “import[ing the invention] into the United States” as an act of infringement. *See* 35 U.S.C. §§ 154, 271 (2000 & Supp. III 2003); World Trade Organization, A Summary of the Final Act of the Uruguay Round, [http://www.wto.org/english/docs\\_e/legal\\_e/ursum\\_e.htm#nAgreement](http://www.wto.org/english/docs_e/legal_e/ursum_e.htm#nAgreement) (last visited Nov. 7, 2006). The same legislation added “offering for sale” a claimed invention in the U.S. as a further act of infringement. *See* §§ 154, 271.

7. *See Deepsouth*, 406 U.S. at 531. The Court wrote:

Our patent system makes no claim to extraterritorial effect; “these acts of Congress do not, and were not intended to, operate beyond the limits of the United States,” . . . and we correspondingly reject the claims of others to such control over our markets. . . . To the degree that the inventor needs protection in markets other than those of this country, the . . . congressional intent [was] to have him seek it abroad through patents secured in countries where his goods are being used.

*Id.* (citations omitted); *see also* MEMC Elec. Materials, Inc. v. Mitsubishi Materials Silicon Corp., 420 F.3d 1369, 1375 (Fed. Cir. 2005); Rotec Indus. Inc. v. Mitsubishi Corp., 215 F.3d 1246, 1251 (Fed. Cir. 2000) (“These extraterritorial activities . . . are irrelevant to the case before us, because [t]he right conferred by a patent under our law is confined to the United States and its territories, and infringement of this right cannot be predicated of acts wholly done in a foreign country.” (quoting Dowagiac Mfg. Co. v. Minnesota Moline Plow Co., 235 U.S. 641, 650 (1915))); U.S. Patent & Trademark Office, Treaties and Foreign Patents, <http://www.uspto.gov/web/offices/pac/doc/general/treaties.htm> (last visited Nov. 3, 2006) (“Since the rights granted by a United States patent extend only throughout the territory of the United States and have no effect in a foreign country, an

other countries.<sup>8</sup> These principles derive support, in part, from *comity*—i.e., respect for the sovereignty and laws of other countries—and, in part, from national self-interest and the exclusionary nature of patent rights as to the territory in which they apply.<sup>9</sup> No country wants the laws of another country to limit commercial activity within its borders, and countries almost never choose to restrict the commercial activities of, or the benefits available to, their own citizens more than the competing activities of, or corresponding benefits for, citizens of other countries.<sup>10</sup> Another explanation, of course, is the difficulty a country ordinarily experiences in enforcing limitations on commercial activities outside its own borders. In short,

---

inventor who wishes patent protection in other countries must apply for a patent in each of the other countries or in regional patent offices.”).

8. See, e.g., Patents Act, 1977, c. 37, § 60(1) (Eng.) (“[A] person infringes a patent for an invention if, but only if, . . . he does any of the following things in the United Kingdom in relation to the invention without the consent of the proprietor of the patent. . . .”); Paris Convention for the Protection of Industrial Property, Mar. 20, 1883; RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW § 415, cmt. i (1987) (“Patents are considered territorial, having legal effect only in the territory of the issuing state.”); GRAEME B. DINWOODIE, WILLIAM O. HENNESSEY, & SHIRA PERLMUTTER, INTERNATIONAL AND COMPARATIVE PATENT LAW § 1.03, 30 (2002) (“[T]he starting point for any study of international patent law [is that] patent laws operate territorially, and patent rights are thus national in scope.”).

9. A patent confers the right to exclude others from making, using, or selling the inventory in the territory to which that right extends. See provisions *supra* notes 5, 6.

10. See, e.g., Paris Convention for the Protection of Industrial Property, Mar. 20, 1883, as amended, Articles 2 and 3, recognizing the territorial application of the patent laws of each of the countries that are signatories to that convention and requiring them to accord the citizens of other countries equal rights in invoking their patent laws as to their respective territories:

Article 2:

(1) Nationals of any country of the Union shall, as regards the protection of industrial property [i.e., patents, etc.], enjoy in all the other countries of the Union the advantages that their respective laws now grant, or may hereafter grant, to nationals; . . . . Consequently, they shall have the same protection as the latter, and the same legal remedy against any infringement of their rights. . . .

(2) . . . [No] requirement as to domicile or establishment in the country where protection is claimed may be imposed upon nationals of countries of the Union for the enjoyment of any industrial property rights.

Article 3:

Nationals of countries outside the Union who are domiciled or who have real and effective industrial or commercial establishments in the territory of one of the countries of the Union shall be treated in the same manner as nationals of the countries of the Union.

the territorial limitations on patent protection are based not only on tradition but also on considerations of national self-interest, sovereignty, and practicality.

Territorial principles can affect patent claims in various ways.<sup>11</sup> Further, like most legal principles, territoriality in patent law can produce seemingly harsh results when applied in unusual circumstances. This is what happened in *Deepsouth*, and that case became the “poster-child” scenario on which Section 271(f) was based.

In *Deepsouth*, Laitram Corporation held a U.S. patent on a machine for removing shells and veins from shrimp. Laitram also held foreign patents on those machines which, for unexplained reasons, it did not discuss or utilize. Deepsouth Packing Company also held U.S. patents for, and produced and sold, machines for de-veining shrimp. In litigation, Laitram’s U.S. patent claims prevailed over those of Deepsouth, and Laitram obtained damages and an injunction against Deepsouth’s production and sale of its de-veining machines.

Cleverly, Deepsouth sought a modification of the injunction to allow it: (i) to continue producing *all* the components of its infringing machines at its factory in the United States; (ii) to assemble those parts in incomplete form in the United States; and (iii) to receive and accept orders for, and ship sets of, those partially assembled machines to foreign countries

---

11. A number of cases, including the recent, highly publicized BlackBerry litigation, provide interesting examples of products otherwise infringing a U.S. patent by sale or use in the U.S. but relying on a component located, or a step performed, outside the United States. Under the Federal Circuit’s recent ruling in that case, a finding of infringement in this context depends on several technical distinctions including: (i) the type of infringing acts alleged and proved (“using” vs. “making/selling” the claimed invention) and (ii) the type of patent claims (“machine” claims vs. “process” claims) at issue. *NTP, Inc. v. Research in Motion, Ltd.*, 418 F.3d 1282, 1316-20 (Fed. Cir. 2005), *reh’g and reh’g en banc denied*, No. 03-1615, 2005 U.S. App. LEXIS 23112 (Fed. Cir. Oct. 7, 2005), *cert. denied*, No. 05-763, 126 S. Ct. 1174 (Jan. 23, 2006). Specifically, the Federal Circuit held that infringement based on *using* or *practicing* the invention: (i) cannot exist as to method (process) claims if any claimed step is performed outside the United States, but (ii) can exist as to system (machine) claims where one of the claimed physical elements is outside the United States as long as the system as a whole is controlled and beneficially used by persons located in the United States. *Id.*

This Article does not address the BlackBerry-type situation because it involves the application of U.S. patents to limit competition in U.S. markets. Conceptually, this is a modest form of extraterritoriality, if it involves extraterritoriality at all; courts can apply the normal remedies for infringement without any special enforcement difficulties, without intrusion into competition in foreign markets, and without discrimination against U.S. producers. Instead, this Article considers the much more aggressive and problematic type of extraterritoriality that Section 271(f) calls for, namely applying a U.S. patent to restrict competition with its holder in foreign markets.

for final assembly and use at those foreign locations. Related circumstances made these steps even more obviously evasive than they might otherwise have been.<sup>12</sup> In Deepsouth's favor, however, none of the parts of the de-veining machines was new or patentable when Laitram applied for its patent, so the patent covered only the complete combination of all the claimed parts. Further, under Deepsouth's proposed modification, there clearly would be no "use" of the patented combinations in the United States. Thus, the case turned on whether the proposed activities would involve "making" or "selling" the complete patented machines *in the United States*.

Drawing on prior case law, a 5-4 majority of the Justices held that "making" or "selling" a machine in the United States required a *complete and operable assembly* of the machine here. Accordingly, the majority held that Deepsouth's continued production, sale, and shipment of sub-assemblies would not infringe Laitram's U.S. patent. The Court granted Deepsouth's requested modification of the injunction and stated that Laitram would have to rely on its foreign patents to protect it from the proposed type of competition by Deepsouth in foreign markets.<sup>13</sup>

Both the majority and the dissenters recognized larger considerations at work in this case. Their differing views illuminate the policy considerations underlying Section 271(f).

Perceptions of "unfairness" drove much of the dissenters' assessment. They stressed what they considered Deepsouth's "evasive" and "iniquitous" conduct in first infringing Laitram's U.S. patent and then trying to skate narrowly around it.<sup>14</sup> More significant, for present purposes, the dissenters acknowledged their disregard for traditional territorial limitations on the monopoly rights conferred by the U.S. patent laws. This disregard was appropriate, they argued, for two reasons. First, extending the reach of the U.S. patent laws to prevent U.S. producers from competing in foreign markets would protect patent holders' sales there and therefore encourage

---

12. Deepsouth shipped each set from the U.S. in three large boxes, each containing a sub-assembly. Together, the boxed sub-assemblies weighed 1.5 tons and comprised a complete de-veining machine. Deepsouth's literature generally referred to what it shipped as "de-veining machines," rather than "parts" or "sub-assemblies," and it said installation of the items as a complete operable machine at the foreign location would take "less than an hour." Deepsouth charged the same price for these partially assembled sets as it had charged for the fully assembled machines.

13. *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518, 527-31 (1972).

14. *Id.* at 533-34. See *infra* text accompanying notes 150-160 for a discussion of the fairness considerations relating to Section 271(f).

innovation—a highly doubtful proposition for reasons discussed below.<sup>15</sup> Second, they asserted a proposition that may have been plausible in the early 1970s but now seems entirely outdated: that U.S. manufacturers generally enjoyed economic advantages relative to their foreign competitors that justified subjecting them to restrictions in international markets not applicable to their non-U.S. competitors.<sup>16</sup> The dissenters also assumed that the holders of U.S. patents were U.S. enterprises. This was true as to 75% of all issuing U.S. patents (and a higher percentage of outstanding U.S. patents) at the time of *Deepsouth* but is now true only about 50% of the time.<sup>17</sup>

In contrast, the five-Justice majority endorsed the traditional view that each country's patent laws govern patent protection in that country's markets. They emphasized that U.S. inventors can secure patent protection in foreign markets by obtaining foreign patents in the countries where protection is desired. The majority agreed with *Deepsouth* that Laitram's U.S. patent gave Laitram "a monopoly only over the U.S. market" but did not confer on Laitram "the bonus of a favored position as a flagship company free of American competition in international commerce."<sup>18</sup>

---

15. *Id.* at 534. See *infra* text accompanying notes 107-120 regarding the doubtful nature of this proposition.

16. *Id.* When the Court decided *Deepsouth*, many U.S. manufacturers enjoyed vastly stronger positions in relation to their foreign competitors than they do today. For example, General Motors had slightly more than 50% of the U.S. auto market (as opposed to approximately 23% in recent periods) in addition to strong positions in locomotives, refrigerators, and foreign auto sales. Similarly, IBM had market shares in computers that varied depending on the market definition but were often in the range of 90%.

17. *Id.* at 532-34. In 1970, residents of foreign countries received only about 25% of the total U.S. patents granted, but in the mid-1980s this percentage had risen to over 30%. Remarks of Representative Kastenmeier, 130 CONG. REC. H10522-34 (daily ed. Oct. 1, 1984) (statement of Rep. Kastenmeier); 130 CONG. REC. H12231-32 (daily ed. Oct. 11, 1984) (statement of Rep. Kastenmeier). According to the U.S. Patent and Trademark Office's ("PTO") statistics, foreign inventors accounted for 20% of the U.S. patents granted in the mid-1960s and for 35% of all patents granted from 1963 through 1989; but that percentage grew steadily and reached 48% in 2002-2005. See, e.g., U.S. Patent Statistics Summary Table, [http://www.uspto.gov/go/taf/us\\_stat.htm](http://www.uspto.gov/go/taf/us_stat.htm) (last visited Nov. 4, 2006); U.S. Patent Statistics Reports, <http://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm> (last visited Nov. 4, 2006). Similarly, foreign individuals and entities owned 33% of all patents granted in 1963 through 1989, but that percentage grew to about 48% in the period from 2002 through 2005. The percentage of U.S. patents being issued to foreign inventor applicants now equals, or will soon equal, 50%. *Id.*

18. *Deepsouth*, 406 U.S. at 523. Criticizing the idea that the holder of a U.S. patent should be the only U.S. producer allowed to compete in foreign markets, the majority stated:

The majority's emphasis on foreign patents as the way to secure protection in foreign markets was non-controversial, and Laitram's failure to rely on its foreign patents remained unexplained. Nevertheless, the majority's narrow definitions of "make" and "sell" seemed hyper-technical given the extreme circumstances of the *Deepsouth* case, and Deepsouth's tactics seemed evasive and unfair. In addition, competition from non-U.S. manufacturers was less intense in 1972 than it has since become, and the holders of U.S. patents were more often U.S. entities. Thus, it may have seemed plausible, as the dissenters argued, that granting U.S. patent protection as to foreign markets—but only against U.S. producers—could provide meaningful protection in those markets for holders of U.S. patents and increase the incentives for U.S. innovation, all without harming U.S. interests. As explained below, this is no longer true.

## B. Enactment of Section 271(f)

After decades of relative weakness, patent law was in resurgence in the early-and mid-1980s.<sup>19</sup> Proponents of strengthened U.S. patent protection

---

... [W]e note that what is at stake here is the right of American companies to compete with an American patent holder in foreign markets. Our patent system makes no claim to extraterritorial effect; "these acts of Congress do not, and were not intended to, operate beyond the limits of the United States," *Brown v. Duchesne*, 19 How., at 195; and we correspondingly reject the claims of others to such control over our markets. *Cf. Boesch v. Graff*, 133 U.S. 697, 703 (1890). To the degree that the inventor needs protection in markets other than those of this country, the [patent laws] reveal a congressional intent to have him seek it abroad through patents secured in countries where his goods are being used. Respondent [Laitram] holds foreign patents; it does not adequately explain why it does not avail itself of them.

*Id.* at 531.

19. From roughly the late 1920s through the 1970s, U.S. patent protection was relatively weak. Courts frequently narrowed or denied patent rights on arguable legal grounds, and patent protection in general was frequently subordinated to widespread distrust of monopolies and anti-competitive activity. The early 1980s brought many changes, spurred mainly by concerns about lagging U.S. innovation and increasing foreign competition, which often came to dominate industries that were based on U.S. research or inventions. *See, e.g.*, 1 DONALD S. CHISUM, CHISUM ON PATENTS, 11-15 (Cumulative Supp. 2006) (noting the repeated swings in the U.S. Supreme Court from pro-patent to anti-patent and back again); COMMITTEE ON INTELLECTUAL PROPERTY RIGHTS IN THE KNOWLEDGE-BASED ECONOMY BOARD ON SCIENCE, TECHNOLOGY, & ECONOMIC POLICY, A PATENT SYSTEM FOR THE 21ST CENTURY 31-38 (Stephen A. Merrill, Richard C. Levin & Mark Myers, eds., The National Academies Press 2004) (2001), *available at* <http://newton.nap.edu/html/patentsystem> [hereinafter NAS REPORT] (pointing out an anti-patent orientation beginning approximately in 1930 and a pro-patent period in the 1980s and since); FEDERAL TRADE COMMISSION, TO PROMOTE INNOVATION: THE PROPER BAL-

characterized the result in *Deepsouth* as a “loophole” in U.S. patent protection and sought a legislative response. The extreme facts and overly clever conduct in *Deepsouth* buttressed the “loophole” argument. Having taken major steps to strengthen U.S. patent protection little more than a year earlier,<sup>20</sup> Congress passed the Patent Law Amendments of 1984.<sup>21</sup> This statute enacted Section 271(f) of the Patent Act and made other pro-patent changes. Until the final Congressional action, the legislation also contained a precursor of the Process Patent Amendments Act, which was subsequently enacted in separate legislation and codified as Section 271(g) of the Patent Act.<sup>22</sup>

---

ANCE OF COMPETITION AND PATENT LAW AND POLICY 14-23 (2003), available at <http://www.ftc.gov/os/2003/10/innovationrpt.pdf> [hereinafter FTC REPORT]. The FTC REPORT notes that: (i) despite the enactment of the Sherman Act in 1890, the U.S. courts gave little attention to the intersection of patent and antitrust law until the early 1900s; (ii) the courts generally refrained, through most of the 1920s, from limiting even “rather substantial overreaching” by patent owners including a variety of post-sale restrictions on patented products; (iii) an “antitrust backlash” began as early as 1917, reached great strength in the decades following 1930, and continued through the 1970s; and (iv) a strong pro-patent swing, including both pro-patent legislation and judicial decisions that strengthened patent rights, began in the late 1970s or early 1980s and continues to this day.

Reflecting the dominance of antitrust considerations prior to the 1980s, the Supreme Court stated in *Deepsouth*:

[W]e must consider petitioner’s claim in light of this Nation’s historical antipathy to monopoly . . . and of repeated congressional efforts to preserve and foster competition. As this Court recently said without dissent: “In rewarding useful invention, the “rights and welfare of the community must be fairly dealt with and effectually guarded. . . . To that end the prerequisites to obtaining a patent are strictly observed, and when the patent has issued the limitations on its exercise are equally strictly enforced.”

406 U.S. at 530-31. See also *infra* text accompanying notes 123-127.

20. In the Federal Courts Improvement Act of 1982, Pub. L. No. 97-164, 96 Stat. 25, Congress created the United States Court of Appeals for the Federal Circuit, consolidating the appeals of essentially all patent cases into that court, along with jurisdiction over a few other types of cases. Congress intended for the Federal Circuit to bring more consistency into patent law, and it was also generally expected that the new court would strengthen patent rights for the purpose of stimulating U.S. innovation.

21. Patent Law Amendments Act of 1984, Pub. L. No. 98-622, 98 Stat. 3383.

22. See 130 CONG. REC. H10522-34 (daily ed. Oct. 1, 1984) (statement of Rep. Kastenmeier); 130 CONG. REC. H12231-32 (daily ed. Oct. 11, 1984) (statement of Rep. Kastenmeier). Under Section 271(g), infringement generally includes importing a product into the U.S. if that product is made by a U.S. patented process.

As enacted, Section 271(f)<sup>23</sup> consists of two parts, which have largely similar effects. They both make it a new form of “infringement” for anyone to produce, in the United States, unpatented parts usable in a larger product that is claimed in a U.S. patent and supply those parts for assembly with other parts abroad. Clause “1” resembles pre-existing Section 271(b) of the Patent Act defining *inducement of infringement*. It applies only where the components supplied from the United States make up all or a “substantial portion” of the combination claimed by the U.S. patent. Clause “2” resembles pre-existing Section 271(c) on *contributory infringement*. It applies where only a single component or a small portion of the total patented combination is supplied from the United States, if the nature of the component and the intent of the supplier more strongly imply the claimed combination will be made abroad. Conspicuously absent from the statute is any definition of its key terms such as “component” and “combined.”

The legislative history of these provisions is sparse and largely uninformative, consisting of brief statements widely scattered among the much more extensive discussions of different provisions in the House and Senate bills.<sup>24</sup> Nevertheless, recent judicial decisions have claimed support in that

---

23. Section 271(f) states:

(1) Whoever without authority supplies or causes to be supplied in or from the United States *all or a substantial portion of the components* of a patented invention, where such components are uncombined in whole or in part, in such manner as to actively induce the combination of such components outside of the United States in a manner that would infringe the patent if such combination occurred within the United States, shall be liable as an infringer.

(2) Whoever without authority supplies or causes to be supplied in or from the United States *any component* of a patented invention that is especially made or especially adapted for use in the invention and not a staple article or commodity of commerce suitable for substantial noninfringing use, where such component is uncombined in whole or in part, knowing that such component is so made or adapted and intending that such component will be combined outside of the United States in a manner that would infringe the patent if such combination occurred within the United States, shall be liable as an infringer.

35 U.S.C. § 271(f) (emphasis added).

24. The only items relevant to the interpretation of Section 271(f) are brief and scattered statements in the following: (i) an October 1, 1984 Section-by-Section Analysis of the non-final bill, S. REP. NO. 98-663 (1984) *as reprinted* in 1984 U.S.C.C.A.N. 5827, which also appears in S. REP. NO. 98-663, cited below; (ii) 130 CONG. REC. H10522-34 (daily ed. Oct. 1, 1984) (statement of Rep. Kastenmeier); 130 CONG. REC. H12231-32 (daily ed. Oct. 11, 1984) (statement of Rep. Kastenmeier); (iii) Patent Law Amendments

legislative history for major extensions of Section 271(f).<sup>25</sup> Closer review shows the absence of such support, and instead strongly suggests that Congress did not intend the Federal Circuit's recent expansions of the scope of Section 271(f).<sup>26</sup> At bottom, four relevant points emerge from the

---

of 1984, S. REP. NO. 98-663 (1984); and (iv) the statement by President Reagan on the signing of the bill into law, 20 WEEKLY COMP. PRES. DOC. 1818-19 (Nov. 9, 1984).

25. See, e.g., *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005). See *infra* note 42 and accompanying text for a further discussion of this case.

26. The portions of the legislative history referring to what became Section 271(f) state, with no further elaboration or explanation, the following (all emphasis added by the author for later reference):

(i) Section 271(f) "will prevent copiers from avoiding U.S. patents by supplying components of a patented product in this country so that the assembly of the components may be completed abroad. This proposal responds to the United States Supreme Court decision in *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518 (1972), concerning the need for a legislative solution to close a loophole in patent law." Section-by-Section Analysis, *supra* note 24, 1984 U.S.C.C.A.N. at 5828. This Analysis is repeated with the remarks of Representative Kastenmeier dated October 1, 1984. See *supra* note 24. A subsequent portion of the Section-by-Section Analysis emphasized the concerns of Congress to stimulate U.S. innovation via revisions to the patent laws, but this remark was also made in connection with unrelated provisions in the same bill: "During the past three Congresses, much has been heard about the need to improve American creativity. The fostering of technological change and the stimulation of innovation have become our goals." *Id.* at 5837.

(ii) Clause "1" of Section 271(f) corresponds to the provisions in Section 271(b) on "active inducement" of infringement, and the "components" for purposes of that clause can be "staple articles or commodities of commerce" suitable for substantial non-infringing use. Clause "2," on the other hand, corresponds to the provisions in Section 271(c) on "contributory infringement," and the components referred to there must not be "staple article[s] or commodit[ies] of commerce suitable for substantial non-infringing use[s]." Section-by-Section Analysis, *supra* note 24, 1984 U.S.C.C.A.N. at 5828, repeated in S. REP. NO. 98-663 (1984), *supra* note 24.

(iii) Section 271(f) is intended to "declare it to be patent infringement to supply components of an invention patented in the U.S. for final assembly abroad if the purpose of the shipment abroad is to circumvent a U.S. patent. . . . This provision is a response to the Supreme Court's 1972 *Deepsouth* decision, which interpreted the patent law not to make it infringement where the final assembly and sale are abroad. . . . The bill simply amends the patent law so that when components are supplied for assembly abroad to circumvent a patent, the situation will be treated the same as when the invention is 'made' or 'sold' in the U.S." S. REP. NO. 98-663 (1984), *supra* note 24.

(iv) "The bill is needed to help maintain a climate in the U.S. conducive to invention, innovation, and investment. Permitting the subterfuge which is allowed under

legislative history: (i) the prominence of the *Deepsouth* fact pattern in Congressional thinking; (ii) concerns about foreign competition, without identification of any meaningful link between the new statutory provisions and any effect on foreign competition; (iii) use of broad language in some parts of Section 271(f), which allows application of the provision to a broad range of U.S. exports not involving the narrow *Deepsouth* fact pattern; and (iv) Congress' apparent expectation that other vague but important terms in the statute would be read and applied narrowly. These points deserve further explanation.

---

the *Deepsouth* interpretation of the patent law weakens confidence in patents among businesses and investors." *Id.* at 3.

(v) The ability of the U.S. "to foster innovation is a central element of our national security, for without technological and scientific developments, we could not maintain our current standard of living or hope for the diminution of unemployment caused by foreign competition." 130 Cong. Rec. H10522-34 (daily ed. Oct. 1, 1984) (statement of Rep. Kastenmeier), *supra* note 24.

(vi) One main aim of the proposed bill was to "to avoid encouraging manufacturing outside the United States." Section-by-Section Analysis, *supra* note 24, 1984 U.S.C.C.A.N. at 5827. This goal was attributed to a section of the bill that contained both the proposed amendments to the process patent infringement provisions (which were dropped from the bill before its enactment but were later enacted in slightly revised form in separate legislation and are now codified as Section 271(g)) and the proposed new Section 271(f). While this goal was certainly widely shared, it may have been thought of more in connection with the process patent amendments than in relation to Section 271(f).

(vii) Section 271(f) and the other measures enacted with it "address specific, narrow concern[s] in the patent law. However, without enactment of these housekeeping-oriented measures, the patent system would not be responsive to the challenges of a changing world and the public would not benefit from the release of creative genius." 130 Cong. Rec. H10522-34 (daily ed. Oct. 1, 1984) (statement of Rep. Kastenmeier), *supra* note 24.

(viii) Only "a minimal amount of controversy" was expressed about the provisions contained in the legislation, and essentially all of that concerned provisions other than those relating to Section 271(f). *Id.*; 130 CONG. REC. H12231-32 (daily ed. Oct. 11, 1984) (statement of Rep. Kastenmeier), *supra* note 24.

(ix) The new Section 271(f) "closes a loophole in existing law which permit[s] copiers to export jobs and avoid liability by arranging for final assembly of patented machines to occur offshore. . . ." Statement by President Reagan on the signing of the bill into law, *supra* note 24; *see also* 20 WEEKLY COMP. PRES. DOC. 1818-19 (Nov. 9, 1984).

First, the recorded statements and debate referred to the *Deepsouth* case with great prominence. The whole of Section 271(f) was identified as a “response” to that case and the “loophole” it illustrated. None of the statements mentions any other case, fact pattern, or problem to be addressed.<sup>27</sup>

Second, the only concrete objectives identified for Section 271(f) were: (i) to promote confidence in U.S. patents and thereby foster U.S. innovation and reduce U.S. unemployment caused by foreign competition and (ii) to “avoid encouraging manufacturing outside the United States.”<sup>28</sup> There was no explanation of how Section 271(f) would reduce foreign competition or discourage manufacturing outside the United States. The legislative history also contains no mention of three main realities of international patent protection: (i) that any benefits the new provision might provide were already provided by U.S. law *if* the U.S. patent claimed any of the components sent abroad—as opposed to just the full combination assembled abroad; (ii) that a method of obtaining patent protection in foreign markets had long been well recognized, namely for the inventor to apply for patents in the respective foreign countries; and (iii) that the alternative approach implemented in Section 271(f) imposes restrictions on U.S. producers that do not apply to producers located anywhere else in the world. It seems likely that Congress’ failure to consider these points reflected the extreme but deceptively simple fact pattern of *Deepsouth*.<sup>29</sup>

Third, despite following the *Deepsouth* facts in most respects, Section 271(f) deliberately extends well beyond those facts in one respect, thereby vastly expanding its reach. As noted above, the U.S. producer in *Deepsouth* supplied 100% of the parts of the patented de-veining machines for assembly abroad, as well as instructions for the essentially trivial final assembly process. The wording of Section 271(f), however, sets an entirely different threshold for infringement under that Section—ranging down to a single unpatented component in many cases.<sup>30</sup> This vastly low-

---

27. See, e.g., *supra* note 26, at (i).

28. The related objectives of preventing “circumvention” and “avoid[ance]” of U.S. patents and closing the *Deepsouth* “loophole” were supportive of the main goals of creating greater respect for U.S. patents and greater incentives for innovation and, thereby, increasing manufacturing and employment in the U.S. rather than abroad. See *supra* note 26 at (iv) and (v), regarding objective (i) of § 271(f), and *supra* note 26 at (vi), regarding objective (ii) of § 271(f).

29. The unusual facts of *Deepsouth* apparently captured the thinking of those involved in the legislative process. Those facts gave no suggestion of any possible complications and were so extreme as to divert thinking from any other aspects of the subject being addressed.

30. See *supra* note 23 and accompanying text.

ered threshold converted a potentially narrow provision applicable only to the extreme circumstances of *Deepsouth* into a far broader provision applicable to a much wider range of exports.

Fourth, contrary to statements in some recent opinions discussed below, the legislative history suggests *narrow* readings of the key undefined terms in Section 271(f). These interpretive questions can affect the scope of that provision in significant respects. For example, Section 271(f) contains no clear, express requirement that the U.S. produced component(s) or the resulting foreign combination be tangible in order for the section to apply. The statute also uses the broad term “patented invention” in referring to the foreign combination. Thus, subject to the meaning of “components” and to some separate logical difficulties,<sup>31</sup> the provision could be read as extending not just to patents for physical products—as in *Deepsouth*—but also: (1) to patents claiming method (process) inventions and/or (2) to other cases where the U.S.-supplied “component,” the foreign-assembled combination, or both, are intangible. A number of statements in the legislative history, however, indicate Congressional and Presidential assumptions that both the components and the foreign-produced combination would be tangible.<sup>32</sup> This reading would exclude all method patents and some other inventions and components from the reach of Section 271(f).

---

31. See *infra* notes 49-55, regarding logical difficulties in applying Section 271(f)’s requirements to method (process) claims.

32. Examples of this assumption include: (i) the repeated use of the terms “component” and “assemble”—terms that usually connote physical objects physically attached to other physical objects—and the similar implications in the phrase “complete the assembly of [the] components abroad”; (ii) the references to “components” as “articles or commodities of commerce,” the use of the term “patented machine” to describe the resulting combination made abroad, and the use of the term “manufacturing” as the activity Congress intended to retain within the U.S. (*see supra* note 26, at (ii), (ix), concerning the terms “article or commodity of commerce” and “patented machine”; *supra* note 26, at (vi), regarding use of the term “manufacturing”); (iii) the repeated references to the fact pattern of *Deepsouth*—since *Deepsouth* involved only physical components assembled abroad into a mechanical combination invention—together with the absence of any references to any other fact pattern; and (iv) the difficulty of conceptualizing how steps, acts, or other intangibles can be supplied from the U.S. and combined abroad with other “components.” See generally *supra* note 26; *infra* notes 37, 51-55, 67 and accompanying text.

### III. APPLICATION OF SECTION 271(F)—UNCERTAINTIES, JUDICIAL DECISIONS, AND THEIR EFFECTS

Lacking definitions of its key terms but reflecting the stark *Deepsouth* fact pattern, Section 271(f) presented substantial scope-related uncertainties: how far outside the context of machine patents and tangible, mechanical components should the new provision apply? For a decade after its enactment, U.S. patent holders seldom invoked Section 271(f), and its uncertainties remained largely dormant. A few decisions then recognized significant limitations on its reach, particularly as to method (process) patents.

Recently, however, the courts have cast aside nearly all limitations, with the exception of one that actually *increased* the perverse incentives that provision creates. Accordingly, under current case law, Section 271(f) provides for aggressively extraterritorial application of the U.S. patent laws against U.S. producers. Moreover, it does so broadly and without any distinctions like those applied in connection with less aggressive extensions of U.S. patents to off-shore equipment or activities.<sup>33</sup> The following summary of the extensions of Section 271(f) reveals their importance and illuminates the current operation and effects of that provision.

#### A. Applicability to Chemical Combinations

*Deepsouth* involved a machine patent that claimed a combination of mechanical parts. Consistent with this context, the word “components” usually refers to mechanical or electrical parts of a larger physical structure. Accordingly, courts might have deemed Section 271(f) inapplicable to patents for chemical combinations. U.S. defendants accused of infringing such patents by supplying particular unpatented ingredients for inclusion in foreign-made mixtures argued that “components,” as used in the

---

33. *See supra* note 11. Its defenders have characterized Section 271(f) as not technically extraterritorial in its application because it penalizes actions taken in the U.S.—namely producing non-patented components in the U.S. and making them available for combination with other components abroad—regardless of whether the combination abroad actually takes place. While technically true, this misses the point: Congress intended Section 271(f) to limit competition in foreign markets, and Section 271(f) benefits its intended beneficiaries (i.e., holders of certain U.S. patents) only to the extent that it does so. Moreover, damages awarded under that provision are generally based on the actual combinations and sales made abroad, and that provision is superfluous to the extent that the combinations are returned to the U.S. because other provisions of the patent laws make such importation an infringing act regardless of Section 271(f). Further, while only U.S. production can lead to violations of Section 271(f), this is clearly not a virtue, but a shortcoming, of the provision that runs directly counter to the arguments made in favor of its enactment. *See supra* text accompanying note 28.

statute, did not reach chemical ingredients. District court decisions in 1988 and 1999 rejected this defense, finding no limitation of Section 271(f) to “machine” patents and holding that chemical ingredients are “components” of a resulting patented mixture.<sup>34</sup> A recent Federal Circuit decision involving chemical catalysts raised additional issues but lends support to this result.<sup>35</sup>

## B. Applicability to Patented Methods/Processes

A second extension of Section 271(f)—and one of considerable importance—concerns method (process) patents. The wording and legislative history of Section 271(f) discourage its application to method claims because the words “components” and “assemble” are not normally understood as referring to *steps* or *acts*, which are the elements of a method (process) claim. As set out above, the legislative history of Section 271(f) refers to the creation of infringing combinations under that section as “manufacturing” or “assembly” of components, suggesting tangible objects and a physical combination invention. At a few points, it also refers to the resulting combinations as “products” or as “machines.”<sup>36</sup> In contrast, neither the statutory wording nor the legislative history contains any of the terminology associated with method inventions or method claims, namely “processes,” “method,” “step,” or “act.” This omission of all terminology relating to method patents seems more significant than it might otherwise be because patented processes were the focus of the immediately preceding segment of the bills that became the Patent Act Amendments of 1984, namely the segment that would have enacted a precursor of the Process Patent Act, now codified as Section 271(g). Even if steps or acts could qualify as “components,” neither *Deepsouth* nor anything else in the legislative history suggested how steps or acts could be “supplied from” the U.S. in “uncombined” form and then combined with other steps or acts abroad. The legislative history does not discuss anything along

---

34. *Lubrizol Corp. v. Exxon Corp.*, 696 F. Supp. 302, 323-25 (N.D. Ohio 1988) (enjoining, without discussion, shipments of the key chemical ingredient abroad for use in the subject chemical combinations); *W.R. Grace & Co. v. Intercat, Inc.*, 60 F. Supp. 2d 316, 320-21 (D. Del. 1999) (recognizing that the case law held Section 271(f) inapplicable to process/method patents but distinguishing it as not addressing chemical compositions; emphasizing the absence from the statutory wording of any exclusion of chemical compositions from the statute’s coverage).

35. *See Union Carbide Chems. & Plastics Tech. Corp. v. Shell Oil Co.*, 434 F.3d 1357, 1358 (Fed. Cir. 2006). There, a Federal Circuit panel held that the shipment of a U.S.-made chemical catalyst abroad for use there in a process claimed in U.S. patent constitutes infringement under Section 271(f). *See* the discussion in the immediately following Section regarding the logical difficulties associated with this conclusion.

36. *See supra* note 32 and accompanying text.

these lines. Rather, the legislative deliberations had an entirely different focus and apparent intent<sup>37</sup>

Accordingly, courts and commentators thought that Section 271(f) simply did not apply to such claims. A 1991 Federal Circuit decision so concluded in a brief and largely unexplained holding.<sup>38</sup> Three reported district court decisions concurred with and applied this rule, each with at least some discussion of the point.<sup>39</sup> A leading authority on patent law affirmed the inapplicability of Section 271(f) to method claims based on its use of the word “components” and without seeming to think additional authority was needed to support that conclusion.<sup>40</sup> Finally, the Federal Circuit’s lan-

---

37. See also *Union Carbide*, 434 F.3d at 1358 (Lourie, J., dissenting). Judge Lourie stated:

The statute [Section 271(f)] itself speaks of supplying “components of a patented invention, where such components are uncombined . . . in such manner as to actively induce the combination of such components outside of the United States,” . . . *the whole tenor of that provision relates to physical inventions, i.e., apparatus or compositions, not methods.*

*Id.* (emphasis added).

38. See *Standard Havens Prods. v. Gencor Indus.*, 953 F.2d 1360, 1374 (Fed. Cir. 1992) (recounting various theories under which actionable infringement (or inducement of infringement or contributory infringement) of a *process* patent might be found where the defendant supplied a product for use abroad in the patented process and, after rejecting all other possible theories, stating that, likewise, it “do[es] not find the provisions of Section 271(f) implicated” in that context).

39. See *Imagexpo, L.L.C. v. Microsoft Corp.*, 299 F. Supp. 2d 550 (E.D. Va. 2003) (denying Microsoft’s motion in limine to preclude introduction of evidence of its foreign sales of software for calculation of damages under Section 271(f) because, although that provision does not apply in connection with method patents, the patent in question was not for a method but for a computer apparatus programmed with software like that provided by Microsoft); *W.R. Grace & Co. v. Intercat, Inc.*, 60 F. Supp. 2d at 320-21, discussed *supra* note 34 and the accompanying text; *Enpat, Inc. v. Microsoft Corp.*, 6 F. Supp. 2d 537 (E.D. Va. 1998) (granting defendant Microsoft’s motion for partial summary judgment on Section 271(f) liability to preclude damages based on foreign use of software supplied from U.S. and allegedly performed patented process because, based on its legislative history, Section 271(f) applies only to the assembly of components into patented products abroad and does not apply to method or process patents).

40. See Donald F. Chisum, *Extraterritorial Application of U.S. Intellectual Property Law: Comment: Normative and Empirical Territoriality in Intellectual Property: Lessons from Patent Law*, 37 VA. J. INT’L L. 603, 607 (1997).

Assessed in terms of economic policy, section 271(f) is ill-conceived. It was presumably an attempt to close a loophole created by *Deepsouth*, but its most immediate effect is to create one more incentive for U.S. companies [that] compete in foreign markets to move their manufacturing facilities abroad. Furthermore, the statute is incomplete. It covers the manufacture and export of unpatented components of patented ma-

guage and analysis in a recent decision addressing a separate question under Section 271(f) seemed to support this negative conclusion.<sup>41</sup>

Surprisingly, in this context, the recent Federal Circuit decision in *Eolas Technologies, Inc. v. Microsoft Corp.*,<sup>42</sup> discussed more fully in the following Section, adopted the opposite position without qualification and without mentioning any of these contrary considerations or authorities. Writing for the panel in *Eolas*, Judge Rader stated that Section 271(f) applied to the method claim as well as the product claim asserted there. His reasoning relied mainly on the use, in both clauses “1” and “2” of Section 271(f), of the broad term “patented invention” instead of narrower language that could have clearly excluded method claims from the scope of that provision.<sup>43</sup> Ignoring any possible limitations inherent in the words “components” and “assemble,” he reasoned that nothing in Section 271(f) expressly limits its application to physical structures or precludes its application to method claims or to the “step” or “act” elements of which methods consist.<sup>44</sup> Further, he said, excluding method claims from the reach of Section 271(f) would amount to distinguishing among types of inventions without any principled basis. This, he argued, would contravene the policy of non-discrimination among types of inventions, as he read it in the 1994 TRIPS Agreement.<sup>45</sup>

---

chines and other structural combinations. It does not cover manufacture and export of a component for use in a patented process—even though many valuable inventions take the form of new processes for using materials or components.

*Id.*

41. *Pellegrini v. Analog Devices, Inc.*, 375 F.3d 1113 (Fed. Cir. 2004); *see infra* note 88.

42. *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005), *cert. denied*, 126 S. Ct. 568 (2005).

43. *Id.* at 1338. Such narrower language would presumably have been “patented machine” or, possibly, “patented machine, manufacture, or composition of matter.” *See* 35 U.S.C. §101.

44. *Eolas*, 399 F.3d. at 1339.

45. *Id.* at 1339. The TRIPS agreement of the 1994 Uruguay Round GATT trade agreements states that patents “shall be available and patent rights enjoyable without discrimination as to the place of invention . . . [or] field of technology.” World Trade Organization, A Summary of the Final Act of the Uruguay Round, [http://www.wto.org/english/docs\\_e/legal\\_e/ursum\\_e.htm#nAgreement](http://www.wto.org/english/docs_e/legal_e/ursum_e.htm#nAgreement) (last visited Nov. 7, 2006). Persons opposing software patents have pointed out the ambiguity of this provision and the provisions in the TRIPS Agreement and the WIPO Copyright Treaty that computer software is to be protected as literary works, i.e., by copyright laws. *See, e.g.*, Justin Mason: Happy Software Prole, TRIPS, WIPO, and the WTO Doing the Right Thing on Software Patents?, <http://taint.org/2004/04/24/053238a.html> (last visited Nov. 7, 2006).

On little more than this,<sup>46</sup> Judge Rader broadly asserted that “steps” and “acts” are the “components” of method/process inventions and that Section 271(f) is fully applicable to method claims.<sup>47</sup> Subsequently, in *Union Carbide Corp. v. Shell Oil Co.*,<sup>48</sup> another opinion by Judge Rader, an overlapping panel of the Federal Circuit, followed *Eolas* and applied Section 271(f) to method claims. As detailed below, however, other Federal Circuit judges have recently criticized the *Eolas* and *Union Carbide* results. This criticism and the other considerations discussed below undermine the *Eolas* and *Union Carbide* conclusions, particularly vis-à-vis potential future review of the question by the Federal Circuit en banc or by the Supreme Court.

The *Eolas* reasoning has an appealing surface simplicity, but it lacks any supporting authority and seems shallow compared to the contrary implications in the statutory wording and legislative history. Moreover, its result leads to two logical problems which several Federal Circuit judges have identified in two recent opinions criticizing *Eolas*.

---

46. Without indicating its relevance, Judge Rader also noted that software code alone qualifies as a patentable process and that it can also be patented in conjunction with physical structure—citing cases decided, respectively, 10 and 15 years after the enactment of Section 271(f). *Eolas*, 399 F.3d at 1339-40. Congress probably would not have shared this view when it enacted Section 271(f) because software was not generally deemed patentable at the time. Furthermore, the independent patentability of an element or “component” of a claim is generally not relevant to the validity or infringement of that claim. It is unclear why such independent patentability would be relevant to whether something could be deemed a “component” for purposes of Section 271(f). After all, none of the separate components of the patent at issue in *Deepsouth* was independently patentable.

47. Judge Rader stated, for the court:

[The] statutory [wording does] not limit Section 271(f) to patented “machines” or patented “physical structures.” Rather every form of invention eligible for patenting falls within the protection of Section 271(f). By the same token, the statute did not limit Section 271(f) to “machine” components or “structural or physical” components. Rather every component of every form of invention deserves the protection of Section 271(f). . . . A “component” of a process invention would encompass method steps or acts. *See, e.g.*, 35 U.S.C. § 112, ¶ 6. A “component” of an article of manufacture invention would encompass a part of that construct.

*Id.* at 1339 (emphasis added). Section 112, ¶ 6, cited here by Judge Rader, has nothing to do with the scope or interpretation of “components” or Section 271(f). It merely indicates that “steps” or “acts” are the elements of method (process) claims.

48. 425 F.3d 1366 (Fed. Cir. 2005), *reh’g & reh’g en banc denied*, 434 F.3d 1357 (Fed. Cir. 2006).

First, for Section 271(f) to apply to method claims, it is also necessary: (i) that the key “component(s)”—i.e., one or more of the claimed “steps” or “acts” that make up the method—be “supplied in or from the United States” in “uncombined” form; and (ii) that this be done for the purpose of “combining” them, abroad, with the other claimed steps or acts to satisfy all the elements of the relevant method claim(s). These statutory requirements are easy to understand and apply with respect to physical components—tangible objects are routinely produced in one place, shipped to another, and combined with other physical objects there. They are essentially incomprehensible, however, as to steps or acts.<sup>49</sup> The facts in *Eolas* tended to obscure this problem but did not solve it.<sup>50</sup> Because of this conceptual problem, dicta in an August 2005 Federal Circuit panel decision expressed disagreement with this *Eolas* holding.<sup>51</sup> Similarly, at least three of the Federal Circuit judges who dissented from the 2006 denial of rehearing en banc in the *Union Carbide* case noted above did so in part be-

---

49. According to Judge Rader’s analysis set forth *supra* note 47, the “components” of a patented method are the steps or acts that comprise that method. It would seem, therefore, that those steps or acts would have to be supplied from the U.S. in uncombined form to a foreign location for combination there. See the italicized portion of the quotation from the Federal Circuit’s recent decision in the highly publicized BlackBerry case set forth *infra* note 51, pointing out the incongruity of the Section 271(f) language as to method/process claims.

50. It was difficult to perceive this problem on the *Eolas* facts because the U.S.-supplied “component” there was computer code, i.e., instructions for computer-accomplished steps or acts that were among the claim elements of the patent. Coded instructions for steps or acts, however, are not the same as the steps or acts themselves, which Judge Rader’s *Eolas* reasoning seemed to require be supplied from the U.S. See *supra* note 47.

51. *NTP, Inc. v. Research in Motion, Ltd.*, 418 F.3d 1282, 1322 (Fed. Cir. 2005), *reh’g and reh’g en banc denied*, 2005 U.S. App. LEXIS 23112 (Fed. Cir. Oct. 7, 2005), *cert. denied*, 126 S. Ct. 1174 (Jan. 23, 2006). The NTP panel refused to apply Section 271(f) to the furnishing of BlackBerry handheld devices to U.S. customers in connection with an international system for wireless e-mail communications that the court found to infringe NTP’s U.S. patent. The NTP panel’s revised decision pointed out the difficulties in applying Section 271(f) to method patents:

Although [the] *Eolas* [decision] was correct to observe that Congress did not expressly limit [S]ection 271(f) to a specific type of invention, we have held that the very nature of the invention may compel a difference . . . . A method, by its very nature, is nothing more than the steps of which it is comprised. . . . [I]t is difficult to conceive of how one might supply or cause to be supplied all or a substantial portion of the steps of a patented method in the sense contemplated by the phrase “components of a patented invention” in Section 271(f). . . .

*Id.* (emphasis added). The panel went on to cite *Standard Havens*, a Federal Circuit decision widely interpreted as holding Section 271(f) inapplicable to method claims. *Id.*

cause of the incongruity of the wording of Section 271(f) in the context of method claims.<sup>52</sup>

Second, the *Eolas* equation of Section 271(f) “components” with “steps” or “acts” in order to apply that provision to method claims should logically preclude many, if not most, potential applications of Section 271(f) to method claims, but Judge Rader and the remainder of the *Union Carbide* panel refused to recognize that logical result. To be specific, perhaps the most common fact pattern where the holder of a U.S. method patent may allege infringement under Section 271(f) is U.S. production of a physical object or substance that is then shipped abroad for use in the claimed processes. The recent *Union Carbide* case provides a good example.<sup>53</sup> In such cases, however, the U.S.-supplied “component” is not a “step” or “act,” as the *Eolas* reasoning would require. Nevertheless, the *Union Carbide* panel reversed the district court’s holding that Section 271(f) could not apply in this context and remanded the case for determination of additional damages under the U.S. patent from use of the process abroad. While Judge Rader spent several paragraphs arguing that Section 271(f) could apply in this context, he never dealt directly with the logical inconsistency between his *Eolas* reasoning and his result in this case.<sup>54</sup> The three-judge dissent from the denial of rehearing en banc in that case forcefully criticized this logical lapse.<sup>55</sup>

*Eolas* survived a petition for rehearing and rehearing en banc in the Federal Circuit and a petition for a writ of certiorari in the Supreme

---

52. *Union Carbide Chems. & Plastics Tech. Corp. v. Shell Oil Co.*, 425 F.3d 1366 (Fed. Cir. 2005), *reh’g and reh’g en banc denied*, 434 F.3d 1357, 1358-59 (Fed. Cir. 2006) (Lourie, J., dissenting).

53. In *Union Carbide*, Shell supplied a tangible, unpatented catalyst from the U.S. and was charged with infringement under Section 271(f) because that catalyst was used abroad in a process claimed in Union Carbide’s U.S. patent. *See* 425 F.3d at 1379-80.

54. *Id.*

55. 434 F.3d at 1358-59. The dissenters stated, in part:

[Section 271(f)] speaks of supplying “components of a patented invention . . . .” The whole tenor of that provision relates to physical inventions, i.e., apparatus or compositions, not methods. . . . A component of a process is a step in the process; it is not the physical material to be used in the process. What the panel opinion here holds is that supplying a component to be used in one of the process steps can create infringement. That is, in my view, an incorrect extension of the statutory language. . . . A material for use in practicing a process is not a component of that process.

*Id.* (Lourie, J., dissenting).

Court.<sup>56</sup> Subsequently, the *Union Carbide* parties settled their dispute after the Federal Circuit denied rehearing in that case and before Shell filed its widely expected petition for Supreme Court review. These events, however, did little to solidify *Eolas* and *Union Carbide* as reliable precedent, and several considerations call into question their long-term viability. These include the earlier case law holding Section 271(f) inapplicable to method claims, the recent criticism of the *Eolas* and *Union Carbide* decisions, the logical difficulties those decisions create, and the wording and legislative history of Section 271(f). Indeed, as noted below, the Supreme Court could effectively overrule *Eolas* and *Union Carbide* by its ruling in the now-pending *Microsoft Corp. v. AT&T Corp.* case.<sup>57</sup>

Note, however, that overruling these two decisions and holding Section 271(f) inapplicable to method claims could reduce, but would come nowhere near eliminating, the perverse incentives Section 271(f) creates for companies that produce, or consider producing, technologically advanced components in the United States for international markets. That is because, under modern patenting practices, most inventions that can be claimed as processes can be, and usually are, also claimed as “products” or “systems,” which the patent laws treat as “machines” or “manufactures.” Accordingly, even if Section 271(f) did not apply to method claims, the risks of Section 271(f) would largely persist for U.S. producers. Thus, corporate managers would still have to take that provision into account when deciding whether to direct capital investments, employee training, and technology-based manufacturing activities to U.S. locations or to foreign locales. To a large extent, its perverse incentives would remain.

### C. Computer Software—Intangibles as “Components” and “Copies” Made Abroad as “Supplied” from the U.S.

Another recent extension of Section 271(f) concerns computer software and, apparently, other intangible but potentially functional types of information.<sup>58</sup> This field is economically and technologically important in its own right. It also illustrates a common pattern of geographically differentiated production that is important to any assessment of Section 271(f).

In contrast with the many types of computer hardware, software is a field where U.S. producers retain strong or dominant international posi-

---

56. *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005), *cert. denied*, 126 S. Ct. 568 (2005).

57. See *infra* note 86 and accompanying text, regarding the possibility that the Supreme Court’s decision in the pending *Microsoft Corp. v. AT&T Corp.* case might affect the applicability of Section 271(f) to method claims.

58. See *infra* text accompanying note 168 for some possible examples.

tions. In this and some other technological arenas, U.S. producers are competing successfully in certain advanced, high-value products even though they have been displaced by lower-cost foreign producers as to many companion components. This pattern of production promotes off-shore assembly of combinations involving one or a few U.S.-produced high-value items with other components produced abroad. This, in turn, increases the likelihood that Section 271(f) will play a prominent role in any dispute involving a U.S. patent covering the overall product, system, or method, whether the patent is held by a U.S. or a foreign entity.

Several complex issues confront the application of Section 271(f) to software supplied from the United States. Two recent Federal Circuit decisions, both against Microsoft Corporation and both involving the same fact pattern, frame these issues: the March 2005 *Eolas* decision, discussed above, and the July 2005 decision in *AT&T Corp. v. Microsoft Corp.*<sup>59</sup>

In each of these decisions, the Federal Circuit applied Section 271(f) broadly, holding it applicable where the U.S.-supplied “component” was intangible computer code that was sent abroad, slightly revised at the foreign location, and copied there onto foreign-made data storage devices for inclusion in foreign-made computer systems that would be sold and used only in foreign countries. Accordingly, while the Supreme Court recently granted review in the *AT&T* case,<sup>60</sup> Section 271(f) now seems fully applicable to all kinds of computer software. Moreover, it is unclear what, other than possible restrictions the Supreme Court may add, will limit its reach as to any other types of intangible information, data, patterns, or other possible U.S.-supplied items of value that play any role in foreign-made combinations—including instructions, data, or materials that facilitate foreign steps in methods patented in the United States.

As discussed below, the resulting risks and burdens on U.S. producers can be very large.<sup>61</sup> The district court judge in one of these recent decisions observed that “[the] novel issue[s] regarding the application of Section 271(f) presented [in these cases have] profound ramifications for Microsoft and other U.S. software manufacturers.”<sup>62</sup>

---

59. 414 F.3d 1366 (Fed. Cir. 2005), *reh'g and reh'g en banc denied*, No. 04-1285, 2005 U.S. App. LEXIS 24112 (Fed. Cir. Oct. 20, 2005); *cert. granted*, 127 S. Ct. 467 (2006).

60. *Microsoft Corp. v. AT&T Corp.*, 127 S. Ct. 467 (2006).

61. *See infra* text accompanying note 122.

62. *AT&T Corp. v. Microsoft Corp.*, 71 U.S.P.Q. 2d (BNA) 1118-19, 2004 U.S. Dist. LEXIS 3340, \*2 (S.D.N.Y. 2004).

### 1. *Typical Software Fact Pattern*

The plaintiffs in the recent software cases<sup>63</sup> held U.S. patents claiming apparatus or methods that required software in combination with hardware or other items. Each of the plaintiffs sued Microsoft in U.S. courts alleging purely domestic infringement in the loading of certain Microsoft software products onto computer systems in the United States.<sup>64</sup> In the same lawsuits, they also sought additional damages under their U.S. patents and Section 271(f) based on Microsoft's transmission of its U.S.-produced software to foreign Microsoft licensees. The licensees would adjust the software for the particular installations and copy the result onto foreign-made hardware in those foreign countries. The plaintiffs did not assert that any of the resulting off-shore combinations were sent back into the United States. The foreign sales typically accounted for approximately 2/3 of the total damages sought from Microsoft in these cases.

In these cases, Microsoft's U.S.-developed software found its way into the foreign-made combinations through two alternative routes. In the first route, the software, in not-quite usable (i.e., not quite "executable") form, was recorded in the United States onto a few "golden master" disks—like CDs or DVDs. Those disks were shipped to foreign entities having license agreements with Microsoft. At the foreign locations, the licensees read the software off of the "golden masters." They then slightly re-arranged the code to make it directly executable, saved it, "replicated" it onto large numbers of storage devices—generally hard-disk drives—and included it inside the foreign-made computer systems. The resulting foreign-assembled combinations contained no tangible item supplied from the United States—such as the "golden masters" themselves.

The second alternative utilized electronic transmission, rather than physical transport of CD-like disks, to take the software abroad. In this method, the U.S.-developed code was encrypted here and then electronically transmitted, such as via the internet, from the United States to the

---

63. See *supra* note 39, regarding two software cases in the district courts besides *Eolas* and *AT&T v. Microsoft*, which are the focus here.

64. Technically, most of the alleged domestic violations were "contributory infringement" or "active inducement" of infringement under 35 U.S.C. § 271(c) or (b), respectively. The AT&T patent covered technology for compressing and de-compressing coded speech signals, and the accused Microsoft product was its Windows operating system, which apparently includes such technology. The *Eolas* patent was directed at software allowing the opening of, and interaction with, separate applications through a web browser, and Microsoft's Internet Explorer was the accused product. See also *Imagexpo, L.L.C. v. Microsoft Corp.*, 299 F. Supp. 2d 550 (E.D. Va. 2003) (involving Microsoft's NetMeeting software as the accused product).

foreign parties under license with Microsoft. There it was downloaded, decoded, decrypted, slightly re-arranged into executable code, saved, and then “replicated” (copied) onto the large numbers of foreign made storage devices in foreign made computer hardware as described above. In this alternative, no tangible or material item was supplied from the United States at all, let alone included in the accused combinations.

These facts raised two main, but overlapping, issues under Section 271(f): (i) did the software qualify as a “component” for purposes of that provision; and, if so, (ii) was the relevant “component” “supplied . . . from the U.S.” where the licensees included only the foreign-made representations of the information (i.e., the copies made and recorded abroad onto the foreign-sourced data storage devices) in the foreign-assembled combinations? These questions are related to the applicability of Section 271(f) to method claims, discussed above; and the Federal Circuit has taken essentially the same approach here as it did there. Specifically, it has answered each of these questions in the affirmative, but it did so with differing amounts of attention and understanding. While the Federal Circuit judges initially overlooked some of the complexities in this area, they now recognize those complexities and disagree forcefully on at least two key points.

## 2. *Did the Software Qualify as a Section 271(f) “Component”?*

Microsoft advanced several arguments and analogies, and some arguably relevant case law, to show that its U.S.-supplied software did not qualify as a “component” under Section 271(f). Most fundamentally, it argued, the software was merely intangible information and instructions, which were not what Congress understood as “components” when it enacted Section 271(f).

Despite these arguments, the Federal Circuit’s *Eolas* ruling, like the district court decisions in that case and two others,<sup>65</sup> held that computer code, whether supplied on a “golden master” or transmitted electronically, qualifies as a Section 271(f) “component.”<sup>66</sup> Each such ruling relied mainly on the absence of statutory language expressly excluding intangibles from “component” status under that provision.

---

65. *Eolas Techs., Inc. v. Microsoft Corp.*, 2004 U.S. Dist. LEXIS 534, 70 U.S.P.Q. 2d (BNA) 1939 (N.D. Ill. Jan. 14, 2004); *AT&T Corp.*, 2004 U.S. Dist. LEXIS 3340; *Imagexpo*, 299 F. Supp. 2d 550.

66. *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005). In its *AT&T* ruling, the Federal Circuit said *Eolas* determined this issue and declined to revisit it. 414 F.3d 1366, 1369 (Fed. Cir. 2005).

Strictly as a matter of analysis and statutory interpretation, this reasoning was weak and the result is highly questionable. None of the cases discussed the Congressional enactment of Section 271(f) or its legislative history in any substantial and meaningful way; none of them recognized the implications of the terms used in the legislative history or the possibility of a narrower meaning in the word “component” itself; none of them acknowledged that software inventions were generally deemed unpatentable when Congress enacted Section 271(f) and that Congress never considered software, information, and other intangibles as possible “components” for purposes of Section 271(f).<sup>67</sup> Indeed, given the 1984 Congressional context and the heavy influence of the *Deepsouth* fact pattern, the absence of language expressly excluding disembodied software or other intangibles such as information or instructions from “component” status cannot be understood as suggesting that they should be included. Rather, it almost surely reflects how very far the possibility of such inclusion was from Congressional thinking at the time.

We should recognize that software can have “component-like” attributes from the standpoints of functionality, structural integration with hardware, and the manner in which consumers sometimes obtain and handle it.<sup>68</sup> Further, while the existing case law and legislative history sug-

---

67. Further, as discussed above, the origins and background of that provision involved strictly physical objects as “components,” and the legislative history referred to “products,” “manufacturing,” “machines,” and “assembly” but never to “intangibles,” “information,” “instructions,” “steps,” “acts,” etc. See *supra* notes 26, 32, 37. Moreover, the background of the 1984 Patent Law Amendments suggests that intangibles, at least those like process steps, would have been mentioned if their possible “component” status under Section 271(f) was within Congressional contemplation. See *supra* note 26, at (iii), regarding the companion provision in the same bill as enacted Section 271(f), which was the precursor of the Process Patents Amendments Act, now codified as Section 271(g).

With further regard to the wording of Section 271(f), none of the decisions recognized the view, expressed less than a year later by three other Federal Circuit judges, that “the whole tenor of [Section 271(f)] relates to physical inventions, i.e., to apparatus or compositions [of matter]....” See *supra* notes 37, 55. In addition, when Congress enacted Section 271(f), software was not generally deemed patentable. *Parker v. Flook*, 437 U.S. 584 (1978); *Gottschalk v. Benson*, 409 U.S. 63 (1972); cf. *Diamond v. Diehr*, 450 U.S. 175 (1981); *Diamond v. Chakrabarty*, 447 U.S. 303 (1980); but see *AT&T Corp. v. Excel Commc’ns, Inc.*, 172 F.3d 1352 (Fed. Cir. 1999); *State Street Bank & Trust Co. v. Signature Fin. Group*, 149 F.3d 1369 (Fed. Cir. 1998). This makes it even less likely that anyone involved in the drafting, debate, or voting on the legislation contemplated information, software, or other intangibles as potential “components” for purposes of Section 271(f).

68. The following attributes make many types of software resemble “components,” as that term is used in everyday speech: (a) software is created for the purpose of being combined with computer hardware to achieve an operable physical system; (b) it is some-

gested information or other intangibles supplied from the U.S. would not qualify as “components” for purposes of Section 271(f), none of the reported cases was exactly on point. In particular, none of them addressed the provision from the U.S. of directly functional information, like computer code, where that information itself was subsequently incorporated as an operating part of a patented combination. Thus, considering only the statutory language and giving each of the words a generous reading, an affirmative answer to the “component” question could not be ruled out.

At bottom, two things seem to have influenced the courts—particularly the Federal Circuit in *Eolas* and *AT&T*—away from a more careful and technical effort at statutory interpretation: (i) the courts’ favorable regard for Section 271(f) and its presumed benefits and their resulting desire to apply it liberally; and (ii) the courts’ appreciation of the importance of software and software innovations, both to particular patented inventions and, more generally, as an area of inventive activity and a subject of commerce.<sup>69</sup> The following Section discusses these two factors in detail, paying special attention to the *AT&T* decision.

A third factor was also at work in *Eolas*: a failure of the Federal Circuit panel to grasp some of the complexities the case presented. When the additional complexities appeared more clearly only five months later in the far simpler *AT&T* appeal, Judge Rader, the author of the *Eolas* opinion, reversed his position on a key issue that he, and presumably the other two members of the *Eolas* panel, had failed fully to grasp in that earlier case.<sup>70</sup> He could not have written the opinion he did in *Eolas* if he had understood that issue as he came to understand it only five months later in *AT&T*.

---

times distinct and separable from the associated hardware and can be added to the computer hardware, deleted, revised, and replaced; (c) it is easily placed on tangible objects such as magnetic disks, CDs, magnetic tape, and/or silicon memory chips, and, in these forms, easily transported, delivered, etc.; (d) it can be purchased separately; (e) when installed, it is highly functional—overshadowing, in some respects, the computer hardware with which it is combined; (f) it exists as physical patterns—essentially hardware—at most or all times, not only in storage devices but also in operating computer systems, including during its use; and (g) it is often a large and complex package that must be highly exact in order to produce the desired performance of the overall physical system.

69. See, e.g., *Eolas*, 399 F.3d at 1339 (emphasizing software as “the key part” of the overall combination).

70. That issue, addressed in the immediately following Section, concerns the foreign reproduction of the U.S.-supplied software followed by the recording of that reproduced code onto the foreign-made hard-disk drives at the foreign location.

3. *Did the Software Copies Made Abroad Qualify as Components Supplied from the United States?*

In its recent *AT&T* decision, a Federal Circuit panel, whose membership overlapped with the *Eolas* panel, addressed a question that was present and essential to the result in *Eolas* but was neither well-articulated nor clearly recognized in that case. This question was whether the foreign-made copies of the U.S.-sourced software should be deemed (i) items supplied from the U.S.—allowing application of Section 271(f)—or, alternatively, (ii) items made abroad—precluding application of that provision. This question is closely related to a dual reality of software: it is both intangible and tangible.

Considering only the intangible aspect of software, the question of foreign-made copies is essentially subsumed in the “component” question. If one deems software to be completely and fundamentally intangible yet fully within the definition of “component” for purposes of Section 271(f), the copy-vs.-original distinction loses much of its potential importance. In this view, and neglecting a few technicalities, the key *intangible* “component” was truly “supplied from” the United States and genuinely “included” or “combined” into the foreign-made computers.<sup>71</sup> The *Eolas* panel took this approach without focus or articulation. As explained below, two—but not all three—members of the *AT&T* panel followed this approach too, ignoring software’s other reality—its tangible embodiment. This concentration solely on the intangible aspect of software is not reality, however, from an operational or technical point of view.

The physical representation of software is fundamental and essential from a technical and operational standpoint. Software must take physical form in order to operate within a computer system, as well as to be saved,

---

71. Even neglecting the foreign replication/copying of the software as sent from the United States, technical changes to that software—such as decryption, decompression, and then final compilation and linking—generally occur at the foreign locations to ensure that the software will operate properly with the hardware with which it will be combined. This means that the item supplied from the U.S. is slightly different from the item included in the foreign combination. No mention of these small technical changes appeared in the Federal Circuit’s *Eolas* or *AT&T v. Microsoft* decisions. This may be appropriate, to the extent that Section 271(f) remains on the books and courts continue to follow the main holdings of these cases. Presumably a party intent on evading Section 271(f) could arrange matters so that some re-adjustment or re-assembly of the component(s) sent from the U.S. would be required after their delivery at the foreign location. To deem that stratagem effective in rendering the provision inapplicable seems arguably inconsistent with what Congress apparently intended to accomplish in enacting that provision, namely establishing infringement despite the need for some amount of foreign assembly work before the claimed combination comes into existence at the foreign location.

transferred, or transmitted.<sup>72</sup> Copying and installing software is a wholly physical process of altering tangible objects. In *AT&T*, the “replication” amounted to a fine-scale, late-stage manufacturing process that added a digital pattern, representing information and instructions, to physical objects by rearranging the physical substance of those objects.<sup>73</sup> The fine-scale re-arranging of the material of the foreign-sourced disk at a foreign location to reflect U.S.-supplied information is easier to characterize, from a physical point of view, as “foreign manufacturing” based on U.S.-supplied information than as inclusion of a “component” supplied from the United States.

For the two-judge majority of the *AT&T* panel, these technical and operational considerations, and indeed the process of statutory interpretation based on Congressional intent, were relatively unimportant. Far more significant, in their view, were two “bigger picture” considerations—essentially the same two that influenced the *Eolas* panel on the “component” question. First, the majority had a high regard for Section 271(f) and its presumed benefits. They characterized Section 271(f) as a “remedial” statute, and expressly gave it a broad and expansive interpretation—ironically relying on a U.S. Supreme Court approach to interpretation of

---

72. The *Eolas* panel referred to this dual reality and used it as one basis for declining to view the intangible aspects of software as incompatible with its treatment as a Section 271(f) “component.” Referring to software’s intangible aspects as “software” and its physical aspects as “hardware,” the *Eolas* panel stated:

[A]s the district court pointed out, process and product—software and hardware—are practically interchangeable in the field of computer technology. On a functioning computer, software morphs into hardware at the touch of a button. In other words, software converts its functioning code into hardware and *vice versa*. Thus . . . the computer transforms the code on the [golden master] into a machine component in operation. Thus, sound policy again counsels against varying the definition of “component of a patented invention” according to the particular form of the part under consideration, particularly when those parts change form during operation as occurs with software code.

399 F. 3d at 1339-40 (citation omitted). In actuality, of course, software within a computer system takes physical form at all times, although that physical form is readily changeable.

73. In the recent cases, those objects were hard disks, and the added patterns were in their fine-scale magnetization. The replicated pattern was provided, subject to the technicalities discussed *supra* note 71, as prescribed by, and representing, U.S.-supplied information—via the “golden master” or via electronic transmission—but the ordinary usage of the word “component” fits the resulting patterned disk better than it fits the information that was used as a model for the pattern.

the Federal Securities Laws that the Court has long since abandoned.<sup>74</sup> On little more basis than this, the majority rejected possible interpretations of Section 271(f) that might “subvert[] [its] remedial nature . . . [and permit] a technical avoidance of the statute by ignoring the advances in . . . industry practices that developed after the enactment of [Section] 271(f).”<sup>75</sup>

Second, like the *Eolas* panel, the *AT&T* majority had a high regard for the importance of software innovations and tried not to exclude them, as a practical matter, from the presumed benefits of Section 271(f). Recognizing that copying of the U.S. components abroad could easily constitute foreign manufacturing rather than “supply[ing the copies] from the U.S.,” the panel majority sought a broader definition of “supply.” In doing so, the panel majority looked *not* to Congress’s understanding or intent when it enacted Section 271(f) and *not* to the usage of that term in other statutory provisions, cases, or in standard references. Instead, it looked to “the way software is typically ‘supplied.’”<sup>76</sup> Based on that consideration, it concluded that “copying . . . is part and parcel of software distribution” so that “the act of copying is subsumed in the act of ‘supplying.’” Accordingly,

---

74. Quoting a statement in the legislative history that described Section 271(f) as a significant “housekeeping” provision, the majority in *Eolas* re-characterized it as a “remedial” provision that “should be construed broadly to effectuate its purposes.” 414 F.3d at 1371 (emphasis added). The majority stated that, at its enactment, Section 271(f) was “touted as a ‘housekeeping-oriented’ measure, without which ‘the patent system would not be responsive to the challenges of a changing world and the public would not benefit from the release of creative genius,’” and then quoted from *Tcherepnin v. Knight*, 389 U.S. 332, 336 (1967), one of several Supreme Court decisions from the 1960s under the Federal Securities Laws that called for a broad and flexible application of the securities laws. *Id.* By the mid-1970s, the Supreme Court’s approach to interpretation of the securities laws had narrowed tremendously. *See, e.g.*, *Chiarella v. United States*, 445 U.S. 222 (1980) (rejecting Court of Appeals’ extension of prohibition against insider trading under Section 10(b) of Securities Exchange Act of 1934 and SEC Rule 10b-5 thereunder beyond the traditional context of persons in relations of trust and confidence with purchasing or selling party and persons having an affirmative duty to speak); *Santa Fe Indus., Inc. v. Green*, 430 U.S. 462 (1977) (rejecting the Court of Appeals’ broad reading of Section 10(b) of Securities Exchange Act of 1934 and SEC Rule 10b-5 thereunder as reaching certain breaches of fiduciary duty in connection with securities transactions; holding deception or manipulation necessary for any violation); *Ernst & Ernst v. Hochfelder*, 425 U.S. 185 (1976) (rejecting Court of Appeals’ broad application of Section 10(b) of Securities Exchange Act of 1934 and SEC Rule 10b-5 thereunder as reaching certain negligent acts in connection with purchases or sales of securities; scienter held necessary for any violation).

75. 414 F.3d at 1371 (emphasis added). The two-judge majority continued: Section 271(f), “if it is to remain effective, must . . . be interpreted in a manner that is appropriate to the nature of the technology at issue.” *Id.*

76. *Id.* at 1369.

the panel majority deemed “all the resulting copies,” for purposes of Section 271(f), “essentially [to have] been supplied from the United States.”<sup>77</sup> The panel also rejected any different treatment for the software supplied on the “golden masters” as opposed to that transmitted electronically.<sup>78</sup>

This two-judge majority decision of the *AT&T* panel provoked a strong dissent, based on the foreign copying issue, from Judge Rader, the author of the recent *Eolas* decision. That is ironic because Judge Rader’s own opinion in *Eolas* had fully endorsed the application of Section 271(f) to the same essential facts as were presented in *AT&T*. In particular, the same foreign copying issue was present—though the parties apparently did not emphasize it—in *Eolas*, but neither Judge Rader nor any other member of the *Eolas* panel made any mention of it there. In effect, Judge Rader had reached the same conclusion in *Eolas* as the one he strenuously rejected in *AT&T*.<sup>79</sup>

Judge Rader’s dissent is notable for its traditional, rather than expansive, approach. For one thing, he refused to ignore the physical processes involved in software replication and installation. Effectively, he declined to treat the tangible embodiment of the software as completely unimportant compared to its intangible aspects. Accordingly, he viewed copying or “replicating” the software abroad as foreign manufacturing and therefore incapable of being “subsumed” in the act of “supplying” it from the U.S.<sup>80</sup>

He also emphasized the traditional view that patent protection in foreign markets requires inventors to secure patents in the respective foreign countries. The panel majority’s ruling, he said “disregards the existing international scheme of patent law” with potentially serious consequences extending well beyond the software industry.<sup>81</sup> Pursuant to a stipulation in the case, he noted, Microsoft paid damages for every copy of its accused software loaded onto computers in the United States. The matter before the court, he said, concerned instances where the software was installed

---

77. *Id.* at 1369-71.

78. *Id.* Liability under Section 271(f), the majority said, “does not depend on the medium used for exportation.” To treat these methods of supply differently, it said, would exalt “form over substance.” *Id.* at 1370.

79. The complexity of the *Eolas* appeal and the parties’ apparent failure to emphasize the foreign copying issue help explain the oddity of Judge Rader’s opposite conclusions. In *Eolas*, the court devoted most of its attention to difficult issues of claim construction, anticipation, obviousness, and inequitable conduct. The court’s rulings on Section 271(f) in *Eolas* came as a relatively short segment at the end of a long opinion and will affect the ultimate result in the case only if the anticipation, obviousness, and inequitable conduct questions are all decided against Microsoft.

80. *Id.* at 1373-74.

81. *Id.* at 1372.

“in Düsseldorf or Tokyo” for foreign sale and use. In those cases, he said, foreign installation may constitute infringement of AT&T’s rights, but only if AT&T held a German or a Japanese patent.<sup>82</sup>

Finally, Judge Rader objected to the potentially massive liability Microsoft faced, compared with the very small number of “golden masters” or electronic transmissions it had sent abroad. “One act of supplying,” he said “cannot give rise to liability for multiple acts of [foreign] copying.” He therefore decried the “endless liability in the U.S.” that the majority’s decision imposed “for products manufactured entirely abroad.”<sup>83</sup>

Microsoft sought Supreme Court review in the *AT&T* case, as it had done, unsuccessfully, in *Eolas*. After first inviting the Solicitor General to present the views of the United States, the Supreme Court recently granted Microsoft’s petition for review.<sup>84</sup> The two questions before the Court are: (i) whether computer software qualifies as a “component” for purposes of Section 271(f); and, if so, (ii) whether “supplied . . . from the U.S.” includes foreign-produced copies of the U.S.-sourced software that are loaded onto computer hardware at foreign locations.<sup>85</sup>

As this article went to press, only Microsoft and the various amici supporting it had filed their briefs on the merits of the case. Presenting the views of the United States, the Solicitor General’s (“SG’s”) brief carries special weight. As in his September, 2006 amicus brief recommending that the Court grant certiorari in the case, the SG argues for essentially the same positions that Judge Rader took in his *AT&T* dissent, namely: (i) that software qualifies as a Section 271(f) “component” but (ii) that foreign-made copies of software supplied from the United States do *not* qualify as “supplied” from the United States for purposes of that provision. The case does not directly present, and it appears unlikely that the Court will directly address, the separate question, discussed above, of whether Section 271(f) applies to method (process) claims. Nevertheless, a narrow interpretation of Section 271(f) “components” as including only tangible objects would tend to discourage application of that provision to method claims.<sup>86</sup> Unfortunately, any of the results the Supreme Court is likely to

---

82. *Id.*; see also *id.* at 1373.

83. *Id.* at 1372-73.

84. *Microsoft Corp. v. AT&T Corp.*, 127 S. Ct. 467 (2006)

85. See *Petition for a Writ of Certiorari, Microsoft Corp. v. AT&T Corp.*, 2006 WL 403897 (Feb. 17, 2006) (No. 05-1056).

86. Under Judge Rader’s *Eolas* reasoning, the “components” of a method invention are “steps” or “acts.” Since those are intangible, a requirement that Section 271(f) components be tangible would seem to rule out application of that provision to method inventions. Judge Rader’s reasoning in *Union Carbide*, however, might lead to the opposite result. There the Federal Circuit panel applied Section 271(f) to a method claim where

reach in this case will have at least some ironic and undesirable aspects and will fail to solve the fundamental defects of Section 271(f).<sup>87</sup>

#### D. Components Designed but Not “Produced” in the United States—U.S. Companies Moving Production Off-Shore

Another recent Federal Circuit case, *Pellegrini v. Analog Devices, Inc.*,<sup>88</sup> rounds-out our review of Section 271(f) in operation. While *Pellegrini* and a few similar decisions<sup>89</sup> recognized limitations on the application of Section 271(f), the nature of those limitations underscores—indeed strengthens—the perverse incentives Section 271(f) creates against invest-

---

the “component” supplied from the U.S. was a tangible substance that operated as a catalyst in the claimed chemical process. See *supra* notes 53-54 and accompanying text regarding the logical inconsistencies in the *Eolas* and *Union Carbide* reasoning.

87. The worst result, in terms of economic consequences, would be to uphold the ruling by the Federal Circuit panel majority. This would maximize the perverse incentives Section 271(f) creates for current and prospective U.S. producers and the discrimination in favor of off-shore production.

A ruling in accordance with the views stated by the SG and Judge Rader would lead to a less harmful result but one still ironically inconsistent with a stated objective of Section 271(f) and far less favorable to the U.S. economy than the fact pattern and result in *Deepsouth*. Specifically, one stated objective of Section 271(f) was to “avoid encouraging manufacturing abroad.” See *supra* note 26. Finding no infringement because foreign-made copies are not “supplied from the U.S.” would protect U.S. software producers from Section 271(f) liability in many cases, but would encourage (a) copying of U.S. made software abroad onto (b) foreign made hardware there and (c) final assembly there of the overall computer systems or software combination products. In *Deepsouth*, by contrast, all of the “manufacturing” except for the trivial final assembly of the U.S.-made parts was done in the U.S. with the associated employment, tax-generation, and other benefits to the U.S. economy.

The other possibility, a ruling that software cannot qualify as a “component” for Section 271(f) purposes, would likewise reduce the adverse consequences for the U.S. economy, but would re-subject Section 271(f) to one of several criticisms that a leading patent law scholar has leveled at it. Such a ruling would probably involve a determination that only tangible, physical items can qualify as Section 271(f) “components.” Thus, software and all types of data would be excluded, as would steps and acts, the elements of method (process) claims. While good for U.S. software and data providers, such a conclusion would re-create what Professor Chisum has identified as an essentially irrational feature formerly thought to characterize the coverage of Section 271(f), namely its applicability to claims to apparatus and compositions of matter but not to process (method) claims. See *supra* note 40. In short, only repeal of Section 271(f) will eliminate the inconsistencies and adverse consequences of Section 271(f).

88. 375 F.3d 1113 (Fed. Cir. 2004).

89. *Id.* at 1115-17; see e.g., *Rotec Indus. Inc. v. Mitsubishi Corp.*, 215 F.3d 1246, 1257-58 (Fed. Cir. 2000) (reaching a result similar to that in *Pellegrini* on slightly less dramatic facts).

ing in productive facilities in the United States and in favor of moving U.S. jobs and technology to other countries.

Pellegrini held a U.S. patent on brushless drive circuits for electric motors. The defendant, Analog Devices, was organized and headquartered in the United States. Analog conceived and developed integrated circuit chips for use with other components in combinations that allegedly fell within Pellegrini's patent. Analog controlled and arranged all aspects of the allegedly infringing activity from the United States, but it did not manufacture the chips in the United States, let alone combine them into infringing combinations here. Rather, after designing and perfecting the chips in the United States, Analog both manufactured them at its own factory in Ireland and contracted for their production under proprietary arrangements at two different facilities in Taiwan. It then assembled and tested the combinations in several non-U.S. locations. Analog provided financing and management from the United States, set production levels in the United States, and sent instructions to the off-shore locations for the production and disposition of the chips. It received purchase orders from, and invoiced, customers worldwide and received payments for all of the products sold worldwide. Analog sold and shipped most or all of its chips, however, from Analog's foreign facilities to foreign purchasers for use in foreign countries.<sup>90</sup>

Relying on Section 271(f), Pellegrini charged Analog with infringement and sought damages for world-wide sales of the resulting combinations that included the chips. Analog moved for summary judgment of non-liability for all chips sold to foreign locations on the ground that Section 271(f) applies only where some significant part or parts are physically produced in, and shipped from, the United States. The Federal Circuit panel unanimously agreed with Analog despite the seemingly broad phrasing of the key clause in Section 271(f).<sup>91</sup> The court held that the location of production was key to whether Section 271(f) could apply.<sup>92</sup> Further,

---

90. 375 F.3d at 1115, 1117.

91. *Id.* at 1117-18. The statutory language states an infringer includes anyone who "without authority supplies or causes to be supplied in or from the United States [the key components] . . ." § 271(f)(1)-(2). In effect, the court ruled that "in or from the United States" modifies only "supplies or . . . supplied" and not "causes."

92. The court stated:

[Section] 271(f) is clear on its face. It applies only where components of a patent[ed] invention are physically present in the United States and then either sold or exported. . . . The plain language of § 271(f)(1) focuses on the location of the accused components, not the accused infringer. . . . [T]here can be no liability under § 271(f)(1) unless components are shipped from the United States for assembly.

the court said, it made no difference whether Analog produced the chips at its own factory in Ireland or contracted with other offshore companies for their production. The court stated that Analog's successful defense was "not based on the use of a contractor, but rather on [foreign] manufacture."<sup>93</sup> In other words, Section 271(f) attacks only U.S. production followed by shipment abroad. Companies—foreign or U.S.-based and regardless of their activities in the United States—can avoid all risks and exposure under Section 271(f) if they produce their components anywhere but in the United States. At least one other recent Federal Circuit decision supports this conclusion.<sup>94</sup>

### E. Effects of Section 271(f) Extensions

In summary, despite some dissenting voices, the Federal Circuit is applying Section 271(f) expansively. Under recent rulings, that provision reaches all types of products and all types of patents, and the triggering "components" can be anything in a broad range of tangibles and intangibles, including a broad but ill-defined range of acts, substances, and other items that comprise or facilitate steps in claimed processes. Moreover, while *Pellegrini* limited Section 271(f) in one respect, that limitation benefits only those entities that carry out their productive activities *outside* the United States. It provides no comfort whatsoever for entities producing in the United States. Indeed, by emphasizing that a company—whether based in the United States or abroad—can achieve complete safety under Section 271(f) by locating its production facilities abroad or contracting-out production of key components to foreign producers, *Pellegrini* increased the incentives for companies *not* to invest in, and not to conduct, production in the United States.

The effects of Section 271(f) in recent cases have been dramatic. The \$565 million damages award (including some interest) in *Eolas*, for example, was the second largest in U.S. patent law history at the time. On remand, the court will presumably increase this award, adding additional interest, if it finds the patent valid and enforceable. Of particular importance here, *two-thirds* of that total dollar amount is potentially recoverable under the U.S. patent and in a single U.S. lawsuit solely because of Section 271(f). Similarly, inclusion of Section 271(f) claims in *Union Carbide* doubled the damages exposure there as well as increasing the reach of the resulting injunction.

---

*Pellegrini*, 375 F.3d at 1117 (emphasis added).

93. *Id.* at 1118.

94. *Rotec*, 215 F.3d at 1257-58.

As important as the dollar amounts involved is the wide publicity that has attended the *Eolas* ruling and other recent decisions under Section 271(f). Awareness of substantial patent awards on the part of corporate managers and legal advisers can substantially influence corporate attitudes regarding the risks of patent infringement suits. For example, the *Kodak-Polaroid* patent litigation of the late 1980s and early 1990s resulted in heavy monetary and injunctive consequences adverse to Kodak and in extensive related publicity. Many people credit corporate America's awareness of those consequences with a fundamental shift in the thinking of U.S. corporate executives and their advisors regarding the strength of U.S. patents and the dangers of infringement litigation.<sup>95</sup> The wide publicity of *Eolas* and other recent Section 271(f) decisions, together with the heavy damages awards and broad injunctive remedies that courts may impose in those cases may have similar effects. The result could be enhanced corporate concerns regarding the dangers of infringement under Section 271(f), with a consequent chilling effect on investments in U.S. productive facilities and a further shift of technology sector jobs to foreign countries.

As noted above, judicial review in this area is incomplete. The far-reaching *Eolas* and *AT&T* decisions reflect the views of only a few of the Federal Circuit's twelve judges. Recently, other Federal Circuit judges have disagreed with the *Eolas* holding that Section 271(f) applies to method patents.<sup>96</sup> Moreover, the Supreme Court recently granted Microsoft's petition for review of the Federal Circuit's decision in *AT&T*. Unfortunately, the greatest plausible cut-backs to Section 271(f) that Federal Circuit en banc review or Supreme Court review could effect would reduce but not eliminate the perverse effects of Section 271(f).<sup>97</sup>

---

95. FTC REPORT, *supra* note 19; ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW THE BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS AND WHAT TO DO ABOUT IT 112-115 (2004) [hereinafter INNOVATION AND ITS DISCONTENTS].

96. See the discussion of the Federal Circuit's revised opinion in the BlackBerry case *supra* note 51, and the discussion of the dissent by several Federal Circuit judges from the denial of rehearing and rehearing en banc in *Union Carbide Chemicals & Plastics Technology Corp. v. Shell Oil Co.*, 434 F.3d 1357, 1358-59 (Fed. Cir. 2006), *supra* note 55 and accompanying text.

97. In view of the Federal Circuit judges' recent statements, it seems plausible that en banc review by the Federal Circuit in some future case under Section 271(f) might reverse the rulings that the provision can apply to method claims. See, for example, the statements quoted *supra* notes 51 and 55. Supreme Court review in some future case, of course, could go beyond that, but the questions presented to the Court in the pending *AT&T v. Microsoft* case are narrower, namely: (i) whether software object code can be a "component" for purposes of Section 271(f) and (ii) if so, whether copies of that object

#### IV. ASSESSING SECTION 271(F)

This Part examines the merits and demerits of Section 271(f). The criteria used are: (i) the objectives of the patent laws—namely the promotion of the “progress of science and useful arts”;<sup>98</sup> (ii) considerations of fairness and efficiency in patent enforcement; and (iii) related, widely recognized goals of U.S. trade and economic policy such as promoting U.S. competitiveness in international markets and enhancing employment and investment in the United States.<sup>99</sup> A preliminary question confronting any such evaluation is whether we have enough information on which to base such an assessment. Accordingly, this Part addresses the following matters:

- Whether the information available to us with regard to the operation of Section 271(f) is sufficient to support an evaluation in which we can have any confidence;
- The extent to which Section 271(f) generates beneficial incentives for invention and/or commercialization of new technology;
- The risks and burdens U.S. producers face as a result of Section 271(f);
- Other possible arguments in favor of Section 271(f), including considerations of fairness and efficiency in patent enforcement; and
- A “bottom line” assessment.

Thereafter, a short final Part of this Article addresses some new areas where we may expect Section 271(f) to be applied in the future and some special aspects of that provision in contemporary patent enforcement.

---

code are “supplied” from the United States when those copies are created overseas by replicating a separate master version supplied from the United States.

For reasons mentioned above, even the broadest cut-backs of Section 271(f) that the Federal Circuit or the Supreme Court might plausibly make would reduce, but not eliminate, the perverse incentives that Section 271(f) creates. *See supra* text accompanying note 57.

98. U.S. CONST. art. I, § 8, cl. 8.

99. The legislative history identifies promotion of U.S. manufacturing and reduction of U.S. unemployment as objectives of Section 271(f). *See supra* note 28 and accompanying text.

**A. Do We Have Sufficient Information on Which to Base an Evaluation of Section 271(f)?**

Ideally, an assessment of Section 271(f) would have the benefit not only of a legal and historical review but also of economic analyses of that provision in operation. If possible, this would include, among other things: (i) data and case studies on how Section 271(f) has operated in different circumstances and technological areas, including the nature of the U.S.-supplied components in relation to the patented inventions and of why those components were not themselves covered by a U.S. patent (which would have made Section 271(f) superfluous); (ii) a comparative economic analysis of the U.S. patent system (with and without Section 271(f)) and at least a few important foreign patent systems (which lack any provision similar to Section 271(f)); (iii) an assessment of the costs and effectiveness of the patent enforcement mechanisms in other countries (which is the traditional method of obtaining patent protection in foreign markets); and (iv) analyses of whether and why alternative foreign sources of the U.S.-supplied components have not been readily available and how long this condition has lasted in particular product and geographic markets. This and similar information might provide a definitive picture of how Section 271(f) affects U.S. interests in such areas as innovation, trade, technology investment, facility siting, jobs creation, and international competitiveness. As far as the author knows, such analyses are not currently available.

For several reasons, however, the absence of such analyses need not prevent us from making important judgments as to Section 271(f) or from having confidence in those judgments. First, a wealth of information on the operation of the U.S. patent system has recently become available as a result of several recent, wide-ranging studies. The associated reports illuminate the U.S. patent system and its operation, including its shortcomings and changes that might improve its performance. The U.S. Federal Trade Commission sponsored and prepared one of these, releasing the associated report in 2003 (“FTC Report”).<sup>100</sup> The National Research Council, an arm of the National Academies (formerly the National Academy of Sciences), sponsored another and published the resulting report in 2004 (“NAS Report”).<sup>101</sup> A number of other studies and reports on the U.S. patent system,

---

100. FTC REPORT, *supra* note 19.

101. NAS REPORT, *supra* note 19. This report was the result of initiatives of the Governing Board of the National Academies and its Board on Science, Technology, and Economic Policy. They commissioned a Committee on Intellectual Property Rights in the Knowledge-Based Economy, and, under their auspices, multiple meetings, public fora, and workshops were held; nine policy-related empirical studies (selected from more than

generally less extensive than these two but nevertheless substantial, have likewise appeared in the last two years.<sup>102</sup>

These studies included the commissioning of a number of scholarly inquiries and reports on aspects of the patent system as well as the receipt of numerous presentations from academics, corporate managers, intellectual property directors, legal and technical experts, and others. Additional information on the patent system has been published in connection with these two studies and in connection with the extensive hearings on patent reform in the two most recent sessions of Congress. Commentators have published various other reports, articles, and at least one book in the last two years on the U.S. patent system and its operation and problems.<sup>103</sup> None of these studies, reports, or articles seems to have focused substantially on Section 271(f). Nevertheless, they provide extensive, detailed, and apparently reliable information regarding the operation, enforcement, and effects of the patent laws, much of which is relevant to an assessment of Section 271(f).

A second aid in assessing Section 271(f) is the incentives-based rationale of the patent laws: the patent laws are all about incentives. Moreover, courts have fleshed out the legal application of Section 271(f) to a significant degree now, and its basic operation is relatively clear in most

---

eighty proposals) were sponsored; testimony was received from invited speakers including economists, legal scholars and practitioners, judges, former PTO officials, and business managers; and information was gathered in other ways.

102. In addition to the two reports noted above and the commentary and hearings mentioned *infra* note 103, each of the following organizations, among others, sponsored a substantial study and report on the U.S. patent system that has been first published within the last two years: the Brookings Institution, the Council on Foreign Relations, the Congressional Research Service/Library of Congress, and the Intellectual Property Law Section of the American Bar Association. Copies of these reports are generally available on the respective sponsoring organization's website.

103. Apart from the FTC and NAS Reports and the separate studies/reports referred to *supra* notes 19 and 102, responses to the FTC and NAS reports have been published by organizations such as the American Intellectual Property Law Association and the Intellectual Property Law Section of the American Bar Association. In connection with the NAS and FTC Studies and their recommendations for reform, the House Subcommittee on the Courts, the Internet, and Intellectual Property and the Senate Subcommittee on Intellectual Property held a number of hearings throughout 2005 and 2006 on patent reform and problems with the U.S. patent system. The House Subcommittee staff and others have published and/or introduced several proposed bills. This activity, including the main points of the testimony and presentations made in these hearings, has been widely reported by the Bureau of National Affairs (BNA), Commerce Clearinghouse (CCH), and others. A considerable range of additional commentary has also been published recently on the same subject including numerous articles and at least one widely discussed book by two economists, *INNOVATION AND ITS DISCONTENTS*, *supra* note 95, at 112-15.

respects. Accordingly, one can understand the risks, rewards, and incentives Section 271(f) is likely to produce for inventors, investors, and producers in at least a general way and on average. In addition, large degrees of uncertainty are inherent in the application and enforcement of patents and, to an even greater extent, in the early stages of invention and commercialization of technological advances. As far as the resulting incentives are concerned, these uncertainties swamp much or all of the details that one might analyze in careful empirical studies of how Section 271(f) has operated in specific cases. This is especially true because the incentivizing effects of the patent laws must operate at the early inventive and investment stages of innovation, whereas patent enforcement can occur, and Section 271(f) might actually operate in court, only later and after these large uncertainties have had their effects. Accordingly, inventors, backers, and corporate managers can only take into account the general and average effects of Section 271(f) at the early inventive and investment stages when the incentivizing effects of the patent laws must operate, if they are to operate at all.

Third, the real effects of Section 271(f)—whether in stimulating inventive activity and commercialization of technological advances or, on the other hand, inducing companies to invest and produce off-shore—depend on how persons in the inventive, industrial, and investment sectors and their legal advisors *perceive* that section and its role in the overall patent system. Like us, these persons lack the empirical data and economic analyses that would be desirable for a full, rigorous analysis. Accordingly, such sophisticated analyses are, in an important sense, irrelevant to how Section 271(f) is actually influencing incentives, production planning, and choices of locations for investment, facilities construction, and jobs creation. The more relevant factors, in this regard, are the publicized results of Section 271(f) in operation and the advice that executives receive from their legal and other advisors.

Fourth and finally, detailed analyses—indeed any analyses at all—of economic effects or impacts on innovation have been starkly absent from both enactments and amendments of the patent statutes and major judicial interpretations of the patent laws. No meaningful analysis accompanied the enactment of Section 271(f).<sup>104</sup> A generally similar situation has characterized many legislative actions revising the U.S. patent laws. Congress generally acts, in this and many other areas, on the basis of very limited

---

104. The majority and dissenting opinions in *Deepsouth* each provided more substantial discussions of the relevant issues and policy considerations than appeared in the legislative history of Section 271(f). See *supra* notes 13-18, 24, 26, 29 and accompanying text.

analysis. Usually, their data comprise no more than general testimony and/or reviews of prior court decisions, and the accompanying committee reports sometimes fail to make clear whether important lines of existing case law are being approved or disapproved. Further, political pressures tend to play larger roles than economic or other analyses. Likewise, as some of the judges on the Federal Circuit have noted, that court usually decides major issues of patent law without the benefit of economic analyses and without any data on the likely effects on innovation or economic activity, even though the purpose of the patent laws is to affect those very matters.<sup>105</sup>

Thus, despite our lack of empirical data and economic analyses, we have the same information that inventors, managers, investors, and their legal advisors use when they include Section 271(f) in their decisions af-

---

105. See, e.g., *Hilton Davis Chem. Co. v. Warner-Jenkinson Co.*, 62 F.3d 1512, 1529-32 (Fed. Cir. 1995) (en banc) (Newman, J., concurring) (noting (i) the importance of the doctrine of equivalents to the “national interest in technologic innovation” and to the effectiveness of the patent system in incentivizing innovation but (ii) the lack of available “policy exploration, economic analysis, [etc.]” in these areas and (iii) the failure of the Federal Circuit, the parties, or the amici curiae to discuss these subjects in that important case. That and other cases similarly note the lack of express clarifications in major codifications of the patent laws as to whether major lines of prior case law are being reaffirmed or rejected); see also *Warner-Jenkinson Co. v. Hilton Davis Chem. Co.*, 520 U.S. 17 (1997) (rejecting arguments that the re-codification of the Patent Act in 1952 and the various earlier changes in claiming requirements had undercut or overturned a long line of mainly older cases affirming the viability of the doctrine of equivalents where the legislation and its legislative history were silent on that question). Compare the disagreements among the Federal Circuit’s decisions and among its judges on the proper approach to the fundamental process of patent claim construction, which is described and largely resolved in the recent en banc ruling in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005) (en banc) but without any economic or other real-world analysis being provided to aid in that decision. Finally, note the recorded statement by Chief Judge Michel of the Federal Circuit at a 2002 conference on patent reform, regarding recent litigation on the issue of the doctrine of equivalents in the leading case of *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 234 F.3d 558 (Fed. Cir. 2000) (en banc):

Now you might have thought . . . where there was a concern about the relative needs to promote adequate incentives—or you could say fairness to inventors—on the one hand with the need for competitors to have adequate predictive value and certainty on the other hand, that somebody, at least *amici* and, one would hope, also the parties, would have given us some . . . meaningful data about that. Now I read all the briefs . . . and I can’t remember anything that I would consider empirical data. . . . If you trace back the pedigree, I suspect that you will find that in a great many cases there never was any meaningful economic or qualified data.

NAS REPORT, *supra* note 19, at 86 n.14.

fecting innovative efforts, investments, and the locations of productive facilities and activity. Moreover, this amounts to far more information than Congress had in 1984 in enacting Section 271(f). Twenty years of experience and “hindsight” give us a reasonable basis for assessing that provision, particularly when compared with the lack of analysis that underlay its enactment.

### **B. Benefits—Does Section 271(f) Help Incentivize Innovation?**

The basic aim of the U.S. patent laws is to “promote the progress of science and useful arts,” both by incentivizing invention and commercialization of technological advances—including the investments needed for those purposes—and by encouraging full public disclosure of discoveries that would otherwise remain secret.<sup>106</sup> Section 271(f), however, can provide no significant incentives or encouragement along these lines.

Initially, of course, Section 271(f) contributes nothing to protection of U.S. patent holders with respect to exploitation of their inventions in U.S. markets. It was never intended to provide additional protection in domestic markets, and its terms add nothing to the other provisions of the Patent Act in that regard.<sup>107</sup> Two other points then prove the absence of any incentivizing benefits from Section 271(f). First, at the relevant times, that provision provides no plausible promise of useful protection against competition—or even of the right to exploit the invention—in foreign markets. Second, in the relatively unusual cases where one or more foreign markets might actually be material in incentivizing the inventor or his/her financial backers, those persons will have obtained alternative—and substantially better—protection in the form of patents in the relevant jurisdiction, thereby making Section 271(f) superfluous. Even easier, if one or more key component(s) of an invention is itself new and patentable, the inventors likely will have rendered Section 271(f) superfluous simply by claiming that or those key components, as well as the larger invention, in their U.S. patents.

---

106. The U.S. patent laws are based on Article I, Section 8, Clause 8 of the U.S. Constitution, which provides that Congress shall have the power “To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” To promote the progress of science and the useful arts, the patent laws primarily incentivize invention and the commercialization of technological advances and secondarily encourage the disclosure of information about inventions and how they work. Such disclosure assists persons other than the inventor to understand and learn from the patented invention and thereby further advance the art. NAS REPORT, *supra* note 19, at 35-36, 39, 63-64.

107. See *supra* notes 5, 6 and accompanying text.

The first of these points—that Section 271(f) can provide no meaningful prospect, at the relevant times, of protection against competition in foreign markets—may not have been completely clear when that provision was enacted in 1984. Developments in international competitiveness and legal developments such as the *Pellegrini* case, however, have now made that clear.

The basic problem, as *Pellegrini* shows, is that Section 271(f) offers holders of U.S. patents protection in foreign markets only against competition from products containing components produced within the United States. Thus, as to any foreign market, Section 271(f) provides no protection whatsoever against any components or resulting combinations made in that country, or any other country, as long as the key component(s) were not made in the United States. In fact, as *Pellegrini* illustrates, Section 271(f) provides no protection against competition in any foreign market, even by a U.S. company that designs the key component(s) and the full combination in the United States, provides financing and management from the United States, directs production and sales from the United States, combines the components into “infringing” combinations abroad, and sells those combinations in foreign markets, as long as that company *either* produces the relevant component(s) in its own non-U.S. facilities *or* obtains them from foreign producers. The same lack of protection, of course, applies to companies headquartered outside the United States. They can compete with the U.S. patent holder in all foreign markets free of any restraint from the U.S. patent laws as long as their production occurs in some country other than the United States.

In 1972, the *Deepsouth* dissenters postulated that U.S. manufacturers generally had competitive advantages over producers in other countries.<sup>108</sup> If this were true, restraining competition only from U.S. producers might provide the holder of a U.S. patent with a meaningful degree of protection in foreign markets. For some years following World War II, of course, U.S. manufacturers did have advantages over their foreign counterparts in many types of production. Now, however, a trip to Wal-Mart or to a car dealership, camera shop, hardware store, TV or computer store, etc., or a perusal of the *Wall Street Journal* shows that any such general manufacturing advantages disappeared long ago in most if not all fields.

*Pellegrini* illustrates this disappearance: the components at issue were semiconductor integrated circuit chips, a type of product and a technology that were developed mainly in the United States. Yet the U.S. company accused of infringement sourced them in two ways, neither of which in-

---

108. See *supra* note 16 and accompanying text.

volved U.S. production. There may still be a few areas where U.S. producers have strong positions world-wide, such as software and, perhaps, aircraft manufacturing and certain pharmaceuticals. But the strengths of programmers in Bangalore, Eastern Europe, and other foreign locations are well known, as are the competitive strength of Airbus and other foreign aircraft manufacturers, the difficulties of McDonnell-Douglas, Lockheed-Martin, and other former U.S. aircraft manufacturers, and the strengths of European pharmaceutical manufacturers. Further, the continuing viability of some U.S. producers may be partly the result of historical circumstances (including distribution arrangements, installed base, network effects, etc.), not of advantages in the manufacturing of components. In short, foreign producers in most fields are now strong and often enjoy substantial cost advantages over U.S. producers—and Section 271(f) gives U.S. patent holders no protection whatsoever against any of them.

A timing consideration further undercuts any protection Section 271(f) might seem to offer in foreign markets. The incentivizing effects of the patent laws must operate at the early inventive and investment stages of innovation if they are to bear any fruit. Possible monetary returns, and the need for effective patent protection, in contrast, come years later. Thus, the patent laws can incentivize innovation only if meaningful protection appears likely to be available *some years after* the initial inventive efforts and associated investments. Traditional territorial patent protection fits very well with this timing requirement, but any “protection” under Section 271(f) fails this test miserably.

To be specific, a foreign patent, if obtained, lasts for a fixed term of reasonable duration, typically twenty years from the date of the original patent application. Likewise, a U.S. patent claiming key patentable components provides a similar term of protection. Any protection available under Section 271(f), however, is very different. If no foreign producers could offer competition in foreign markets for a given product today, that has no bearing on whether one or more foreign producers could do so five or more years in the future, when inventive efforts and investment might pay off and protection from competition would be needed. In other words, while traditional patent protection is known to last long enough to promote returns on innovation, one would not expect protection under Section 271(f) today still to be effective when it will be needed. Thus at the relevant times, any possibility of protection from Section 271(f) will appear not only weak and speculative in the short term but also wholly speculative or unlikely as to the period when it might be helpful.

The second point mentioned above is equally important: in the cases where prospects of returns from one or a few foreign markets actually

would operate to incentivize an inventor who has applied or will apply for a U.S. patent, that inventor very probably will obtain a foreign patent or patents. Note, initially, that Section 271(f) can apply only where a U.S. patent exists, so we must assume the inventor is applying or will apply for a U.S. patent and that it will be granted. This situation then poses two questions: (i) whether foreign markets would actually help incentivize that inventor or investor; and (ii) if so, what those persons can and will do as to getting patents in such countries.

In order to play a role in incentivizing innovative efforts, a specific foreign market must be a focus of actual interest on the part of the inventor or investor at the time of invention or investment. This is often the case with regard to the U.S. market because it is the most important economically for most inventions. Focus on the U.S. market is particularly strong on the part of the persons that, in furtherance of U.S. interests, we most want to incentivize, namely U.S. inventors and investors.<sup>109</sup> After the United States, and particularly after Japan and the countries participating in the European Patent Office agreement, market sizes—and interest in the corresponding patents—drop off rapidly.<sup>110</sup> Thus, while inventors may

---

109. The patent laws are intended to promote progress in science and the useful arts, and this, in concept, can be achieved not only by incentivizing U.S. individuals and companies but also by incentivizing foreign inventors and investors by offering them, if successful, patent-based monopolies in the U.S. markets. The U.S. patent laws do that because they do not favor U.S. inventors or patent owners over non-U.S. inventors or owners as to the availability of U.S. patents or as to the remedies available for infringement. But, where the inventor/owner is a non-U.S. person or company, the financial rewards of the invention provided from the U.S. markets are more likely to go outside the U.S., and potential U.S. producers are barred from competing here. Accordingly, the cost/benefit relationship is not as favorable for U.S. interests as where the incentive works successfully on U.S. entities. The even-handedness of the U.S. patent laws is more a matter of international agreement and reciprocity than a matter of maximizing progress in innovation: the U.S. grants patent rights to foreign inventors in return for foreign countries giving U.S. inventors and patent owners equal rights to obtain and enforce patents granted by those foreign countries. See *supra* note 10 summarizing certain provisions of the Paris Convention.

110. One indication of this is the relative numbers of applications made for patents to be issued by various countries. Statistical comparisons are difficult for several reasons, including: (i) the existence of international and regional patent cooperation arrangements producing multiple avenues for patent applications and, in some cases, single patent grants effectively covering numerous countries; (ii) variations in the allowable subject matter coverage of individual patents causing multiple patents to be used in some countries where a single patent might cover the same subject matter in others; and (iii) differing pendency periods for applications filed in different countries. Generally, however, the European participants in the European Patent Office together with the United States and Japan account for between 70% and 88% of the world-wide patent applications submitted annually, patents granted annually, and patents in force, with all other countries account-

have special interests in particular markets, it is unlikely that more than a very few markets other than the U.S. market will assume actual significance for incentivizing purposes.<sup>111</sup>

Applying for patents in many countries can increase the costs the inventor or his/her backers must bear, but several things mitigate this as to the inventor who applies for a U.S. patent. First, as noted just above, it is unlikely that more than a very few markets, other than the United States, will assume actual significance for incentivizing purposes. Therefore, we need not consider the costs of obtaining patents in large numbers of jurisdictions other than the United States but only in a few.

Second, by far the biggest cost in securing a patent is for legal costs for the preparation of the written description, claims, and other portions of the application.<sup>112</sup> Once that is done for one country, assumed here to be the United States, the identical or nearly identical document, perhaps with translation, is generally used in any other countries where patents are sought, substantially reducing the incremental cost of applying in a second

---

ing for the remainder. *See, e.g.*, EUROPEAN PATENT OFFICE, U.S. PATENT & TRADEMARK OFFICE & JAPAN PATENT OFFICE, TRILATERAL STATISTICAL REPORT (2005), *available at* [http://www.trilateral.net/tsr/tsr\\_2005/](http://www.trilateral.net/tsr/tsr_2005/) [hereinafter TRILATERAL STATISTICAL REPORT]; ORGANISATION FOR ECONOMIC CO-OPERATION & DEVELOPMENT, COMPENDIUM OF PATENT STATISTICS (2005), *available at* <http://www.oecd.org/dataoecd/60/24/8208325.pdf> [hereinafter COMPENDIUM OF PATENT STATISTICS].

111. *See supra* notes 109, 110. Further, U.S. inventors and patent owners generally have greater familiarity with U.S. markets and U.S. sources of financing than with those in foreign countries. This familiarity further contributes to their concentration on U.S. markets in evaluating their chances of monetary success in their inventive and investing activities. Indeed, based on some experience with inventors and considerable experience in venture capital financing, development joint ventures, and early stage acquisitions, it is the author's view that a majority of U.S. inventors and smaller U.S. companies consider U.S. patents but not foreign patents in considering the likelihood of returns from their inventive or IP-related investing activities. Further, if foreign patents are considered at all, interest is limited to a very few foreign countries. It is rare indeed for the markets in numerous foreign countries each to play a meaningful part in incentivizing these persons with respect to inventive efforts, early stage financing, or early-stage venture or acquisition planning. A similar attitude is evident even in the patenting practices of some major U.S. companies such as IBM, which has recently described its tendency to patent a high percentage of its inventions in the U.S. but to patent its inventions much more selectively in foreign countries. *See* Thomson, Rankings of the Most Innovative Companies by Patent Filings, <http://scientific.thomson.com/free/ipmatters/bti/8199727> (last visited Nov. 5, 2006) (quoting Takako Yamakura, Public Relations Manager, Technology & Research Communications at IBM).

112. "Legal counsel [fees] represent[] the vast majority of [the cost of obtaining a patent]." NAS REPORT, *supra* note 19, at 38.

or third country.<sup>113</sup> In addition, a prospective patentee must respond to objections that the U.S. Patent and Trademark Office (“PTO”) raises due to pieces of prior art. The prospective patentee can usually reuse that response to similar objections that a foreign patent examiner makes, further lessening the incremental costs of obtaining foreign patents. Moreover, even without these savings, the total costs of obtaining a patent in the United States is, for a variety of reasons, simply “well above the range of those [costs] in other industrial countries.”<sup>114</sup>

In addition, the Patent Cooperation Treaty (“PCT”), sponsored by the World Intellectual Property Organization (“WIPO”), further eases the burdens of obtaining patent protection for such foreign markets as an inventor may select. As of October 2006, the United States and 132 other countries were parties to the PCT.<sup>115</sup> Under its provisions, inventors filing a patent application in the United States can make a single, cheap, international filing and thereby: (i) obtain an effective filing date, in all participating countries, for purposes of priority and avoidance of statutory bars; and (ii) postpone the decision as to whether to incur most of the relatively modest incremental costs of preparing and filing foreign applications for a substantial period, typically thirty (30) months. This international filing will preserve inventors’ priority rights as to as many countries as they desire for a single modest fee, and will produce an international search report for the inventors and their backers to review.<sup>116</sup> They can then use the thirty months and the search report to determine whether the invention seems commercially viable and protectible before committing, on a country-by-country basis, to incur the relatively modest incremental costs of prosecuting an actual application in one or more countries in addition to the United States. If the invention shows promise in that period, it should be easier for inventors and/or backers to secure any additional financial support they might need to cover the incremental costs of the additional applications.

Accepting these modest incremental burdens gives the inventor and his/her backers considerably more protection in the significant foreign

---

113. See *id.* at 121, for the criticism of the “best mode” disclosure requirement in U.S. patent law that, because it has no counterpart in the laws of other countries, the common practice of simply translating a first-filed foreign patent application for use in the U.S. cannot be used for the best mode disclosure, which increases U.S. filing costs. This problem is absent where, as we assume here, the original application is prepared for filing in the U.S.

114. *Id.* at 68.

115. 1 WORLD INTELLECTUAL PROPERTY ORGANIZATION, PCT APPLICANT’S GUIDE (Annex A 2006), available at [http://www.wipo.int/pct/guide/en/gdvol1/annexes/annexa/ax\\_a.pdf](http://www.wipo.int/pct/guide/en/gdvol1/annexes/annexa/ax_a.pdf).

116. See, e.g., COMPENDIUM OF PATENT STATISTICS, *supra* note 110, at 57.

markets than Section 271(f) can possibly provide. While there are some gaps in effective foreign patent protection, these are primarily in countries having small domestic markets. Moreover, patents issued in foreign countries will protect the inventor against non-U.S. competition as well as against components manufactured in the United States. Such protection will last for the full patent term, not just until non-U.S. producers decide to compete. Thus, assuming the particular foreign market or markets actually figured into the inventor's or financial backer's thinking sufficiently to affect their incentives and conduct, rationality certainly calls for them to seek genuine patent protection for those markets.

Yet another consideration supports this assessment: Section 271(f) does not prevent a separate inventor from applying for and obtaining a patent on the same invention in the relevant foreign countries. The inventor's U.S. application will afford him/her certain priority advantages in obtaining a foreign patent if he/she seeks one. The publication or issuance of the U.S. patent (or other events) might bar a patent application in the foreign country filed after that date by a separate inventor. Before filing the U.S. application and for at least eighteen months thereafter, however, the inventor must assume that he/she may relinquish foreign patent protection unless he/she blocks other patents in those countries by pursuing protection there. Thus, if particular foreign countries are so significant as to figure in the inventors' or backers' actual thinking, they will take at least the first steps to obtain foreign patents.

Statistics bear out these conclusions. Pursuant to U.S. procedural changes mandated by the 1994 TRIPS Agreement, the U.S. PTO publishes most U.S. patent applications eighteen months after they are filed, destroying the secrecy that would otherwise last unless and until the patent issues. An exception to such automatic publication applies if the applicant disclaims the filing of any foreign counterpart patent application(s).<sup>117</sup> Because it preserves secrecy, the applicant has a substantial incentive to file such a disclaimer if he/she will not file any such foreign applications. Yet, tabulations of these disclaimers show that, on average across technologies and years, applicants preserve secrecy as to only about 10-11% of U.S. applications by filing such disclaimers.<sup>118</sup> Moreover, knowledgeable observers estimate that a substantial portion of this 10-11% of applications for which the prospective patentee files no foreign counterparts pertain to marginal inventions not deemed worth patenting abroad.<sup>119</sup> This suggests

---

117. 35 U.S.C. § 122(b).

118. *Id.* at 64; see FTC REPORT, *supra* note 19, ch. 4 at 27.

119. NAS REPORT, *supra* note 19, at 64.

that at least one foreign counterpart application is filed in connection with a very high percentage of all U.S. patent applications and that the percentage is even higher among applications representing inventions thought to have substantial value.<sup>120</sup>

To summarize our two main points, Section 271(f) provides patent holders no protection in U.S. markets and does not offer meaningful prospects of patent protection in foreign markets, particularly at the inventive and investment stages of innovation when patent-based incentives must operate. In contrast, foreign patenting of inventions being patented in the United States promises real patent protection for those instances where prospects of protection in foreign markets might produce incremental incentives for innovation; it is now user-friendly, entails modest incremental costs that need not be incurred or committed to at an early stage, and is actually carried out in a large majority of cases, particularly where the inventions are thought to have significant value. In short, Section 271(f) contributes nothing toward the incentivizing benefits the U.S. patent laws are intended to provide.

### C. Risks and Burdens for U.S. Producers

The failure of Section 271(f) to protect U.S. patent holders in foreign markets does not translate into low risks for U.S. producers who sell components abroad. Indeed, Section 271(f) has become a substantial burden and disincentive for U.S. producers of technologically advanced products for international markets.

The seeming paradox of significant added risks without any significant protection results from Section 271(f)'s applicability only against U.S. producers. This leaves all actual and potential foreign producers free to compete with the patent holder in foreign markets and makes Section 271(f) ineffective as a protective measure for U.S. patent holders. It does not, however, reduce the monetary or injunctive burdens Section 271(f) can impose on companies that do produce components in the U.S. and export their products to other countries.

The impact of Section 271(f) on U.S. producers can be severe. In *Eolas*, for example, 2/3 of the huge \$565 million (including some interest)

---

120. The 89-90% figure for the fraction of U.S. patent applications for which foreign filings are not disclaimed would include both (i) cases where the foreign filing is made first and (ii) cases where the applicant makes the initial PCT filing but then, based on the international search report and any commercial developments during the subsequent thirty months, decides not to complete any filings in foreign countries, in addition to (iii) cases where the U.S. filing is made first with plans to file abroad which are subsequently carried out.

district court judgment represented damages based on foreign sales that were recoverable under the U.S. patent laws only because of Section 271(f).<sup>121</sup> In other words, the award under Section 271(f) dwarfed the traditional U.S.-based damages. In the recent *Union Carbide* case, the claims for foreign sales advanced under Section 271(f) did not have quite as dramatic a role. There, Section 271(f) damages merely doubled *Shell's* damages exposure (increasing it by about \$110 million) as well as extending the scope of the injunctive relief in the case. In cases of willful infringement, of course, treble damages can be awarded, as well as the typically large attorney fees and costs incurred in patent litigation.<sup>122</sup>

Several other aspects of patent enforcement—many of them not unique to the Section 271(f) context—further increase the risks that Section 271(f) poses for U.S. producers. These include:

- 1) The strengthening and proliferation of U.S. patents in the last twenty-five years;
- 2) The frequently uncertain scope and validity of patents;
- 3) Technological and industrial trends of increasing product complexity and international sourcing of “components”; and
- 4) The likely availability of onerous permanent injunctions against unsuccessful infringement defendants.

As explained below, these factors compound Section 271(f)'s heavy monetary consequences and create a truly substantial set of risks for U.S. producers of technologically advanced “components” for world markets. Because these risks are easy to avoid by moving production offshore, they create strong incentives that operate directly contrary to U.S. interests. The following paragraphs briefly review these factors.

### *1. Strengthening and Proliferation of U.S. Patents*

Completely apart from Section 271(f), U.S. patent rights have become dramatically stronger in the last twenty-five years. For decades prior to the 1980s, their legal and economic weakness was widely recognized,<sup>123</sup> causing Justice Jackson to disparage the trend he feared could soon mean “the

---

121. See *supra* text accompanying note 95.

122. See 35 U.S.C §§ 284-85; *infra* note 137 regarding the magnitude of attorney fees in patent infringement cases.

123. See sources cited *supra* note 19 and accompanying text; INNOVATION AND ITS DISCONTENTS, *supra* note 95, at 97, 99-110.

only patent that is valid is one which this Court has not been able to get its hands on.”<sup>124</sup>

Beginning around 1980, however, a combination of factors has vastly strengthened U.S. patents and patent rights.<sup>125</sup> The results have included much increased strength for patents as legal and economic instruments, much greater respect for patents on the part of business executives as well as legal advisors and academics, and a broad new recognition among corporate managers that large monetary liabilities and potentially disastrous injunctions are likely possibilities in cases of patent infringement. This across-the-board strengthening of patent rights has made patents a much greater, and much more widely recognized, source of business risks for companies producing technologically advanced products.<sup>126</sup>

At the same time as U.S. patents have gained strength legally, they have proliferated greatly in numbers, thereby further increasing the patent-

---

124. *Jungerson v. Ostby & Barton Co.*, 335 U.S. 560, 572 (1949) (Jackson, J., dissenting); *see, e.g., Ling-Temco-Vought, Inc. v. Kollsman Instrument Corp.*, 372 F.2d 263, 271 (2d Cir. 1967) (Kauffman, J., concurring in part and dissenting in part); *Reiner v. I. Leon Co.*, 285 F.2d 501, 503 (2d Cir. 1960).

125. Among other things, the U.S. Court of Appeals for the Federal Circuit, created in 1982 with exclusive appellate jurisdiction over essentially all U.S. patent cases, ended the inconsistency of patent rights among the circuits and the hostility many courts had shown to patents and patent enforcement. In numerous decisions since its creation, the Federal Circuit has established more supportive standards of validity and enforceability for patents, narrowed the defenses to claims of patent infringement, and generally increased the availability of effective remedies for holders of U.S. patents. *See supra* note 20; FTC REPORT, *supra* note 19, ch. 1 at 18-22, 30; NAS REPORT, *supra* note 19, at 31-37, 50. This broad and multi-faceted strengthening of U.S. patents by the Federal Circuit's rulings is one of the principal subjects of the recent and widely discussed book INNOVATION AND ITS DISCONTENTS, *supra* note 95, at 104-26. Separate federal legislation has further strengthened U.S. patents and the rights of their holders in other ways. *See, e.g.,* Patent Misuse Reform Act of 1988, 35 U.S.C. § 271 (amending Section 271(d) to cut back the misuse defense); Process Patent Amendments Act, Pub. L. No. 100-418, 102 Stat. 1563 (adding Section 271(g), thereby extending to holders of process patents the right to preclude importation of products made abroad by means of their U.S. patented process); Patent Law Amendments of 1984, Pub. L. No. 98-622, 98 Stat. 3383 (enacting not only Section 271(f) but also provisions to protect against certain types of obviousness-based invalidation of patents because of work of other members of the same research team); Drug Price Competition and Patent Term Restoration Act of 1984, Pub. L. No. 98-417, 98 Stat. 1585 (extending the terms of certain patents on inventions subject to regulatory review); Cooperative Research & Technology Enhancement (“CREATE”) Act of 2004, Pub. L. No. 108-453, 118 Stat. 3596 (removing additional types of cooperatively-developed prior art as an obstacle to the issuance of new patents).

126. FTC REPORT, *supra* note 19, ch. 1 at 14-22, 30, ch. 3 at 16, 36; INNOVATION AND ITS DISCONTENTS, *supra* note 95, at 104-26; NAS REPORT, *supra* note 19, at 31-37, 50.

based risks and uncertainties facing technology companies. Between 1981 and 2000, for example, the number of U.S. patents granted annually more than tripled.<sup>127</sup> There are now nearly two million unexpired U.S. patents.<sup>128</sup> This proliferation reflects several causes, in addition to rapid technological advances, including: (i) the widespread corporate adoption of patent accumulation and licensing as a method of revenue generation; (ii) a virtual “patent arms race” of “defensive patenting,” particularly in the electronics, software, computer, and telecommunications industries;<sup>129</sup> and (iii) the basic ease of obtaining U.S. patents.<sup>130</sup> In any event, many com-

---

127. See, e.g., FTC REPORT, *supra* note 19, ch. 5 at 24-25; see also NAS REPORT, *supra* note 19, at 16, 42-51. Indeed, according to the U.S. PTO’s *Annual Performance and Accountability Reports* published on its website, the number of utility patent applications filed annually is up by nearly a factor of four from 1978 to 2005.

128. Based on annual issuance statistics presented on the PTO’s website and the patent term of seventeen years from issuance, which applied until a few years ago, there should be more than two million issued and unexpired U.S. patents outstanding. This, however, is something of an overestimate for three reasons: (i) a relatively small number of the patents issued in the last seventeen years were, no doubt, issued subject to terminal disclaimers pursuant to 37 C.F.R. § 1.321(a) and (b) and therefore had shorter lives; (ii) some of these patents have expired for failure of the holders to pay the applicable maintenance fees (maintenance fees were first imposed pursuant to the Act of Dec. 12, 1980, Pub. L. 96-517, 94 Stat. 3015, and were applicable to applications filed on or after that date); and (iii) a relatively small number of those patents have been held invalid in infringement or declaratory judgment suits or in re-examination proceedings. The second of these factors is no doubt the most important in numerical terms, but such failures to pay maintenance fees are presumably rare as to economically significant patents, and a procedure exists whereby many such prematurely expired patents can be revived. See 37 C.F.R. § 1.378(c). Taking into account failures to pay maintenance fees and other factors, a recent report indicated that a few years ago something in excess of 1.63 million issued U.S. patents then remained in force. TRILATERAL STATISTICAL REPORT, *supra* note 110.

129. FTC REPORT, *supra* note 19, ch. 2 at 26-28, 30-31, 34-36, ch. 3 at 33-40, 52-54; NAS REPORT, *supra* note 19, at 31. “Defensive patenting” is the deliberate accumulation of as many patents as possible out of fear that, without a substantial patent portfolio with which to bargain and potentially retaliate, a company can easily be barred from participation in important product areas by competitors exercising their patent rights. See FTC REPORT, *supra* note 19, ch. 2 at 34, 36 (referring to “defensive patenting” which was spurred by the practices of Japanese companies in: (i) filing large numbers of patent applications on small innovations; and (ii) “agglomerating huge [patent] portfolios which they swapped [i.e., cross licensed] with [other large Japanese companies] but which they were unwilling to trade with [i.e., license to] outside players”).

130. See, e.g., FTC REPORT, *supra* note 19, ch. 3 at 34-36 (referring to the ease of obtaining U.S. patents); *id.* ch. 1 at 8-9 (referring to the “plethora of presumptions and procedures [that] tip the scales in favor of the ultimate issuance of a patent, once an application is filed” and to the PTO’s obligation to issue a patent unless the PTO can demonstrate that the claimed invention is not patentable—which effectively creates a presumption that each requested patent should be issued); see also NAS REPORT, *supra* note 19, at 54 n.34, quoting two patent scholars:

panies that formerly applied for few if any patents now are making concerted efforts to apply for, and obtain, as many patents as possible.<sup>131</sup>

## 2. *Uncertainty as to Patent Coverage and Validity*

Uncertainty as to patent scope, validity, and enforceability is an inherent—and frequently deplored—aspect of patent rights.<sup>132</sup> This uncertainty is significant, for present purposes, because it often prevents technology companies from determining whether patents cover products they are producing or want to produce and, if the patents do, whether they are valid and enforceable. Accordingly, even as to known patents, the exercise of care cannot insulate U.S. producers from substantial risks of infringement, including the potentially large increases in damages exposure and more burdensome injunctions under Section 271(f). The consequences of uncertainty, of course, extend far beyond Section 271(f). However, the burdens in relation to potential benefits and the effects of recent technological and

---

One of the oddest things about the U.S. patent system is that it is impossible for the [PTO] to ever finally reject a patent application. While patent examiners can refuse to allow [a requested patent], and can even issue what are misleadingly called “Final Rejections,” the patent applicant always gets another chance to persuade the patent examiner to change her mind.

131. See sources cited *supra* notes 129-130.

132. See, for example, the comments of various participants in the February 18, 2005 “town meeting” hearing in San Jose, California, the first of four such hearings on patent reform co-sponsored by the National Academies’ Board on Science, Technology, and Economic Policy, the American Intellectual Property Law Association, and the Federal Trade Commission. Among the comments were the following:

We have a patent system that is fundamentally is [sic] not predictable enough. You can’t just pick up a patent, see what’s claimed, and predict what’s going to happen when you get into court. We know—painfully—patent law isn’t simple enough,” Armitage said.

Armitage also criticized that it takes years to determine something is patentable, “long after the technology is obsolete. It just isn’t relevant if the patent system takes longer than the life cycle of the product.”

Joseph J. Rolla, deputy commissioner for patent examination policy at the PTO, said the agency has more than 100 pending re-examination proceedings that are over four years old. More than 400 pending proceedings are two years or older, and almost 600 proceedings have been pending for a year.

Joyce. E. Cutler, *Patents/Reform: Patent Reform Is Hot Topic for Innovators, Litigators*, Pat. Trademark & Copyright J. (BNA) (Feb. 25, 2005). Substantial segments of the recent FTC Report address the uncertainty of U.S. patent rights and its negative consequences for innovation and economic activity. See, e.g., FTC REPORT, *supra* note 19, ch. 3 at 37-41, 53-54.

trade developments have particular relevance in this context. Accordingly, a brief review of some of the contributors to the uncertainty is useful here.

a) Complexity of Patent Claims and Defenses

Patent claims and defenses, and patent litigation generally, are notoriously complex and unpredictable. Statistics show that the reversal rate for district court claim construction rulings—and on trial court judgments in patent cases generally—is in the range of 40%.<sup>133</sup> In one recent case,<sup>134</sup> two successive panels of the Federal Circuit announced opposite constructions of the same relevant claim terms in the same patent in successive appeals even though claim interpretation is purely a matter of law.<sup>135</sup> The history of the *Eolas* and *Blackberry* cases, discussed above, provides good examples of—and sad commentaries on—this complexity and unpredictability.<sup>136</sup> The very large average legal fees in patent lawsuits further testify to the high levels of complexity in such proceedings.<sup>137</sup>

---

133. See, e.g., Kimberly A. Moore, *Markman Eight Years Later: Is Claim Construction More Predictable?*, 9 LEWIS & CLARK L. REV. 231, 233 (2005) (discussing a survey that found a 34.5% reversal rate in the Federal Circuit on appealed rulings on claim construction alone, which is three times greater than the typical reversal rates in other circuits). Other surveys have found somewhat higher rates of reversal in the Federal Circuit, and it is not unusual for the Federal Circuit to uphold a district court's claim construction rulings but to reverse on, say, validity or enforceability. In fact, this was the result in the *Eolas* appeal. 399 F.3d 1325 (Fed. Cir. 2005).

134. Compare *CVI/BETA Ventures, Inc. v. Tura LP*, 112 F.3d 1146, 1157-58, 1160 n.7 (Fed. Cir. 1997) (construing the same claim terms in the appeal in a manner opposite to that used in the determination of liability that had been affirmed in an earlier appeal), with *CVI/BETA Ventures, Inc. v. Tura LP*, 120 F.3d 1260 (Fed. Cir. 1997). A similar result happened again in *Nystrom v. Trex Co.*, 424 F.3d 1136 (Fed. Cir. 2005), although here it resulted in part from the intervening en banc ruling by the Federal Circuit in *Phillips v. AWH Corp.*, 415 F.3d 1303 (Fed. Cir. 2005), which modestly revised the rules governing patent claim construction.

135. *Markman v. Westview Instruments, Inc.*, 52 F.3d 967 (Fed. Cir. 1995) (en banc), *aff'd* 517 U.S. 370 (1996); *Cybor Corp. v. FAS Techs., Inc.*, 138 F.3d 1448 (Fed. Cir. 1998) (en banc).

136. After years of disputes and litigation, the district court in *Eolas* rejected Microsoft's defenses of anticipation, obviousness, and inequitable conduct, made rulings on claim interpretation, and, based on a jury verdict, entered a very large judgment for *Eolas*. On appeal, the Federal Circuit reversed the district court's rulings on the anticipation, obviousness, and inequitable conduct defenses, remanded the case for trial of these issues, and held that Microsoft had failed to preserve a possible error in claim construction for appeal. See Michael Warnecke, *Web Browser Patent Challenged by Microsoft Is Reaffirmed as Valid by Patent Office*, Pat. Trademark & Copyright J. (BNA) (Sept. 30, 2005). Moreover, in October 2003, the PTO instituted a reexamination of the *Eolas* patent and then issued two successive office actions in February and August 2004, each rejecting all ten of the challenged claims as having been obvious in view of the prior art at the

### b) Unknown Patents

The uncertainties as to known patents, of course, pale in comparison with those in two other categories of patents: (i) those issued patents and published-but-unissued patent applications that may have been missed by any recent search; and (ii) those additional pending patent applications that

---

time of the application. *Id.* Some observers considered these actions indicative of likely eventual invalidation of the patent. In September 2005, however, after further filings by the parties, the PTO issued a "Notice of Intent to Issue a Reexamination Certificate" for the patent, reaffirming the view that the patent is valid. *Id.*

A similar story of uncertainty can be seen in the closely watched BlackBerry litigation. After substantial litigation in that case, the district court entered summary judgment of infringement and a jury found the infringement willful in November 2002. In August 2003, after further proceedings, the district court entered a final judgment awarding damages and an injunction against further manufacture, use, importation, or sale of all BlackBerry systems, software, and handheld devices. In December 2004, however, the Federal Circuit vacated the injunction, reversed some of the district court's claim construction rulings, and set aside the jury verdict. In August 2005, after additional motions, the Federal Circuit withdrew this fifty-nine page decision and issued a new seventy-four page decision, reversing the district court's rulings of infringement of the method claims but affirming the infringement rulings as to the apparatus claims. Meanwhile, in January 2003, the PTO had ordered reexamination of the patent claims. According to published reports, the PTO subsequently issued initial rulings rejecting all of the claims in the subject patents, including the sixteen claims at issue in the litigation, and entered a final action finding one of the patents invalid. In October 2005, both the Federal Circuit and the U.S. Supreme Court denied further stays of injunctive relief pending final resolution of the case. With a broad new injunction against operation of the BlackBerry system expected, its owner-operator settled the infringement claims with a payment of more than \$600 million, although the PTO reexamination proceedings continued and could well ultimately invalidate all the claims in suit.

These examples of continuing uncertainty as to patent coverage and validity are unfortunate, but not unusual, given the complexity of many issues affecting the existence, extent, and enforcement of patent rights.

137. Surveys of attorney fees in patent cases that go to trial have reported that the average legal fees in such cases vary greatly depending on the amount "at risk." One survey found that the legal fees on each side of small patent infringement cases (< \$1 million "at risk") average about \$500,000, that the average per side in medium-sized cases (\$1 million - \$25 million "at risk") is about \$2.0 million, that the average per side in larger cases (> \$25 million "at risk") is \$4.0 million, and that the costs on each side can exceed \$20 million in the largest cases. NAS REPORT, *supra* note 19, at 68-70. Similar ranges, with substantial variations by region of the country, have been reported in other surveys. See, e.g., KIMBERLY MOORE ET AL., PATENT LITIGATION & STRATEGY 1 (Thompson-West, 2d Ed. 2003). David Simon, the Chief Patent Counsel at Intel, has been quoted as saying that, for Intel, "legal fees are a huge hit. 'Any one patent lawsuit will cost us in the area of \$20 million-plus, given the complexity of our technology.'" Victoria Slind-Flor, *Simon Says - Intel's Chief Patent Counsel, David Simon, Has Some Tough Rules for Outside Counsel*, IP LAW & BUS., December 10, 2004, available at <http://www.law.com/jsp/article.jsp?id=1102543080110>.

the PTO has not yet issued nor published. Patent searches are never certain to identify all the patents that may cover a given product or invention. The discovery, well after one or more searches, of highly relevant patents is not unusual. Likewise, the second category—pending but unissued and unpublished patents—still creates substantial risks despite changes to the U.S. patent system that have resulted in most pending U.S. patent applications being published eighteen months after their effective filing dates.<sup>138</sup>

c) Poor Patent “Quality”

Accompanying the proliferation of U.S. patents described above and explaining some of the uncertainty as to patent-based rights and obligations is the widespread perception that many patents being issued by the PTO should never have been issued because they claim unpatentable “inventions.” Clear, quantitative data on patent quality are hard to obtain, but anecdotal information and the testimony of most of the patent experts that address this question indicate U.S. patents are issued too readily and that patent quality has become a very serious problem affecting substantial percentages of issued patents.<sup>139</sup> As reasons for this quality problem, they cite: (i) the ballooning numbers and complexity of applications being filed; (ii) the PTO’s and courts’ generally weak application of the obviousness grounds for denying patents; (iii) the PTO’s inadequate prior art database in some areas where many patent applications are being filed; and (iv) the distressingly short time the PTO examiners have to review

---

138. Pursuant to the American Inventors Protection Act of 1999, Pub. L. No. 106-113, 113 Stat. 1501 (codified at 35 U.S.C. § 122), the PTO recently began publishing U.S. patent applications eighteen months after filing *unless* the applicants disclaimed any intention to file any counterpart applications in any foreign country. Despite this practice, pending, unpublished U.S. patent applications remain a substantial source of uncertainty and risk for several reasons: (i) the U.S. patent system, pursuant to 35 U.S.C. § 102(b), allows a “grace period” of one year for filing the application after placing a product “on sale” or “in public use” or after describing it in a written publication, thereby allowing later applications than most non-U.S. patent systems; (ii) PTO processing times for patent applications have ballooned to well over three years in most art fields and much more than that in some; and (iii) publication can be avoided for the entire period of pendency (i.e., until issuance of the patent) by submitting a disclaimer of intent to file corresponding applications in foreign countries—which is the main situation where Section 271(f) might claim any justification in any event. *See* 35 U.S.C. § 122.

139. FTC REPORT, *supra* note 19, Exec. Summary at 5-10, ch. 1 at 34-35, ch. 3 at 20-21, 37-40, 53, etc.; INNOVATION AND ITS DISCONTENTS, *supra* note 95 at 142-50.

each application, find the relevant prior art, and determine patentability as well as adequacy of disclosure, etc.<sup>140</sup>

3. *Technological and Industrial Trends—International Sourcing of “Components” and Product Complexity*

Apart from legal matters, trends in technology and production practices have contributed greatly to the risks posed by Section 271(f). These include: (i) increased world-wide sourcing of product components; and (ii) the increasing complexity of many new products, which causes them to include large numbers of relatively newly developed “components.”

Business literature for decades has discussed the ever increasing levels of international procurement of components for all sorts of products. The case law suggests the same thing, although one cannot deem Section 271(f) cases as representative because some international aspects of the products in question are essential for the section to have any relevance at all. Nevertheless, cases like *Pellegrini* illustrate what are now common patterns of international component design, sourcing, and assembly.<sup>141</sup> There is simply no doubt that procurement of components from a variety of international sources is now commonplace, particularly for more complex and technologically advanced products. These patterns of international sourcing of components multiply the possibilities for application of Section 271(f).

Increasing product complexity has further increased the impact and burdens of Section 271(f). Particularly in the fields of electronics, semiconductors, computers, telecommunications, software, and the like, single new products often incorporate very large numbers of what one can reasonably term “components.” Indeed, evidence developed in the NAS Report<sup>142</sup> showed that, in semiconductors, computers, telecommunications, and related fields, “it is common for there to be *hundreds* of patentable elements in one [new] product.”<sup>143</sup> The Federal Circuit’s reasoning in *Eolas* and other recent cases implies that an even larger number of individual “elements” of each of those patentable inventions would qualify as Section

---

140. See generally NAS REPORT, *supra* note 19, at 47-54, 61-62, 80, 87-90, 104; FTC REPORT, *supra* note 19, Exec. Summary at 5-10, ch. 1 at 34-35, ch. 3 at 20-21, 36-37, 53-54, ch. 4 at 8-15; INNOVATION AND ITS DISCONTENTS, *supra* note 95, at 145-50.

141. See *supra* text accompanying note 90 regarding the international sourcing and assembly of the products involved in that case.

142. See *supra* note 101 regarding the origins of the NAS Report and the information developed there.

143. NAS REPORT, *supra* note 19, at 37 (emphasis added) (citing studies done in 2000 and 2001).

271(f) “components.”<sup>144</sup> In any event, clearly there are very large numbers of such “components” in many new products. This too makes the application of Section 271(f) to such products much more likely than it would otherwise be and also more complex and difficult to assess.

#### 4. *Likely Injunctions*

In many contexts, the possibility of a permanent injunction against further making, using, selling, etc. a product or “component” because of patent infringement is considered more threatening than even a substantial damages award.<sup>145</sup> The potential for entry of injunctions, in addition to awards of damages, is present under Section 271(f) as it is under the other types of infringement of U.S. patents.<sup>146</sup> Moreover, despite the Supreme Court’s recent *eBay, Inc. v. MercExchange* decision, entry of a broad permanent injunction against the defendant’s continued production and sale activities is quite likely in Section 271(f) cases where infringement is found and the defendant fails to prove, by clear and convincing evidence, that the patent is invalid or unenforceable; this is particularly true if the patent holder or an exclusive licensee is competing in the market for the claimed product.<sup>147</sup> The potential harshness of this result, together with the

---

144. The recent Federal Circuit cases indicate that each patentable invention involves multiple “components,” for purposes of Section 271(f). *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325, 1339-40 (Fed. Cir. 2005) (holding that each structural element of a patented “machine,” “manufacture,” or “composition of matter” would qualify as a “component,” as would each “step” or “act” of a patented method); *AT&T Corp. v. Microsoft Corp.*, 414 F.3d 1366, 1368-69 (Fed. Cir. 2005) (agreeing). *See also supra* notes 46-51 and accompanying text.

145. Injunctive awards can, among other things, shut-down the defendant’s production of the relevant product(s) or components, decimate the defendant’s market participation, block performance of its supply commitments, render useless investments made and/or facilities built for production of the relevant item(s), and/or injure its participation in markets for products not covered by the patent(s) in question but related to the claimed product(s) or processes for sales or distribution purposes.

146. 35 U.S.C. § 283; *see, e.g.*, *Lubrizol Corp. v. Exxon Corp.*, 696 F. Supp. 302 (N.D. Ohio 1988) (enjoining further shipments of the key chemical ingredient abroad for use in the claimed chemical combinations).

147. *eBay, Inc. v. MercExchange, LLC*, 126 S. Ct. 1837 (2006). In this case, the district court found *MercExchange*’s patent infringed and not proved invalid but nevertheless declined to issue a permanent injunction against *eBay*, finding that *MercExchange*’s: (i) willingness to license; (ii) failure to practice the patents itself; and (iii) comments to the media regarding its intent to enforce and license its patent rights all weighed against any presumption that it would suffer irreparable harm without an injunction. *MercExchange, L.L.C. v. eBay, Inc.*, 271 F. Supp. 2d 789 (E.D. Va. 2002). The Federal Circuit reversed this denial of a permanent injunction stating, *inter alia*: “Because the ‘right to exclude recognized in a patent is the essence of the concept of property,’ the general rule is that a permanent injunction will issue once infringement and validity have been ad-

strong bargaining position it can give patent holders claiming infringement, have been widely criticized.

Obviously, the burdens of an injunctive order depend heavily on the facts of the particular case, with some injunctions presumably imposing only modest burdens as a practical matter. Likewise, *eBay* makes it less certain that a broad permanent injunction will be imposed whenever Section 271(f) infringement is found in the U.S. production of unpatented components for use in foreign-assembled products. The likelihood of onerous injunctive relief nevertheless remains important. When considering the risks of Section 271(f) claims, companies that can locate their productive activities either in the United States or abroad must consider not only (i) the increased possibility and larger size of infringement damages awards that producing domestically would entail but also (ii) the possibility that investments in productive facilities and people here might be wasted, even for foreign sales, and the company's ability to supply foreign markets jeopardized, because of an injunction under Section 271(f), whereas little or no such risk would exist if the investments were made in productive facilities located abroad.<sup>148</sup> Indeed, injunctions entered under

---

judged." 401 F.3d 1323, 1338 (Fed. Cir. 2005). The Supreme Court reversed, stating that both courts had erred: the district court by concluding that the factors it mentioned were sufficient to deny a permanent injunction and the Federal Circuit by applying a nearly absolute presumption in favor of such broad, permanent injunctive relief. *eBay*, 126 S. Ct. at 1840-41. The Court mandated detailed facts-and-circumstances application of the traditional four-factor test for entry of an injunction but declined to spell out how the factors should be applied either on the facts of that case or in patent cases more generally. *Id.* The two concurring opinions, representing a total of seven of the justices, indicated that permanent injunctions can still be expected in a large proportion of cases where the patent is found valid and infringed and the patent holder competes in supplying the claimed product. *Id.* at 1841-43.

148. In order for Section 271(f) to add anything to the otherwise applicable rights and obligations: (a) the components in question must not be the subject of U.S. or the relevant foreign patents and (b) no effective foreign patent enforcement can be available as to the "infringing" combinations in the locations where they are assembled, sold, or used. If these points are true, production of the components in question in the country where the combinations are assembled or in any other country will necessarily face little or no meaningful exposure to patent liability. For example, assume as alleged that Pellegrini's patent would be found valid and enforceable and that the combinations in question, when made abroad, fell within its claims. Further assume that Pellegrini had sued not only Analog Devices, which produced the subject circuit chip components in Ireland and had them produced in Taiwan, but also a separate company that produced similar chips in the U.S. for shipment, combination, and sale solely abroad. The result would have been that Analog Devices would be free to continue supplying the chips for foreign markets, but that the company that produced the same chips in the U.S. for combination and use abroad would have been enjoined from continuing that activity.

Section 271(f) are uniquely far-reaching in one sense: they prohibit production in the United States of unpatented components for inclusion in “infringing” combinations anywhere in the world, including in countries that have no patent laws, countries where the combination product in question is not patentable subject matter, countries where a patent on that product exists but is invalid or unenforceable, and countries where no relevant patent application was ever filed.

In summary, Section 271(f) is anything but benign for U.S. producers. When it comes into play, it can multiply the monetary and injunctive exposure of U.S. producers to claims of patent infringement. Its terms necessarily put U.S. producers at a disadvantage relative to their foreign competitors. The recent judicial extensions of Section 271(f) have greatly expanded its reach. At the same time, commercial and technological developments have increased the likelihood that it can be invoked as to newly developed products. In addition, the increased number and enhanced strength of U.S. patents have increased the risks inherent in infringement litigation across the board, both in cases where the risk multiplier of Section 271(f) applies and in cases where it does not. Overall, the incremental risks and burdens created by Section 271(f) are substantial for producers of technologically oriented components for world markets.

Particularly serious are the resulting incentives. As *Pellegrini* makes clear, current and prospective U.S. producers can avoid the risks and burdens of Section 271(f) by the simple stratagem of locating their productive activities outside the United States. In the case of intangible “components” such as software and information, this might require the relocation of activities not normally referred to as “manufacturing” or “production.”<sup>149</sup> Off-shoring of productive activities, of course, is sometimes attractive simply because of the reduction in production costs it often allows. Given the recent legal, technological, and commercial developments extending the reach of Section 271(f), however, the avoidance of its risks must in-

---

149. Part V of this Article suggests some types of information other than software might soon be deemed Section 271(f) “components” under the Federal Circuit’s expansionist approach to that provision. Software and information may be made in final or near final form by writers, designers, and researchers in activities not normally termed “manufacturing.” Conceivably a patent plaintiff trying to apply Section 271(f) to an information “component” might argue that U.S.-developed information is “supplied” from the U.S. even if it finds its way from a U.S. researcher to an off-shore facility where its tangible form of expression is altered and is subsequently distributed in that new form to other foreign locations and combined there with other components such as, say, medical diagnostic machines. If such an argument should prevail, it might be necessary to move the writing or research facility out of the U.S. in order to avoid the operation of Section 271(f) as to the use of the information with other components in world markets.

creasingly influence advice given to corporate managers as to where new production facilities should be located. In fields where patent worries are particularly strong, such as electronics, telecommunications, computer systems, etc., relocation of activities currently conducted in the United States may be part of such advice.

#### **D. Other Possible Arguments in Favor of Section 271(f)**

While Section 271(f) creates substantial problems and cannot incentivize innovation, its defenders might cite two separate considerations as possible reasons for its retention: (i) the knowledge/intent elements of infringement under Section 271(f) and the possible “unfairness” of such infringement in view of those elements; and (ii) efficiency considerations in the enforcement of U.S. patent rights. On examination, neither consideration meaningfully supports retention of Section 271(f).

##### *1. Knowledge Elements and Unfairness*

Defenders of Section 271(f) might say that the knowledge or intent required on the part of a U.S. producer in order to infringe under that section mitigates the burdens it imposes on U.S. producers. Similarly, given the knowledge/intent elements, they might argue that infringement under Section 271(f) requires “unfair” conduct, which is properly subject to sanctions.

Both clauses of Section 271(f) (together with the case law on inducement of infringement and contributory infringement) limit that section’s application to circumstances where the U.S. producer of components: (i) knows or has reason to know that the components it produces and sells are likely to be combined abroad into a combination product; and (ii) has notice of the relevant U.S. patent and the ability to surmise that the foreign-made combination may fall within one or more of the patent’s claims.<sup>150</sup> The extreme case of such knowledge, of course, was *Deepsouth*, where the U.S. producer made all the arrangements, knew of the patent, its coverage and validity, and had already been enjoined from the its proposed acts.<sup>151</sup> The Supreme Court dissenters in that case characterized the U.S. pro-

---

150. The wording of those clauses is set forth *supra* note 23. A number of cases discuss the elements of inducing infringement and contributory infringement and the “lack of clarity” as to the type or degree of intent needed for inducement. *See* *Insituform Techs., Inc. v. Cat Contracting, Inc.*, 385 F.3d 1360, 1378 (Fed. Cir. 2004); *Hewlett-Packard Co. v. Bausch & Lomb, Inc.*, 909 F.2d 1464, 1469 (Fed. Cir. 1990); *cf.* *Golden Blount, Inc. v. Peterson Co.*, 438 F.3d 1354, 1364 (Fed. Cir. 2006). In any event, intent is often inferred from circumstantial evidence. *Water Techs. Corp. v. Calco, Ltd.*, 850 F.2d 660, 668-69 (Fed. Cir. 1988).

151. *See supra* text accompanying notes 12-13.

ducer's conduct as unfair and, indeed, "iniquitous"; this perception contributed greatly to their view that it should be condemned.<sup>152</sup> Most Section 271(f) cases involve nothing like those levels of knowledge, intent, and overall direction on the part of the U.S. producer. In any event, as several considerations show, the knowledge/intent elements in Section 271(f) do not substantially mitigate its burdens and risks, do not justify it on fairness grounds, and do not render its perverse incentives insubstantial.

For one thing, as we have seen,<sup>153</sup> considerable uncertainty clouds the scope and validity of issued patents. This is significant here because a belief that a particular patent would not cover the foreign combination or that the patent was invalid or unenforceable might well not protect a U.S. producer against a finding of infringement of that patent under Section 271(f).<sup>154</sup> Making matters worse for the U.S. producer is the large number of patents that may cover new innovative products, often dozens of existing patents for each individual new product in fields such as semiconductors, software and computer systems, telecommunications, etc.<sup>155</sup> With significant uncertainties as to such large numbers of patents, it is unrealistic to think that a U.S. producer of technologically advanced components

---

152. See *supra* note 14 and accompanying text.

153. See *supra* text accompanying notes 138-140.

154. See, e.g., *Eolas Techs., Inc. v. Microsoft Corp.*, 399 F.3d 1325 (Fed. Cir. 2005) (remanding the case to try the issues of anticipation, obviousness, and inequitable conduct). In the pending reexamination proceeding, the PTO examiner has indicated his preliminary belief that the *Eolas* patent is invalid. The court opinions do not suggest that Microsoft would have any defense to the huge incremental liability under Section 271(f) if those defenses are ultimately rejected even though it could prove that it had a good faith belief that one or more of the defenses was applicable—as the Federal Circuit has said may be the case—or that the patent would not be read as covering the foreign-made combinations. See *Eolas*, 399 F.3d 1325. Such matters presumably would prevent findings of willful infringement and awards of enhanced damages under 35 U.S.C. § 284, but the district court's third-of-a-billion-dollar incremental award under Section 271(f) did not include any such enhancement. See also *Golden Blount*, 438 F.3d at 1364 (affirming finding of actionable inducement of infringement where accused party had received notification from patentee alleging infringement of the asserted patent, claimed a good faith belief of non-infringement of any valid claims, but had not insisted on competent written opinions of counsel affirming non-infringement or invalidity). The *Golden Blount* court held that "the only intent required [on the part of the defendant] is the intent to cause the acts that constitute infringement." *Id.* (quoting *Hewlett-Packard*, 909 F.2d at 1469).

155. See NAS REPORT, *supra* note 19, at 37 (noting that, in semiconductors and other complex product technologies, "it is common for there to be hundreds of patentable elements in one product," and consequently, absent cross license arrangements and other mitigating efforts, it would usually be impossible to develop or commercialize new products without infringing multiple patents of others).

can generally avoid infringement under Section 271(f) by exercising reasonable care.<sup>156</sup>

Yet another difficulty is a matter of timing. When a company must make investment decisions for the location of productive facilities, it or its advisors cannot possibly know what pending U.S. patents will be published or will issue during construction and start-up of those facilities or their operation. Thus, even if a U.S. producer could assess all issued and pending patents and form reliable judgments that none of them validly cover the combinations in which the company's components would be used, exposure under Section 271(f) would still be quite possible by the time the facilities enter operation and for some time thereafter.

Further, any fairness-based argument in favor of Section 271(f) is substantially undercut by the double failure (or inability) of the U.S. patent holder to obtain direct patent protection either: (i) in the form of U.S. patents on the U.S.-supplied components in question; or (ii) in the form of relevant foreign patents for the combination product. If the U.S. patent holder could have obtained either one of those patents but did not do so, his/her claim of unfairness would ring hollow in view of the relative ease of obtaining such protection. On the other hand, if the components and combination were unpatentable here and in the relevant foreign country, respectively, barring U.S. component producers from competing in the foreign market is difficult to justify as "fair."

Indeed, the most apparent unfairness relevant to Section 271(f) seems to be in the operation of that section itself—particularly given current patterns of U.S. patent awards and international competition. The operation of that provision only against U.S. producers, in itself, seems unfair. It is worse now that the granting of patents to non-U.S. inventors and owners differs fundamentally from what was assumed by the Supreme Court in *Deepsouth* and by Congress when it enacted Section 271(f).<sup>157</sup> Until the

---

156. Further, an effort by the U.S. producer to reduce its Section 271(f) exposure would require substantial conservatism on its part in assessing the scope and enforceability of the possibly applicable issued patents and published applications. This would expand the areas where U.S. producers are barred from innovating and competing, thereby chilling production and competition by U.S. producers.

157. Both the majority and the dissent in *Deepsouth Packing Co. v. Laitram Corp.* assumed, consistent with the facts there, that the question addressed concerned U.S. patents held by U.S. individuals or entities. The majority, for example, asserted what they considered a traditional and desirable ability of U.S. producers: "[W]e note that what is at stake here is the right of American companies to compete with an American patent holder in foreign markets. Our patent system makes no claim to extraterritorial effect . . ." 406 U.S. 518, 531 (1972). The dissenters assumed the same thing: "[The majority's result allows] an infringer to set up shop [in the U.S.] next door to a patent-protected inventor

1980s, the large majority of U.S. patents were issued to U.S. inventors and owned by them or by other U.S. entities. Currently, however, half of U.S. patents are being issued to foreign inventors and owned by foreign entities,<sup>158</sup> with Japanese companies comprising a majority of the ten top grantees of U.S. patents.<sup>159</sup> Accordingly, Section 271(f) may soon be invoked by, for example, a Japanese entity as the owner of the U.S. patent for a combination product or for a method/process. The target of the Section 271(f) claim, of course, would have to be one or more U.S. producers of components or of items that facilitate use of the method. Damages could be claimed under the U.S. patent laws for sales of the relevant combination products or use of the method in any country in the world.<sup>160</sup> Given the absence of a provision like Section 271(f) in Japanese patent law, the U.S. patent laws ironically would protect a Japanese company from competition by U.S. producers in markets around the world even though other Japanese producers, indeed all non-U.S. producers, would be under no such restrictions. There can be no justification for such a result.

## 2. *Efficiency in Patent Enforcement*

Another possible argument in favor of Section 271(f) is that it can promote efficiency in the collection of damages in infringement suits involving product sales in multiple countries. The holder of a U.S. patent can sometimes combine claims for domestic infringement with claims under Section 271(f) based on sales in foreign countries. A little consideration, however, shows that this provides no justification for Section 271(f).

Patent enforcement is notoriously expensive and problematic. For a holder of patents in multiple countries, enforcement against infringers in each of those countries has sometimes required separate actions in the respective jurisdictions.<sup>161</sup> In actuality, of course, one or a few infringers

---

whose product enjoys a substantial foreign market and deprive him of this valuable business." *Id.* at 534 (Blackmun, J., dissenting) (quoting *Laitram Corp. v. DeepSouth Packing Co.*, 443 F.2d 936, 939 (5th Cir. 1971)). The limited legislative history of Section 271(f) shows the same assumption. *See supra* notes 24-32 and accompanying text.

158. *See supra* note 17.

159. *See, e.g.*, U.S. Patent & Trademark Office, Calendar Year 2005 - Preliminary List of Top Patenting Organizations, <http://www.uspto.gov/go/taf/top05cos.htm>

160. For Section 271(f) to be relevant, the U.S. patent would have to cover only the combination and not the component(s) or the item that facilitates use of the method. That being the case, any Japanese patent would presumably not cover the component(s) or item either.

161. *See* Briefs of Appellant, Intellectual Property Owners Ass'n as Amicus Curiae Supporting Appellant, United States as Amicus Curiae Supporting Appellant, Federal Circuit Bar Ass'n as Amicus Curiae Supporting Appellant, American Intellectual Property Law Ass'n as Amicus Curiae Supporting Appellant, *Voda v. Cordis Corp.*, Misc. Dkt

may be by far the most significant. Particularly in such a case, one or two suits may either: (i) reap the bulk of potential recoveries; or (ii) shape the resolution and settlement of claims for sales activity in many or all of the relevant jurisdictions. That will not necessarily be the case, though, and the possible need for infringement suits in more than one jurisdiction—with correspondingly increased costs—cannot be dismissed.

As the *Eolas* case demonstrates, Section 271(f) might seem, on the surface, to substitute for suits in multiple jurisdictions, allowing enforcement of patent rights in multiple countries in a single, more efficient proceeding.<sup>162</sup> The basic problems, however, are that: (i) Section 271(f) does not apply the respective patent laws of foreign jurisdictions as to sales in those respective countries but applies an abbreviated form of the U.S. patent laws as to those foreign sales; and (ii) Section 271(f) is available only against U.S. producers. Thus, the standards applied are not those that are applicable as against non-U.S. infringers—who, in fact, would be under no restrictions at all unless an enforceable patent exists in the relevant jurisdiction, in which case Section 271(f) would be largely superfluous—and discrimination against U.S. producers is inherent in such proceedings. The only way to avoid discrimination and accurately reflect the actual underlying patent rights would be to apply the respective national patent laws as to acts in the respective countries.

A few patent holders have tried to sue infringers in a single U.S. court action under: (i) their U.S. patents and the U.S. patent laws for infringing acts committed here; and (ii) their foreign patents and foreign patent laws for acts in the respective foreign countries. The U.S. federal district courts probably have jurisdiction to adjudicate the foreign patent law claims against U.S. producers, in addition to any claims against them under U.S. law, either under the supplemental jurisdiction statute, 28 U.S.C. § 1367(a), or under the general diversity statute, 28 U.S.C. § 1332. In the relatively few cases where such combined claims have been asserted, however, the U.S. courts have generally used one of several rationales to decline to hear the foreign law claims, even though they involved the same

---

No. 785, 122 F.App. 515, 2005 U.S. App. LEXIS 4394 (Fed. Cir. Feb. 22, 2005) (granting petition for permission to appeal), *appeal dismissed*, 125 F.App. 287, 2005 U.S. App. LEXIS 5594 (Fed. Cir. Mar. 2, 2005), *available at* <http://patentlaw.typepad.com/patent/2005/06>.

162. In *Eolas Techs., Inc. v. Microsoft Corp.*, for example, Eolas obtained a judgment from the district court against Microsoft awarding damages for both: (i) domestic infringement; and (ii) creation of the allegedly infringing combination products abroad, effectively covering sales in many different countries as well as sales in the United States. 399 F.3d 1325 (Fed. Cir. 2005).

or similar products and the same or very similar patent claims to those being adjudicated under the U.S. patent laws.<sup>163</sup>

Practical problems may make it infeasible to adjudicate patent infringement in multiple countries under multiple laws in a single proceeding. These include: (i) the complexity of the patent laws; (ii) the differences among the patent laws of different countries; (iii) the need to apply instruments, documents, and laws in different languages; and (iv) the procedural complications and difficulty of gathering information from multiple countries, etc.<sup>164</sup> Such complications may be an argument for the harmonization of international patent laws, which is a common topic of discussion in patent law circles. Judging by negotiations to date, however, internationalizing the U.S. patent laws is not an acceptable answer to this problem as far as foreign countries are concerned. Use of Section 271(f) to accomplish a limited internationalization of U.S. law—but only against U.S. producers and only as to circumstances where U.S.-produced “components” are assembled abroad—seems logically unsupportable and, in any event, would injure U.S. interests and complicate patent rights substantially.

#### **E. Bottom Line Assessment**

The matters addressed above seal our assessment of Section 271(f). That provision fails to stimulate innovation and therefore fails to advance the objectives of the U.S. patent laws. It constitutes a threat and burden for U.S. producers of technologically advanced products that non-U.S. producers do not face. Since companies—U.S. and foreign alike—can avoid the burdens and risks of Section 271(f) simply by locating their productive facilities outside the United States or contracting with foreign producers, that provision creates substantial perverse incentives *against* investment, job creation, employee training, and production of innovative products in the United States. The patent laws are all about creating desirable incentives. In that respect, Section 271(f) is seriously counterproductive.

As we have seen, recent Federal Circuit decisions have greatly extended Section 271(f), but uncertainty remains as to whether the exten-

---

163. These rationales have included (i) determinations that the claims did not involve a “common nucleus of operative fact,” (ii) the discretion to decline to exercise supplemental jurisdiction pursuant to 28 U.S.C. § 1367(c), and (iii) the *forum non conveniens* doctrine. See Briefs of Appellant and for the United States, the American Intellectual Property Law Ass’n., & the Federal Circuit Bar Ass’n. as Amici Curiae Supporting Appellant, *supra* note 161; *Ideal Instruments, Inc. v. Rivard Instruments, Inc.*, 434 F. Supp. 2d 598, 630-31 (N.D. Iowa 2006); *cf.* *SRAM Corp. v. Sunrace Roots Enter. Co.*, 390 F. Supp. 2d 781, 784-85 (N.D. Ill. 2005).

164. See generally briefs discussed *supra* note 161.

sions will last. Supreme Court review in *Microsoft v. AT&T* might broadly confirm—or reverse—the application of Section 271(f) to: (i) intangible “components” such as information, patterns, and software; and (ii) foreign-made combination products that include only foreign-made copies or representations of the U.S.-originated items such as information, patterns, software, designs, etc. It might also influence the applicability of Section 271(f) to method claims, although that issue is not directly presented.<sup>165</sup> Broad clarification, however, may be slow in coming. Recent statements by Chief Justice Roberts, like the Court’s recent decision in *eBay*, indicate a preference for narrow, rather than broad, rulings.<sup>166</sup> Accordingly, we may not have real clarity on these questions in the near future.

Even if the Court rolls back the recent extensions, the discrimination against U.S. producers inherent in Section 271(f) will remain as to many types of components and combination products, and that provision will still lack any incentivizing benefits for innovation. Further, the perverse incentives against investment and production in the United States would be reduced only modestly by, for example, a broad determination that Section 271(f) does not apply to method claims.<sup>167</sup> Complete repeal of Section 271(f) is the preferred alternative.

## V. FUTURE APPLICATIONS OF SECTION 271(F) AND SPECIAL ASPECTS OF THAT PROVISION IN CONTEMPORARY PATENT ENFORCEMENT

Although it is an undesirable provision of law, Section 271(f) may remain on the books for a considerable period in the future and may have important applications that are not yet evident. Accordingly it is interesting, in technical as well as economic and policy terms, to consider some new ways in which Section 271(f) may be applied in the near future, if it is not repealed or drastically curtailed.

---

165. See *supra* notes 84-85 and accompanying text regarding the Supreme Court’s recent grant of review in *AT&T v. Microsoft* and the matters the Court may decide in that case.

166. In his remarks at the May, 2006 Georgetown Law School Commencement, Chief Justice Roberts’ main theme was that the Court should never rule more broadly than necessary to resolve the case before it. See, e.g., E.J. Dionne Jr., *The Chief Justice Sets a Standard*, WASH. POST, June 20, 2006, at A-17. Whether based on this precept or not, the Court issued a very narrow and uninformative ruling in *eBay, Inc. v. MercExchange, LLC*, 126 S. Ct. 1837 (May 15, 2006), its most recent patent law decision.

167. See *supra* text accompanying note 57 regarding the reduction of the provision’s perverse incentives that would result from such a ruling.

### A. High-Value Components and Information Components

Factors influencing the application of Section 271(f) include: (i) the modern globalized patterns of production and assembly of components; (ii) the product areas where U.S. producers remain ahead of, or competitive with, their foreign counterparts in supplying components; and (iii) the large numbers of patents being issued on inventions that, themselves, involve large numbers of distinct components.

One pattern we have seen involves foreign assembly of relatively low-value components made in various countries with one or a few high-value components produced in the United States.<sup>168</sup> This pattern makes it likely that liability under Section 271(f) will be asserted in any dispute involving a U.S. patent claiming the foreign-assembled combination or a portion thereof.

Sophisticated semiconductor microchips used as components of larger foreign-assembled devices fit this pattern in those areas where U.S. producers are prominent in producing such chips. U.S. produced software is another likely example, given the strengths of U.S. software producers and the ease with which such intangible components can be supplied from the United States to any location in the world for inclusion in larger systems and devices. Other U.S. supplied components to which Section 271(f) might be applied could include computer virus signatures, software updates or patches, and data compilations that enable patented inventions to operate or to update their operation.

### B. Genomics and Stem Cells

Moving to the edge of the computer arena, genomic analysis systems and methods might easily lead to infringement charges under Section 271(f) against U.S. producers of several divergent types of “components.” For one thing, silicon microchips having thousands of specialized receptors to detect specific DNA base sequences will apparently be one key component of automated systems for providing economical genomic analysis and typing of individuals. Specialized software for processing the raw data and analyzing it against genotypes in the population, disease correlations, and/or therapeutic sensitivity correlations would likely be another important component. A database of population genetics, drug sensitivity correlations, disease correlations, etc.—some or all of which might be subject to periodic updating—might also be an important component of such systems and methods. Given U.S. strengths in advanced microchips, analytical software, and genetic research, all three of these types of “com-

---

168. See *supra* text accompanying note 59.

ponents” might well be produced in the United States. At the same time, a variety of hardware items, from fluid-handling equipment to computer hardware, electrical systems, display/communications devices, etc, will also be required. Many of these may be produced, and the entire machines may be assembled, abroad.

The patents on such systems and/or methods might cover implementations with any of a range of analytic microchips, any of a variety of analytical software products, and any of a range of genetic databases or updates thereof. Variations of these key components might or might not be patented. The owner of a U.S. patent on all or a part of the system or method might use Section 271(f) to bar U.S. producers from supplying any of a range of microchips, software products, and/or genetic databases for use in such systems or methods anywhere in the world. To the extent they recognize this possibility, technically capable companies, wherever based, would think twice before investing in U.S. facilities and training of U.S. workers to produce such microchips, software, or databases. Indeed, since the data “components” might come directly from research activities, such companies might decide to locate their research and development activities, in addition to production facilities, outside of the United States.

As a final point of speculation, it is widely hoped that stem cells may become important bases for treatment of a range of human ills. If so, such cells or cell lines may become “components” of therapeutic methods that will be practiced abroad or of therapeutic systems that will be assembled and used abroad. Depending on the circumstances, the holder of a U.S. patent on part or all of such a method or system might use Section 271(f) to assert infringement on the part of the U.S. developer/supplier of the cells or cell lines and/or information used in connection with them even though the patent holder had not, and perhaps could not have, patented the cells or information.

### **C. Patent Trolls**

In the last decade, numerous entities have been formed to identify and acquire unused or “quiescent” patents<sup>169</sup> and assert them, via demand letters, threats of suit, or actual litigation, against companies whose products or processes may fall within the patent claims. The organizers of such entities have usually been patent attorneys and have often arranged financial

---

169. Such “quiescent” patents may be obscure and/or owned by companies in bankruptcy or by estates of the deceased inventors. They may have expired for failure to pay maintenance fees—so long as they are still subject to revival by appropriate application to the PTO. See 37 C.F.R. § 1.378(c), which makes it possible in many cases to revive a patent that has expired for failure to pay maintenance fees.

backing by venture capitalists or similar investors.<sup>170</sup> Critics have labeled these entities “patent trolls,” a term that has gained wide currency.<sup>171</sup>

Technology companies and other critics emphasize that trolls play no part in either creating the underlying inventions or commercializing or producing the products covered by the patents they assert. Overall, the critics say, trolls unfairly exploit the weaknesses of the U.S. patent system against entities that are performing the hard work of innovation.<sup>172</sup> In this

170. See FTC REPORT, *supra* note 19, ch. 2 at 31, ch. 3 at 37-41; NAS REPORT, *supra* note 19, at 31. Both reports describe the emergence of patent enforcement entities backed by venture capital, private equity, or hedge fund investors hoping to profit by acquiring and enforcing theretofore unasserted patents.

171. See, e.g., Cutler, *supra* note 132 (describing the presentations and discussion at the four “town meetings” on patent reform sponsored by the National Academies’ Board on Science, Technology, and Economic Policy, the Federal Trade Commission, and the American Intellectual Property Law Association in connection with the recommendations for reform of the U.S. patent laws made in the NAS Report and FTC Report). For example, the Patent, Trademark & Copyright Journal reported regarding one of those “town meetings”:

Much debate centered on patent “trolls,” a term of derision for attorneys and individuals who send letters to and sue companies alleging patent infringement and seeking royalties. “Right now injunctions are being used to effectively deprive [companies in the industry of] the ability . . . to deliver value to customers, to deliver innovation,” [Microsoft’s William] Poole said. Patent trolls, he said, “don’t ever actually commercialize an invention, [but] they . . . hold up someone that is trying to commercialize that invention, that is trying to deliver the value to consumers. And they hold them up via threat of injuncti[ons]. And we think this is wrong.”

*Id.* T. Andrew Culbert of Microsoft said:

The troll concept refers to the person on the bridge who doesn’t contribute to innovation but . . . manages to abuse the system, use the system in a way it was not intended to be used, but then tax[es] the people who do actually . . . innovate and who want to deliver those innovations to the end users.

*Id.*; see also Joe Beyers, *Rise of the Patent Trolls*, CNET NEWS.COM, Oct. 12, 2005, [http://news.com.com/riseofthepatenttrols/2010-1071\\_3-5892996.html](http://news.com.com/riseofthepatenttrols/2010-1071_3-5892996.html). Beyers observed:

The shakedown is on. In the aftermath of the dot-com bust, a new kind of business with a simple, yet potentially lethal, model has emerged. Call them “patent trolls.” These operators have no products or customers. Yet they wield the power to bring the companies that actually make and sell products to their knees. . . . In recent years, trolls have raised massive amounts of money. . . . Trolls have an inherently unfair advantage in extracting value from operating companies due to the nature and uncertainty of the patent judicial system.

Beyers, *supra*.

172. See *supra* note 171. According to media interviews, trolls often search through large numbers of outstanding patents seeking those that: (i) have broad claims that cover,

view, trolls harm innovation and the U.S. economy and amount to a misuse of the patent system.<sup>173</sup>

The typical troll business plan probably does not rely substantially on Section 271(f). Trolls can apparently find many useful quiescent patents and many possible infringers arguably subject to large infringement damages and/or license fees by relying solely on domestic production, sales, and use.

Nevertheless, Section 271(f) does have a vastly disproportionate value and usefulness for patent trolls compared to other patent holders. This special value reflects a timing-related difference between: (i) what is required—early-on—to incentivize inventive efforts and investment; and (ii) what makes a patent valuable, years later, in the hands of a potential en-

---

or might be read as covering, products that have come into substantial use by others; but (ii) lack any history of licensing or assertion, i.e., are quiescent and ignored. According to critics, the acquired patents, though broad, are often invalid, but the activities of trolls are nevertheless a significant burden on many U.S. producers because of the costs and risks of unpredictable patent litigation. For the target of a troll, the economics of litigating validity is generally unattractive because: (i) the troll has little at risk and can rely on the statutory presumption of validity, 35 U.S.C. § 282; whereas (ii) the accused infringer must bear the burden of proof and must demonstrate invalidity by “clear and convincing” evidence; and (iii) as an operating company, the accused infringer is likely to suffer severe disruption and injury if the patent should be found valid and infringed. Where the trolls are particularly concerned about weaknesses of the patents they are asserting and negotiations do not lead to the license fees they seek, trolls may sue a few target companies at a time but, depending on the circumstances, and demand relatively modest royalties of each one. Their apparent aim in such cases is to demand low enough royalties that the target companies will settle and pay rather than going to the expense of litigating the validity of the subject patents. *See, e.g.*, FTC REPORT, *supra* note 19, ch. 2 at 31, ch. 3 at 37-41.

173. Numerous published accounts refer to the particular problems such trolls present for companies that produce technologically advanced products. In many technologies, such as semiconductors, computer hardware and software, and electronics generally, routine cross-licensing of patents is essential to allow the introduction of new products. Absent such routine licensing, the large numbers of existing patents (sometimes termed “patent thickets”) may totally block new product introductions or, in other instances make innovation very difficult. Fortunately, ready cross-licensing, with or without cash payments to balance the perceived worth of the rights licensed by each side, has long been routine. That is largely because no enterprise actually producing and introducing products could succeed without regularly obtaining licenses to patents held by others in return for licensing its own. In contrast, trolls have no commercial operations that could be threatened by the patent portfolios of other companies and no interest in obtaining licenses of such patents. Accordingly, they are not constrained in threatening to shut down the operations of producing companies via injunctive orders under their acquired patents—assuming they would be upheld. *See generally* sources cited *supra* note 172. The more aggressive stance and different incentives of the trolls have been troubling to a number of U.S. technology companies. *Id.*

forcer. Years after outstanding patents have issued, a patent troll can review large numbers of them to see which: (i) have claims that may cover products or processes that have, by then, come into substantial use; and (ii) are available cheaply. Importantly, the troll can evaluate the potential recovery from asserting those patents based not only on the size of the U.S. market for the claimed product or process but also taking into account whether one or more U.S. producer supplies one or more component for inclusion in the claimed product or process as assembled or practiced abroad. If so, Section 271(f) is ideal for capturing, in a single U.S. lawsuit, large incremental damages from worldwide sales—and for threatening injunctions against worldwide sales of components made in the United States.

Whether Section 271(f) claims have played a material role, to date, in negotiations between trolls and U.S. producers of technologically advanced components is unclear, because the negotiations and settlements entered into by trolls are generally kept confidential. Recent testimony by the founder-principal of a large alleged troll, however, showed his awareness of the particular usefulness of Section 271(f). Apparently for that reason, he strongly opposed limitations on, or repeal of, Section 271(f).<sup>174</sup> It seems likely that the increased bargaining power and enhanced recoveries Section 271(f) often provides can make the difference between a marginal patent troll and one that is profitable and attractive to private equity and other investors.<sup>175</sup>

---

174. See Neil Graham, *Legislation/Patents: House Continues Debate on Patent Reform; Momentum Builds As Differences Sharpen*, PAT. TRADEMARK & COPYRIGHT J. (BNA) (May 6, 2005) (describing the April 28, 2006 testimony of Nathan Myhrvold before the House Subcommittee on Courts, the Internet and Intellectual Property). Myhrvold, formerly Chief Technology Officer at Microsoft, founded Intellectual Ventures, which is often identified as a mysterious entity resembling a large and wealthy patent troll. See, e.g., Nicholas Varchaver, *Who's Afraid of Nathan Myhrvold?*, FORTUNE, June 26, 2006, available at [http://money.cnn.com/magazines/fortune/fortune\\_archive/2006/07/10/8380798/index.htm](http://money.cnn.com/magazines/fortune/fortune_archive/2006/07/10/8380798/index.htm). In that testimony, Myhrvold opposed limitations on Section 271(f) as well as limitations on the pre-*eBay* rule of automatic injunctions in patent cases where infringement has been found. Automatic injunctions, like Section 271(f) in appropriate cases, greatly enhance the leverage of trolls *vis-à-vis* putative infringers.

175. The *Eolas* example shows how advantageous Section 271(f) can be in increasing the pay-off from a single U.S. lawsuit. See 399 F.3d 1325 (Fed. Cir. 2005). In *Eolas*, two-thirds of the huge damages award entered by the district court was based on foreign sales made relevant and admissible solely because of Section 271(f). *Id.* Presumably Section 271(f) would be particularly important to the economics of a troll that concentrates on industries such as electronics, telecommunications, computer hardware and software, biotechnology, etc., where many technologically advanced components are often combined into larger patentable products or processes.

Most commentators consider the activities of patent trolls and their impact on the U.S. economy to be undesirable overall. Accordingly, any adverse effects on the economics of their operations as a consequence of narrowing or repealing Section 271(f) would seem to be, overall, positive for innovation and for the U.S. economy as a whole.

## VI. CONCLUSION

Although intended to promote innovation, the patent laws can also impede innovation and competition and threaten social welfare. As both the NAS Report and the FTC Report emphasize, it is important to limit the adverse effects of patents if they are to produce net benefits to society.<sup>176</sup>

Section 271(f) provides no significant incentives for innovation, investment, or disclosure while creating perverse incentives for existing and prospective U.S. producers and investors. Even the good news about that provision is bad: actual or prospective U.S. producers can avoid its discriminatory burdens and achieve relative safety by moving their productive facilities to other countries or contracting-out production to foreign suppliers. But whenever a company chooses one of these options, the U.S. loses technology, sector jobs, tax revenues, and portions of our technological and industrial base.

---

176. See, e.g., *Lab. Corp. of Am. v. Metabolite Labs., Inc.*, 126 S. Ct. 2921, 2922 (2006) (Breyer, J., dissenting from dismissal of writ of certiorari) (“[S]ometimes too much patent protection can impede rather than ‘promote the Progress of Science and useful Arts’”); *id.* at 2929 (discussing the “important debate” as to whether our current patent system reflects the proper balance it should embody); FTC Report, *supra* note 19, Executive Summary at 2-12, 14-16, ch. 1, passim. In *Great Atlantic & Pacific Tea Co. v. Supermarket Equipment Corp.*, the Supreme Court explained:

Every patent is the grant of a privilege of exacting tolls from the public. The Framers plainly did not want those monopolies freely granted. The invention, to justify a patent, had to serve the ends of science—to push back the frontiers of chemistry, physics, and the like; to make a distinctive contribution to scientific knowledge . . . . “[Recognition of excessive patent rights] tends rather to obstruct than to stimulate invention. It creates a class of speculative schemers who make it their business to watch the advancing wave of improvement, and gather its foam in the form of patented monopolies, which enable them to lay a heavy tax upon the industry of the country, without contributing anything to the real advancement of the arts. It embarrasses the honest pursuit of business with fears and apprehensions of concealed liens and unknown liabilities to lawsuits and vexatious accountings for profits made in good faith.”

340 U.S. 147, 154-55 (1950) (quoting *Atl. Works v. Brady*, 107 U.S. 192, 200 (1882)).

Section 271(f) was presumably well intentioned and might not have done much harm in the economic, technological, and legal environment of the 1970s and early 1980s. Even then, however, it amounted to a misguided overreaction to the unusual circumstances presented in *Deepsouth*. Present day patterns of international production and competition, revolutions in technology, and the proliferating numbers of patents have multiplied the circumstances where litigants can invoke that provision. These same factors together with judicial decisions expanding the reach of Section 271(f) and the increased strength and availability of patent remedies have likewise increased the impact of Section 271(f) when it is invoked. U.S. producers have also become more vulnerable to the discrimination Section 271(f) brings to patent enforcement. At the same time, advances in international patent cooperation have reduced the always flimsy justifications for Section 271(f).

Section 271(f) threatens U.S. interests and lacks any redeeming benefits. Congress should repeal that provision without further delay. Failing that, its application should be cut back as drastically as possible.



# LAW AND THE SCIENCE OF NETWORKS: AN OVERVIEW AND AN APPLICATION TO THE “PATENT EXPLOSION”

By Katherine J. Strandburg, Gábor Csárdi, Jan Tobochnik,  
Péter Érdi, & László Zolányi<sup>†</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1294
II. NETWORK SCIENCE: WHAT IT IS AND WHY LEGAL SCHOLARS SHOULD CARE .....	1300
A. WHAT NETWORK SCIENCE IS .....	1300
1. <i>How to Describe a Network: Common Structural Signatures of             Diverse Networks</i> .....	1301
a) Not Your Average Node: “Scale-Free” Degree Distributions .....	1302

---

© 2006 Katherine J. Strandburg, Gábor Csárdi, Jan Tobochnik, Péter Érdi, and László Zolányi

<sup>†</sup> Katherine J. Strandburg is an Associate Professor, DePaul University College of Law. Gábor Csárdi and László Zolányi are doctoral students, Department of Biophysics, KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, Budapest, Hungary, and Center for Complex Systems Studies, Kalamazoo College. Jan Tobochnik is a Dow Distinguished Professor of Natural Science, Department of Physics and Center for Complex Systems Studies, Kalamazoo College. Péter Érdi is a Henry R. Luce Professor, Departments of Physics and Psychology and Center for Complex Systems Studies, Kalamazoo College and Head, Department of Biophysics, KFKI Research Institute for Particle and Nuclear Physics of the Hungarian Academy of Sciences, Budapest, Hungary.

The authors would like to thank David Adelman, Stacey Dogan, Rochelle Dreyfuss, Brett Frischmann, Jay Kesan, Michael Meurer, David G. Post, Joshua Sarnoff, participants at the Symposium on Complexity and Advanced Analytics Applied to Business, Government and Public Policy, University of Michigan-Dearborn, October 2004; the 2005 March Meeting of the American Physical Society; the Intellectual Property Scholars Conference at Cardozo Law School, August 2005; the Midwest Law and Economics Association meeting, October 2005; the Colloquium on Technology, Innovation, and Intellectual Property Policy, University of Arizona Law School, November 2005; a faculty workshop at the University of Illinois College of Law; and the NetSci International Conference on Network Science, Bloomington, IN, May 2006, for helpful comments on this project. We are also grateful to Katherine Strandburg’s research assistants, Douglas Burda and Elizabeth Levine, for substantive comments and assistance with preparing the manuscript.

b)	Building Bridges: The Betweenness Metric .....	1305
c)	Six Degrees of Separation: The “Small World” Property ..	1305
d)	Transitivity and Local Network Structure .....	1306
2.	<i>How Does Your Network Grow? “Preferential Attachment” and Other Factors in Network Evolution .....</i>	1307
3.	<i>Network Structure and Collective Behavior.....</i>	1310
B.	WHY LEGAL SCHOLARS SHOULD CARE ABOUT NETWORK SCIENCE .....	1310
1.	<i>The Implications of Heterogeneity .....</i>	1311
2.	<i>The Importance of Network Structure in Determining Individual Responses to Legal and Social Change .....</i>	1313
3.	<i>The Complicated Influence of Local Network Structure on Collective Behavior.....</i>	1314
III.	<b>NETWORK CLUES TO A DECREASING PATENTABILITY STANDARD .....</b>	1318
A.	BACKGROUND.....	1319
1.	<i>The Patent System and Its Discontents.....</i>	1319
2.	<i>Patent Issuance and Prior Art Citations .....</i>	1323
3.	<i>Relationship of this Study to Previous Uses of Patent Citation Data.....</i>	1325
B.	THE EVOLVING PATENT CITATION NETWORK: IS PATENTING GETTING OUT OF HAND? .....	1329
1.	<i>Concerning the Pace and Breadth of Innovative Progress .....</i>	1331
2.	<i>The Evolving Citability Distribution .....</i>	1333
3.	<i>The Increasing Stratification of Patent Citation Patterns.....</i>	1337
IV.	<b>OF SLEEPER PATENTS AND INNOVATIVE PROGRESS: HOW INNOVATION BUILDS ON PRIOR TECHNOLOGY .....</b>	1340
V.	<b>FUTURE DIRECTIONS AND SOME PRELIMINARY RESULTS ..</b>	1342
A.	PATENT CLASSIFICATION AND THE MEANING OF ANALOGOUS ARTS .....	1343
B.	THE SHRINKING PATENT CITATION NETWORK .....	1344
C.	PATENT THICKETS AND POTENTIAL ANTI-COMPETITIVE LICENSING PRACTICES .....	1346
VI.	<b>CONCLUSIONS.....</b>	1348

## I. INTRODUCTION

Networks are powerful tools for understanding the interconnectedness of the modern era. The internet no doubt motivates our fascination with networks, as it not only ties us together technologically in ways that were not previously possible, but also leads us to conceptualize the world in a

more relational way.<sup>1</sup> Looking through this conceptual lens, one finds that networks are everywhere. A network, consisting of "nodes" and "links," may be a group of individuals linked by friendship or commerce, a group of computers linked by network cables, a nervous system, a system of roads or airline flights, a collection of cracks in the earth, a group of patents and the citations between them, or another of a virtually limitless variety of systems of connected "things."

Consistent with the increased salience of the network concept, the past few years have seen an explosion of interest in "network science," which attempts to move beyond the network metaphor to an analysis of commonalities between different networks. Natural and social scientists in fields from physics to sociology to biology have seized upon networks as important analytical and explanatory tools.<sup>2</sup> Popular books and articles discuss the extent to which people, places, and things are "linked"<sup>3</sup> or separated by only "six degrees."<sup>4</sup> Prominent universities have inaugurated inter-disciplinary centers and programs relating to the study of networks.<sup>5</sup> Most recently, a few legal scholars have begun to apply some of the techniques and insights of network science to the empirical and theoretical study of law.<sup>6</sup>

---

1. *See generally* DARIN BARNEY, *THE NETWORK SOCIETY* (2004) (critiquing the ubiquitous application of the network metaphor to modern society).

2. For overview discussions, see, for example, ALBERT-LASZLO BARABASI, *LINKED: THE NEW SCIENCE OF NETWORKS* (2002); S. N. DOROGVTSEV & J. F. F. MENDES, *EVOLUTION OF NETWORKS FROM BIOLOGICAL NETS TO THE INTERNET AND WWW* (2003); *MODELS AND METHODS IN SOCIAL NETWORK ANALYSIS* (Peter J. Carrington, John Scott & Stanley Wasserman eds., 2005) [hereinafter Carrington, Scott & Wasserman]; ROMUALDO PASTOR-SATORRAS & ALESSANDRO VESPIGNANI, *EVOLUTION AND STRUCTURE OF THE INTERNET: A STATISTICAL PHYSICS APPROACH* (2004); DUNCAN J. WATTS, *SIX DEGREES: THE SCIENCE OF A CONNECTED AGE* (2002); Reka Albert & Albert-Laszlo Barabasi, *Statistical Mechanics of Complex Networks*, 74 *REVS. MOD. PHYS.* 47 (2002); M.E.J. Newman, *The Structure and Function of Complex Networks*, 45 *SIAM REV.* 167 (2003).

3. BARABASI, *supra* note 2.

4. WATTS, *supra* note 2.

5. *See, e.g.*, Center for Advanced Study, University of Illinois, Age of Networks program, <http://www.cas.uiuc.edu/networks.html> (last visited Nov. 22, 2006); Cornell University Institute for the Social Sciences, Theme Project 2005-08, Getting Connected: Social Science in the Age of Networks, [http://www.socialsciences.cornell.edu/0508/networks\\_desc.html](http://www.socialsciences.cornell.edu/0508/networks_desc.html) (last visited Nov. 22, 2006); Northwestern University Institute on Complex Systems, <http://www.northwestern.edu/nico> (last visited Nov. 22, 2006).

6. The application of network science to law is in its infancy. While many law review articles deal with the topic of network effects or discuss the internet and other networked infrastructure, the paucity of articles applying network science or modeling from a statistical physics, computer science, or social network theory perspective is evident.

This Article seeks to accomplish two tasks with respect to the application of network science to the study of law. First, in Part II, it seeks to persuade readers that legal scholars should jump on the network bandwagon in greater numbers because of the important conceptual advances and analytical tools that network science provides. Specifically, network science highlights the potential importance of heterogeneity and local relationship patterns in determining collective behavior. Part II also underscores and begins to address the difficulty of predicting collective behavior from individual interactions or transactions. As the science of networks continues to develop, it promises to provide new approaches to the analysis of empirical data and new tools for modeling the expected results of legal

---

For example, a LEXIS search for law review articles citing either of the authors of the most well-known popular works on network science (BARABASI, *supra* note 2 or WATTS, *supra* note 2), uncovers only about ten articles making more than a passing reference to network science. Yet even this small group demonstrates the breadth of potential applications of network science to legal scholarship. *See, e.g.*, Ian Ayres & Katharine K. Baker, *A Separate Crime of Reckless Sex*, 72 U. CHI. L. REV. 599 (2005); Mark Cooper, *From Wifi to Wikis and Open Source: The Political Economy of Collaborative Production in the Digital Information Age*, 5 J. TELECOMM. & HIGH TECH. L. 125 (2006); Claire Moore Dickerson, *Corporations As Cities: Targeting the Nodes in Overlapping Networks*, 29 IOWA J. CORP. L. 533 (2004); Donald C. Langevoort, *Internal Controls After Sarbanes-Oxley: Revisiting Corporate Law's "Duty of Care as Responsibility for Systems,"* 31 J. CORP. L. 949 (2006); Michael J. Madison, *Social Software, Groups, and Governance*, 2006 MICH. ST. L. REV. 153 (2006); Andrea M. Matwyshyn, *Material Vulnerabilities: Data Privacy, Corporate Information Security, and Securities Regulation*, 3 BERKELEY BUS. L.J. 129 (2005); Andrea M. Matwyshyn, *Of Nodes and Power Laws: A Network Theory Approach to Internet Jurisdiction Through Data Privacy*, 98 NW. U. L. REV. 493 (2004) [hereinafter Matwyshyn, *Of Nodes and Power Laws*]; Lawrence E. Mitchell, *Structural Holes, CEOs and Informational Monopolies: The Missing Link in Corporate Governance*, 70 BROOK. L. REV. 1313 (2005); Daniel F. Spulber & Christopher S. Yoo, *On the Regulation of Networks as Complex Systems: A Graph Theory Approach*, 99 NW. U. L. REV. 1687 (2005); Lior Jacob Strahilevitz, *A Social Networks Theory of Privacy*, 72 U. CHI. L. REV. 919 (2005). For a seminal study of the network of legal citations, see David G. Post & Michael B. Eisen, *How Long is the Coastline of the Law? Thoughts on the Fractal Nature of Legal Systems*, 29 J. LEGAL STUD. 545 (2000). We are also aware of related work by Seth J. Chandler, *The Network Structure of Supreme Court Jurisprudence* (University of Houston Law Center No. 2005-W-01, 2005), available at <http://ssrn.com/abstract=742065>; Gavin Clarkson, *Objective Identification of Patent Thickets: A Network Analytic Approach, in Essays on Intellectual Asset Management 74* (May 2004) (unpublished D.B.A. thesis, Harvard Business School); Tracey L. Meares & Kelsi Brown Corkran, *When 2 or 3 Come Together* (Univ. of Chicago, Public Law Working Paper No. 107, 2005), available at <http://ssrn.com/abstract=835664>; Thomas A. Smith, *The Web of Law* (San Diego Legal Studies Research Paper No. 06-11, 2005), available at <http://ssrn.com/abstract=642863>. *See also* YOCHAI BENKLER, THE WEALTH OF NETWORKS 241-61 (2006) (using network science studies to gain insight into the political significance of the structure of the World Wide Web).

change. These analytical approaches and modeling tools complement the existing empirical methods and theoretical models of law and economics and other interdisciplinary approaches.

Second, in Parts III, IV, and V, the Article illustrates the application of a network approach to empirical data by describing the results of a network science study of patent citations. The overall goal of the study is to gain a better understanding of the evolution and structure of technical relationships between patents in the United States patent system, and to investigate current patenting behavior so as to gain insight into innovation policy and law. The United States patent system has been the subject of increasing criticism in recent years.<sup>7</sup> There is a perception that recently issued patents sweep more broadly than their inventors' contributions justify, and that some are being issued for inventions that should be deemed obvious. Critics also contend that patents increasingly impose transaction costs in areas, such as scientific research, software, and business methods, in which patents may not be needed to promote technological advances. Recently proposed legislation<sup>8</sup> and renewed Supreme Court attention to patent law reflect these criticisms.<sup>9</sup> Responses to the criticism highlight the contributions of intellectual property to the nation's economy and ar-

---

7. See, e.g., FEDERAL TRADE COMMISSION, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY (2003), available at <http://www.ftc.gov/os/2003/10/innovationrpt.pdf> [hereinafter FEDERAL TRADE COMMISSION]; NAT'L RESEARCH COUNCIL OF THE NAT'L ACADEMIES, A PATENT SYSTEM FOR THE 21ST CENTURY (Stephen A. Merrill, Richard C. Levin & Mark B. Myers, eds., 2004); ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PATENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS AND WHAT TO DO ABOUT IT (2004); Rochelle C. Dreyfuss, *Pathological Patenting: The PTO As Cause or Cure*, 104 MICH. L. REV. 1559 (2006) (critiquing Jaffe and Lerner for its emphasis on patent invalidity and pointing out other potential sources of problems with the patent system); Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anti-commons in Biomedical Research*, 280 SCI. 698 (1998).

8. See Patent Reform Act of 2005, H.R. 2795, 109th Cong. (2005).

9. See *Ill. Tool Works, Inc. v. Indep. Ink, Inc.*, 126 S.Ct. 1281 (2006); *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 126 S.Ct. 2921, 2921 (2006) (Breyer, J., dissenting); *eBay, Inc. v. MercExchange, L.L.C.*, 126 S.Ct. 1837, 1842 (2006) (Kennedy, J., concurring); *KSR Int'l Co. v. Teleflex, Inc.*, No. 04-1350 (U.S. argued Nov. 28, 2006); *Microsoft Corp. v. AT&T Corp.*, 414 F.3d 1366, cert. granted, 127 S. Ct. 467 (2006) (No. 05-1056); *Merck KGaA v. Integra Lifesciences I, Ltd.*, 545 U.S. 193 (2005); see also Marcia Coyle, *Justices Ponder Heavy Patent Docket*, NAT'L L. J. (Dec. 12, 2005), available at <http://www.law.com/jsp/article.jsp?id=1134122713321> (arguing that increased Supreme Court attention to patent law is due to "[a] constellation of factors, according to scholars and litigators, not the least of which is unhappiness with the U.S. Court of Appeals for the Federal Circuit.").

gue that increasing numbers of patents do not necessarily signify an erosion of patent quality.<sup>10</sup>

Part III of this Article employs a network science approach to interpreting the burgeoning patenting of the past few decades.<sup>11</sup> In our study, the “nodes” of the network are United States patents and the “links” are citations of one patent by another.<sup>12</sup> Patent citations indicate technological relationships between the patented technologies. For the most part, a citation from one patent to another indicates either that the later patent builds upon the technology of the earlier patent or that the claimed inventions are closely enough related that the earlier technology was material to determining whether the later patent should issue. The citation network thus provides a “map” of the relationships between patented technologies which can be explored using network science techniques.

We have performed a detailed study of the evolution of the patent citation network since 1975. In looking at how the network structure has evolved, we distinguish between three possible explanations for the recent increase in patenting: faster technological progress; increasing breadth of patented technologies due either to expansion of patentable subject matter or to new fields of invention; and a lowering of the legal standard for patentability leading to the issuance of patents for more trivial innovative steps. These possibilities are not mutually exclusive, of course, nor are they the only factors that might affect citation patterns. The question we ask here is whether one can explain the observed evolution of the patent network solely in terms of an increasing pace and breadth of technological advance or whether some additional factor is needed to understand the observed network growth and structure.

Our analysis leads us to conclude that increasing pace and breadth of innovation alone are unlikely to explain the way in which the patent citation network has evolved in recent years. This is primarily because, as we explain in detail in Part III, the citation network has undergone a measurable increase in stratification since the late 1980s, following a modest decrease earlier in the decade. By “increasing stratification” we mean that

---

10. See, e.g., *Perception Gap Hindering Efforts to Improve Patent System*, Dudas says, 71 PAT. TRADEMARK & COPYRIGHT J. (BNA) 374 (2006).

11. Our patent citation study is similar in flavor to recent studies of citations in scientific journal articles, see Sidney Redner, *Citations Statistics from 110 Years of Physical Review*, 58 PHYSICS TODAY 49 (2005), and of the “web of law” consisting of cases and other legal authorities and the citations linking them. See Smith, *supra* note 6; Chandler, *supra* note 6.

12. United States patents also cite scientific literature and foreign patents. Our data do not include these additional links.

there is an increasing disparity in likelihood of citation between the patents most and least likely to be cited again. In the empirical analysis described below, the probability that a patent will be cited again depends upon the patent's age and number of previous citations received. Because econometric studies have indicated that patent citations are linked to the technological importance of the patented technology,<sup>13</sup> a reasonable interpretation of this observation is that patents are being issued for increasingly trivial advances.<sup>14</sup> The timing of the increasing stratification roughly correlates with increasing reliance by the U.S. Court of Appeals for the Federal Circuit on the widely criticized “motivation or suggestion to combine” test for nonobviousness,<sup>15</sup> and with other indications of an increasingly lax legal standard of patentability.<sup>16</sup>

In Part IV of this Article we describe how network science illuminates the question of how innovation proceeds by combining prior technologies. We examine the dependence of patent “citability,” or probability that a patent will be cited again, on the age of the patent. Patent citability dies out surprisingly slowly with age *even for patents that have rarely or never been previously cited*. This means that there are “sleeper” patents which are cited after long periods of neglect, perhaps signifying how innovation proceeds not only by building incrementally on previous advances but also by “doubling back” to make new use of previously neglected technology. These results cast doubt on the possibility that the value of an innovation can be identified confidently at the time of patenting.<sup>17</sup> We briefly review

---

13. See, e.g., John R. Allison, Mark A. Lemley, Kimberly A. Moore & R. Derek Trunkey, *Valuable Patents*, 92 GEO. L.J. 435, 449 n.60 (2004) and citations therein.

14. Other possible interpretations are discussed *infra* Part III.B.

15. See U.S. Department of Commerce, MANUAL OF PATENT EXAMINING PROCEDURES § 2143.01 (8th ed., 2006) [hereinafter MPEP]. See, e.g., *In re Rouffet*, 149 F.3d 1350, 1357 (Fed. Cir. 1998). The “suggestion test” has its roots much earlier in the jurisprudence of the Court of Customs and Patent Appeals. See 2 DONALD S. CHISUM, CHISUM ON PATENTS § 5.04[1][e] (105th rel., 2006), for a discussion of the history of the test. However, it took on its modern form and became increasingly important at the Federal Circuit throughout the 1990s. The percentage of opinions involving nonobviousness citing the suggestion test rose from about 10-15% in the late 1980s to about 40-50% in the late 1990s. See *infra* Appendix, Figure 11.

16. See, e.g., John H. Barton, *Non-Obviousness*, 43 IDEA 475 (2003); Glynn S. Lunney, Jr., *Patent Law, the Federal Circuit, and the Supreme Court: A Quiet Revolution*, 11 S. CT. ECON. REV. 1, 11-21 (2004).

17. See, e.g., Allison et al., *supra* note 13 (developing metrics to identify valuable patents); David E. Adelman & Kathryn L. DeAngelis, *Patent Metrics: The Mismeasure of Innovation in the Biotech Patent Debate* 34, 35-40 (Arizona Legal Studies Discussion Paper No. 06-10, 2006), available at <http://ssrn.com/abstract=881842> (critiquing the approach of Allison et al.).

some earlier studies of the innovative process using a network approach and discuss how these approaches might be combined with ours to investigate the innovative process.

Part V of the Article describes several other possibilities for applying network science to the patent citation network to explore questions of legal significance. We first describe how the network concept of path length may be useful in quantifying the “closeness” of patented technologies. Numerous patent law doctrines, including patent validity and patent infringement under the doctrine of equivalents, are based on the perspective of the “person having ordinary skill in the art” or PHOSITA.<sup>18</sup> In addition, the obviousness of a claimed invention is judged in light of prior inventions in analogous arts.<sup>19</sup> Assessing the “closeness” of different technologies is essential to interpreting the concept of an “art” in applying these doctrines and is also of practical importance to patent classification.

We then describe preliminary studies of the “small world” property of the patent citation network and discuss prospects for further study of the implications of the increasing “connectedness” of the network. Finally, we describe network metrics that may be of use in exploring the extent to which closely related patented technologies are complements or substitutes, which is relevant to the problem of patent thickets and to issues of anti-competitiveness in patent licensing.

Part VI offers conclusions and returns to the broader point of this Article that network theory and modeling, as they are applied in social network analysis, computer science, statistical physics, and elsewhere, have great potential to complement other interdisciplinary approaches to the analysis of present and proposed legal doctrine.

## II. NETWORK SCIENCE: WHAT IT IS AND WHY LEGAL SCHOLARS SHOULD CARE

### A. What Network Science Is

Network science, whether it is rooted in the social sciences, computer science, or the natural sciences such as physics or biology, has three general, interrelated, and ongoing goals: (1) to measure, describe, and categorize network structure and the patterns of relationships between network nodes; (2) to understand network evolution and growth and its relationship

---

18. See, e.g., Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1185-91 (2002) (discussing the importance of the PHOSITA in patent law).

19. See, e.g., *In re Clay*, 966 F.2d 656, 658 (Fed. Cir. 1992).

to network structure; and (3) to understand how the collective behavior of entities connected in a network depends on and derives from the network's structure. Many open questions exist at all three levels. The overall intuition behind the interdisciplinary conversations that characterize “network science” is that common structures, growth patterns, and collective behaviors will arise in networks composed of very different kinds of elements and linkages. If this is the case, common concepts and methods will be useful in understanding widely varying networks and in answering the very different substantive questions posed by physicists, biologists, computer scientists, sociologists, and, most recently, by legal scholars.

1. *How to Describe a Network: Common Structural Signatures of Diverse Networks*

Network science seeks at the outset to measure and describe various relational structures found in the social or physical world. In a network in which the node is an individual, the links might be relationships of friendship, kinship, sexual involvement, business association, and so forth. When the node is an airport, the links might be flight paths. When the node is a business entity, the links might be supply and distribution relationships. And when the node is a patent, scientific journal, or legal opinion, the links can be citations. Links can be “directed,” meaning that the nodes on either end of it relate to each other in an asymmetric way. For example, a link between a patent and a reference that it cites has a direction to it: if one reversed the order of the nodes in the link, it would change what the link meant. Links can also be “undirected.” For example, a link indicating friendship is normally symmetric: “A is friends with B” normally means the same as “B is friends with A.”

The descriptive task of network science necessitates thinking about what quantities characterize networks—what metrics provide concise yet illuminating means of describing and comparing relational structures. Narrowly, it aims at providing a means to parameterize individual network structures. The broader descriptive goal is to delineate the similarities and differences between networks, with the hope of categorizing these structures in a logical way that can provide substantive insights into the underlying systems. Part of the fascination of network science is that similar metrics—degree distribution, “transitivity,” network “path lengths,” and others (some of which are discussed in this Article)—may be used to describe networks of very different “nodes” and “links.”<sup>20</sup> Strikingly similar

---

20. See, e.g., sources cited *supra* note 2.

results are obtained for some of these metrics when measuring seemingly disparate networks.

a) Not Your Average Node: “Scale-Free” Degree Distributions

One metric by which one may distinguish networks is the frequency distribution of the average number of connections between network nodes. The “degree” of a node is the number of connections or “links” it has to other nodes. Node degree is an intuitive measure of “importance.” A person in a social network who has many connections to others likely plays an important role in that network, for example. Of course, the meaning of such importance will be very different in a network of friends than it is in a network of legal citations, but the network approach brings to light underlying similarities between very different networks. One way to characterize a network is to study the “node degree distribution.” The node degree distribution charts the fraction of nodes in the given network that have a particular number of connections.<sup>21</sup>

If a network is produced by randomly connecting nodes to one another with a uniform probability of connection, the degree distribution will be symmetrically peaked around a “typical” number of edges determined by that probability of connection.<sup>22</sup> For large networks, the peak will be very narrow, so that most nodes will have approximately the “typical” number of edges. In this sense, the nodes of a random network can be said to be relatively homogeneous. In the random network analyzed in Figure 2, for example, the most common number of connections for a node is twenty, which is also the average and median number. Though there are plenty of nodes with ten to thirty connections, the numbers drop drastically outside the peak width. There are virtually no nodes with 100 connections.

The symmetrical, peaked distribution of Figure 2 is similar to the “normal” distribution or “Bell curve” that characterizes the probability distribution of many things in our everyday experience.<sup>23</sup> (See Figure 3 for an example of a normal distribution.) The fact that the height of human beings, for example, has this kind of normal probability distribution makes it possible to design furniture, showers, cars, and the like for the average

---

21. See, e.g., *infra* Appendix, Figure 1 (plotting the node degree distribution for the patent citation network).

22. See *infra* Appendix, Figure 2.

23. The probability distribution for the node degree of a random network is not precisely a normal or “Gaussian” distribution. Mathematically, the random network degree distribution takes another form typical for probability distributions—the Poisson distribution. Newman, *supra* note 2, at 197-99. For present purposes, the difference is unimportant. What is important is the peaked, symmetrical shape.

person and to expect that nearly all people will be able to use them, perhaps with minor adjustments.

One of the first and most interesting results of network science was the discovery that many networks encountered in the real world are not homogeneous like the random networks with degree distributions as illustrated in Figure 2. Instead, many real world degree distributions are highly skewed, and very broad, with what are called “fat tails.”<sup>24</sup> In these networks, rather than all nodes having roughly the same “typical” number of links, there are a large number of nodes with very few links and a small, but significant, number of nodes with very many links. These highly connected nodes may have hundreds or thousands of times as many connections as the “average” or “typical” node. Depending upon the network, the few highly connected nodes may play a much more important role than the large number of minimally connected nodes. In this kind of situation, it is not enough to make decisions based on the likely behavior of “typical” nodes. Because a significant fraction of nodes are so far from typical, the effects of non-typical nodes must be taken into account in understanding and predicting the behavior of the network. For example, in a communications network, the system might continue to operate despite failures of many “typical” nodes but may be completely defeated by the failure of one highly connected node. Any attempt to understand the collective behavior of such a network by focusing on the “typical” node is doomed to fail.

When the degree distribution is sufficiently broad, it will have a “power law” decaying tail. In some cases, network degree distributions may be so broad that they do not even have well-defined averages or standard deviations. Figure 4 illustrates a power law degree distribution, meaning that the fraction of nodes with a particular number of connections decays as  $d^{-\gamma}$  once  $d$  is large (where  $d$  is the degree of a node). This shows how it becomes more difficult to identify a “typical” node as the decay power  $\gamma$  becomes smaller. When  $\gamma$  is 3, the most likely number of connections is 0, while the median number is about 15 and the average number is about 20. When  $\gamma$  is even smaller, the most likely number of connections is still 0, but the median number is 40 and the average is essentially infinite because there are so many nodes with very large numbers of connections. The likelihood of having a particular number of connections for  $\gamma = 1.5$  is fairly flat and it would be a mistake to say that a node with 10, 40, or 100 connections is “typical.”

---

24. See *infra* Appendix, Figure 1.

Because there is no “typical” number of links for a node in a power law distribution, such networks are often termed “scale-free,” meaning roughly that there is no typical number of connections or “scale” that can be used to describe the network. In such a scale-free situation, knowing, for example, that the average number of links per node is 12 is not useful for predicting the collective behavior of the network because there are significant numbers of nodes with no links and significant numbers of nodes with hundreds or thousands of links.<sup>25</sup> Decisions, plans, or predictions based on the average node degree can go badly awry. An example of this kind of situation is the attempt to estimate the spread of a virus, such as HIV. While the average number of sexual partners per person may be quite small, if, as appears to be the case, the distribution is sufficiently broad (effectively scale-free), there are a small but significant number of individuals who have hundreds of sexual partners.<sup>26</sup> Any predictions of how fast the virus will spread based on the average number of sexual partners will be completely wrong.<sup>27</sup> Moreover, a program of random condom distribution or HIV testing may be much less effective than anticipated if it happens to miss the few most highly active individuals.

Highly skewed and broad degree distributions are common in networks of widely varying types, including the physical connections and hardware of the internet, the hyperlinks of the World Wide Web, the networks of co-authorship of high energy physics papers and of co-appearances of movie actors, networks of sexual contacts, metabolic networks, phone call networks, networks of word uses, protein interaction networks, and many others.<sup>28</sup> Other observed networks, such as the power grid, do not have a scale-free or nearly scale-free form, however. Thus, the

---

25. More precisely, a “scale-free” network is one in which the large degree tail of the node degree distribution decays as a power of the degree,  $P(k) \sim k^{-\alpha}$ , rather than exponentially with the degree, as more well known probability distributions do. Where the degree distribution decays exponentially, it has a well-defined “scale” given by the average number of links. The term “scale-free” has its roots in the physics of critical phenomena, in which scale referred to the typical length scale of correlations between atoms or magnetic spins. At a critical point, the correlations become “scale-free”, meaning that they occur on all length scales and hence no characteristic length scale can be identified. Just as the analysis of social issues becomes more difficult when no typical behavior can be identified, the physics of critical phenomena is difficult because there is no typical length scale. Technically, the term “scale-free” should be reserved for networks in which the node degree distribution has a power law decay. In reality, it is sometimes difficult to tell a true “power law” decay from a very broad, skewed distribution with an exponentially decaying tail.

26. Ayres & Baker, *supra* note 6, at 607-11.

27. *Id.*

28. See Albert & Barabasi, *supra* note 2; Newman, *supra* note 2.

degree distribution is a metric which can be used to categorize networks and to differentiate between them with respect to how they may behave. In the legal context, knowing how networks—such as drug distribution networks (legal and illegal), communications networks, infectious networks, terrorist networks, and the social networks of ordinary citizens—tend to be organized may be very important in devising effective policy and law.

b) Building Bridges: The Betweenness Metric

Node degree is only one possible way to measure the importance of a network node. Depending on the circumstances, the number of connections a node has may be less important than its specific place in the network structure. A node with few links may play a critical part in connecting different parts of a network. The friend who moved to New York may not have been the most popular person in the Chicago social network, but she may be the only connection to New York which the social group has. In that case, she will have a relatively low node degree but a relatively high “betweenness centrality.”<sup>29</sup> The betweenness centrality of a particular node can be measured by counting the number of paths between two points in the network which pass through that node. Chicago’s O’Hare airport, for example, has a high betweenness centrality because many flight paths between two United States cities pass through that airport. For some purposes, such as communicating between disparate parts of a network, for example, a network may be more dependent on high “betweenness” nodes than on those with high node degree. Of course, often, as in the case of O’Hare airport, nodes with high degree centrality also have high betweenness centrality.

c) Six Degrees of Separation: The “Small World” Property

Another commonly observed attribute of real world networks is the “small world” property, in which a relatively small number of “hops” between nodes is needed to connect any two nodes in the network.<sup>30</sup> The small world property can be important in determining the rate at which something—which might be a fad, a piece of information, or a deadly virus—spreads across the network. The significance of the small world property depends on the particular network. In general, it determines how tightly connected a network is. The small world property is often related to the presence of highly connected “hubs” in the network, which connect

---

29. See, e.g., Carrington, Scott & Wasserman, *supra* note 2, at 65-66 (defining this and other network metrics).

30. *Id.*

otherwise widely separated nodes (as in the sexual contact example already mentioned).

This is not always the case, however. A random network with a degree distribution like that shown in Figure 2 can have the small world property, as can a regular network (think of a fishnet, for example), after a small amount of re-wiring to connect distant parts of the network.<sup>31</sup> Re-wiring occurs in a social network, for example, whenever one person moves to a distant location and maintains social contacts in both locations. The person who moves from Chicago to New York broadens not only her own choice of friends, but also the choices of all her friends and relations who may meet New Yorkers through her.

The presence or absence of the small world property is thus an additional metric which one may use to categorize observed networks and to understand their behavior. There are various ways to measure the small world property, which have somewhat different significance. The most common measure is to compute, for each pair of nodes in the network, the minimum path length (number of hops needed to get from one node to the other), and then to identify the largest of this set of lengths as the “diameter” of the network.<sup>32</sup> The comparison between this network diameter and the size of the network determines whether the network has the small world property.

#### d) Transitivity and Local Network Structure

The small world property is one example of the ways in which network science probes beyond both average properties of individual nodes (such as node degree) and overall properties of the network (such as its

---

31. See Duncan J. Watts & Steven H. Strogatz, *Collective Dynamics of “Small World” Networks*, 393 NATURE 440 (1998) (discussing the basic properties of small world networks).

32. Some technical issues arise here. The term “small world” is used in two slightly different senses: to denote the property of short “distance” between nodes and to denote that property joined with the additional element of local clustering (more technically, high “transitivity”). By the “small world property” here we will mean short “distances” between nodes. By short, we mean, technically, that the longest distance between two nodes grows no more than logarithmically with the number of nodes. Intuitively, this means that it is impossible to make a spatial “map” of the network in which nodes that are connected to one another are near to one another. You will always end up with some connections between “far apart” nodes. An additional technical issue arises because many networks are not completely connected—there are isolated nodes or small clusters of nodes. Path lengths between these isolates and the bulk of the network are technically infinite. A network is still considered to have the small world property if it contains a so-called “giant cluster” which contains the vast majority of nodes and if the giant cluster has the small world property.

size) to investigate network structure. Another important measure of network structure is “transitivity” (often called “clustering coefficient” in the physics literature).<sup>33</sup> The transitivity of a node can be measured by determining the fraction of its network neighbors which are connected to one another. In the social context, a node with high transitivity is an individual whose friends are friends. Such individuals tend to be part of a densely connected social network. Often these multiply connected groups are connected by relatively strong ties of friendship or kinship. A node with low transitivity is an individual with friends who do not know one another. These individuals generally have more acquaintances (as opposed to close friends) and often their social contacts span a wider social context.<sup>34</sup> In other contexts, transitivity has different significance. In a citation network, for example, the transitivity of a particular scientific article may indicate its breadth or degree of interdisciplinarity. An article that cites or is cited by articles that cite each other will tend to have a narrower focus than an article that cites or is cited by articles that do not cite one another (and thus tend to be less closely related). Transitivity distinguishes nodes of the same degree. A highly cited article may be important in a single “hot” field or have more general significance. A person may be very popular within a close-knit group or may be more widely “connected.” Transitivity is thus an important means of probing the local structure of a network.

The above examples give a brief glimpse of the ways networks can be characterized by their patterns of node relationships and of how the network structure can affect the collective behavior of the network. The next Section discusses how the structure of a network is related to the way in which it evolves.

## 2. *How Does Your Network Grow? “Preferential Attachment” and Other Factors in Network Evolution*

Degree distributions, small world properties and other measures are useful in describing and categorizing real world networks. The aim of this descriptive task is not simply the “gee-whiz” satisfaction of seeing similar graphs show up for widely differing networks (though physicists, at least, do derive an inordinate amount of pleasure from such observations of universality). The underlying intuition driving the descriptive enterprise is

---

33. See Carrington, Scott & Wasserman, *supra* note 2, at 164-67; Albert & Barabasi, *supra* note 2, at 49; Newman, *supra* note 2, at 183 (giving definitions and discussing transitivity (also known as “clustering coefficient”).

34. See Mark Granovetter, *The Strength of Weak Ties: A Network Theory Revisited*, 1 SOC. THEORY 201, 218 (1983) (discussing the relationship between transitivity and “strong” and “weak” social ties).

that similar structures probably result from similar mechanisms of network growth and evolution, and that networks with similar structures will behave in similar ways. Making the connection between the mechanisms by which a network evolved and the network's eventual structure is the second task of network science.

In many cases, for example, a scale-free degree distribution results from what is called the "preferential attachment" mechanism.<sup>35</sup> Preferential attachment (also called the "rich get richer" effect) describes a process for creating a network in which "popular" nodes, which already have many links, are more likely than others to gain additional links as nodes are added to the network. It can lead to a highly heterogeneous, scale-free degree distribution in which some nodes eventually acquire a very large number of links while others remain relatively unconnected.

Why does this happen? In some cases, the nodes of the network may have a highly skewed distribution of "quality" which becomes reflected in the network connectivity. In a social network sociable people attract friends; in scientific journal citation networks the most important seminal papers are increasingly cited; in transportation networks "all roads lead to Rome" and so forth. Probably more often, however, the very fact that nodes are highly linked may make them more "linkable"—it is useful to fly through Chicago's O'Hare airport because so many other connecting flights already do; there may be reputational benefits to associating with popular people; popular sites garner more links on the World Wide Web.<sup>36</sup> Once one identifies the concept of preferential attachment, it can provide insights into the growth of many different kinds of networks. While the particular reasons that well-connected nodes acquire additional links cer-

---

35. See Albert & Barabasi, *supra* note 2, at 71; Newman, *supra* note 2, at 212-18.

36. The preferential attachment mechanism is related to the economic concept of "network effects" which explains, for example, why a lesser quality technology (such as VHS rather than Betamax for video recorders) can become entrenched. See, for a definition, S. J. Liebowitz & Stephen E. Margolis, *Network Externalities (Effects)*, in *THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW* 671-75 (Peter Newman ed., 1998). Network effects can be a source of preferential attachment. People choose a particular technology because others have chosen it. "Network effects" are important causes of market failure and sub-optimal social outcomes. More generally, preferential attachment can lead to highly skewed outcomes even when attachments are selected entirely at random in the first instance. Disentangling pure preferential attachment ("the rich get richer") from the effects of underlying node quality ("the talented or hard-working get richer") is a difficult and open question in network science. Understanding the reasons for preferential attachment in a particular context may be very important in predicting or understanding the network's collective behavior.

tainly vary from network to network, the general phenomenon of preferential attachment is widespread.

Though preferential attachment is common, in many real situations it is curtailed by cost or congestion. The capacity of a power hub limits the number of substations to which it can connect; the capacity of an airport limits the number of flights it can handle; popular people may run out of time or otherwise wish to limit their numbers of friends. Measuring the degree distribution of a network can thus lead to insights into how the network evolved. If the network is scale-free or very broad, one may search for some mechanism for preferential attachment. Similarly, if the number of connections cuts off at some point, a hunt for a congestion or cost mechanism may be warranted.

Citation networks, such as the patent network discussed in Part III of this Article or a network of scientific journal citations, tend to have very similar degree distributions, which are broad and skewed, but differ from the pure preferential attachment scale-free form. Figure 1 shows the distribution of number of citations received for the patent citation network at different times. It is quite broad—there are many patents that have never been cited, but others that have received nearly 1000 citations. To look for possible scale-free behavior, we plot the distribution function on a “log-log” graph. On such a graph, a power-law decay would show up as a downward sloping straight line. (This is illustrated in Figure 4 at the bottom, which shows the power law distributions on a log-log graph.) The patent citation distribution is clearly quite broad and skewed, but the tail of the distribution is not quite a straight line on the graph. Instead, it curves downward, cutting off the probability of finding a patent that has received an extremely high number of citations.

While the precise mathematical form of these citation network distributions is not completely understood, our results, discussed in more detail in Part III, and those of others<sup>37</sup> suggest that the shape reflects a competition between preferential attachment—which leads to some patents acquiring very large numbers of citations—and the rate at which patents age or go out of date. The preferential attachment leads to a broad, skewed distribution, while the aging of nodes slows their accumulation of citations and cuts off the tail of the degree distribution. If a similar distribution is observed in some new network, it will be reasonable to predict that the

---

37. See, e.g., S. N. Dorogovtsev & J.F.F. Mendes, *Evolution of Networks with Aging of Sites*, 62 PHYSICAL REV. E 1842 (2000); Han Zhu, Xinran Wang & Jian-Yang Zhu, *Effect of Aging on Network Structure*, 68 PHYSICAL REV. E 056121 (2003).

growth of that network involved some mechanism for preferential attachment and for aging.

As more networks are characterized using a variety of metrics, it becomes possible to infer much about the means by which a particular network evolved (or is evolving) by measuring aspects of the network's current structure. It may conversely be possible to predict many aspects of the network structure that will result from particular means of network growth.

### 3. *Network Structure and Collective Behavior*

The third task of network science is to understand the kinds of collective behavior that may emerge when elements of a network interact with one another and how the resulting behavior may depend upon the structure of the network. For example, the flow of information (or the progress of an epidemic) within a social group, the flow of oil through the ground, the susceptibility of an electrical or communication grid to the failure of individual elements, the cost of licensing the necessary patents to commercialize a particular technology, and the propagation and acceptance of a changing behavioral norm may all depend on the relational structure of an underlying network.<sup>38</sup> Network science will attempt to delineate and interpret the common features of network structures and interactions between nodes that result in particular varieties of collective responses. Such understandings may be used to predict the collective effects of changes in network structure, changes in the interactions between neighboring elements, and changing global influences—including changes in legal rules.

#### **B. Why Legal Scholars Should Care About Network Science**

In some sense, legal scholarship, in its descriptive form, *is* network science—the study of how particular social, commercial, infrastructural, and other kinds of networks adapt and react to particular rules of interaction between nodes that might be individuals, corporations, communities, “things,” or institutions. Moreover, the common law itself is a growing network of precedent, as reflected in citations in judicial opinions.<sup>39</sup> More than this, though, normative legal scholarship, legal policymaking, statutory legislation, and common law jurisprudence are in some respects network *engineering*—attempts to produce particular collective social re-

---

38. See BARABASI, *supra* note 2, WATTS, *supra* note 2, and MALCOLM GLADWELL, *THE TIPPING POINT* (2000) for examples of situations in which the local structure of a network can determine the global behavior.

39. See Post & Eisen, *supra* note 6, Chandler, *supra* note 6, and Smith, *supra* note 6 for studies of the network of legal citations.

sponses by adjusting the interactions and incentives experienced by individual network elements. Of course, this observation is important (and not merely semantic) only if thinking about society as a network matters—only if it tells us something different from what we have already gained from our existing conceptual frameworks.

There are good reasons to think that network analysis will matter in this way—and that it may matter a lot. Here we discuss three inter-related concepts highlighted by network science which may have broad impact on legal and policy analysis: (1) node heterogeneity, (2) the importance of network relational structure, and (3) the complicated relationship between local structure and interactions and global, collective behavior.

### 1. *The Implications of Heterogeneity*

Much of our legal doctrine and scholarship—including much law and economics analysis—depends implicitly, but heavily, on assumptions and predictions about average or typical behavior. Thus, in legal decision-making we rely on the “reasonable person” and the “person having ordinary skill in the art;” we seek to deter the “typical” criminal; and we attempt to avoid confusing the “ordinary consumer.” We predict the results of policy changes by considering the rational economic actor (or, perhaps, the “boundedly rational” economic actor).<sup>40</sup> All of these modes of analysis rely on an intuitive assumption that various traits, propensities, preferences, and so forth are distributed among the population according to a more or less normal distribution. In other words, we assume that the average behavior is the most typical behavior and that extreme deviations from the typical are extremely rare—and thus generally unimportant.

This assumption is reasonable for many purposes since many properties are distributed according to a normal distribution—there are no hundred foot tall or one inch tall human beings, for example. However, network science provides a warning about unquestioning reliance on the typicality assumption. It points out that there are common examples of quantities—such as the number of links to a node—that are distributed in a highly skewed and very broad manner—indeed, scale-free or close to it. While we have always known that such broad, skewed distributions were possible, the study of networks and their societal ubiquity brings us face-to-face with the fact that highly skewed, scale-free or nearly scale-free,

---

40. See, e.g., Christine Jolls, Cass R. Sunstein & Richard H. Thaler, *A Behavioral Approach to Law and Economics*, in *BEHAVIORAL LAW AND ECONOMICS* 13, 14-15 (Cass Sunstein ed., 2000) (providing an overview of the significance of bounded rationality).

distributions are not only possible, but are common in some contexts and that taking account of the non-typical may be crucial.<sup>41</sup>

As discussed above, if a network is scale-free or nearly so, it is meaningless to talk about a “typical” node because the average node is not the most typical node—and the most important nodes are likely neither average nor typical. As a result, social interventions, including legal rules, which are aimed at the “typical” member of a social group may be ineffective—it may be necessary to consider a broad spectrum of characteristics to devise a workable rule. Ayres and Baker, for example, have highlighted the importance of the extreme heterogeneity of sexual behavior to designing legal means to deal with the threat of HIV-AIDS.<sup>42</sup>

Thus, the first lesson from network science is that legal analysis should consider the possibility of heterogeneity—in some cases radical heterogeneity—among network elements and environments. Indeed, the realization that a simple and common mechanism like preferential attachment in a network can result in radical heterogeneity should motivate us to consider what other commonplace deviations from “typicality” may arise.<sup>43</sup> Perhaps deviations from the normal distribution are not so abnormal after all.

Currently, there is no complete fundamental understanding of when scale-free distributions occur. Intuitively, however, it seems likely that these distributions are possible when scale is not strongly constrained by some form of cost.<sup>44</sup> For example, the network of World Wide Web hyper-

---

41. See CHRIS ANDERSON, *THE LONG TAIL: WHY THE FUTURE OF BUSINESS IS SELLING LESS OF MORE* (2006) (discussing the prospective effects of consumer heterogeneity on business).

42. See Ayres & Baker, *supra* note 6, at 630-58 (arguing in favor of a crime of “reckless sex” based on the importance of targeting one-time sexual encounters to prevent the spread of sexually transmitted diseases, in part because of the highly skewed distribution of numbers of sexual partners).

43. The statistical physics interest in scale-free distributions, for example, pre-dates network theory and stems from the observation of scale-free behavior at phase transitions and in observations of self-organized criticality. See, e.g., PER BAK, *HOW NATURE WORKS: THE SCIENCE OF SELF-ORGANIZED CRITICALITY* (1996); HENRIK JELDTOFT JENSEN, *SELF-ORGANIZED CRITICALITY: EMERGENT COMPLEX BEHAVIOR IN PHYSICAL AND BIOLOGICAL SYSTEMS*, Cambridge Lecture Notes in Physics (No. 10) (1998); SHANG-KENG MA, *MODERN THEORY OF CRITICAL PHENOMENA* (1976); H. EUGENE STANLEY, *INTRODUCTION TO PHASE TRANSITIONS AND CRITICAL PHENOMENA* (1971); P.C. Hohenberg & B.I. Halperin, *Theory of Dynamic Critical Phenomena*, 49 REV. OF MODERN PHYSICS 435 (1977); H. Eugene Stanley, *Scaling, Universality, and Renormalization: Three Pillars of Modern Critical Phenomena*, 71 REV. MODERN PHYSICS S358 (1999).

44. In statistical physics, for example, interest in scale-free phenomena arose in the study of critical phenomena. Critical points occur when energy and entropy balance in such a way that fluctuations on all length scales have equal free energy and thus all occur

links (which is relatively unconstrained by linkage costs) is approximately scale-free, whereas the railway system (which is constrained by the costs of building additional track “links”) and the power grid (which is constrained by the cost of wiring and the loss of energy in transportation) are not.<sup>45</sup> An observation of extreme heterogeneity should trigger an inquiry not only into its effects but also into its causes. It is conceivable, for example, that there is a relation between the well-known highly skewed distribution of values of technological innovations and a fundamental lack of proportionality between research investment and research result. If this (highly speculative) premise were correct, it might have important implications for the use of patent law to provide incentives to invest. Similarly, a very skewed distribution of network connections which results from some particular combination of legal rules and other norms and constraints might reflect an underlying skewed distribution of talent or effort, but we should recognize that preferential attachment alone is enough to produce such a distribution even if all nodes are identical. Depending on whether the skewed distribution of connections is socially desirable, this recognition may affect the choices of policies to pursue.

## 2. *The Importance of Network Structure in Determining Individual Responses to Legal and Social Change*

A second lesson from network science is the importance of local network context in determining individual responses to legal and social change. Legal analysis often conceptualizes individual legal actors as responding independently to legal rules in the context of global average social forces.<sup>46</sup> Network science demonstrates that individual responses to legal intervention or changing social norms may be determined not only by the average impact of global social forces but also by the specific network structures by which these social forces are mediated in a particular case. Social network studies demonstrate that access to information de-

---

at once. Away from critical points, fluctuations in physical systems have a characteristic scale. See, e.g., DIDIER SORNETTE, *CRITICAL PHENOMENA IN NATURAL SCIENCES: CHAOS, FRACTALS, SELF-ORGANIZATION AND DISORDER: CONCEPTS AND TOOLS* 241-42 (2d ed. 2004).

45. See Newman, *supra* note 2, at 178-79.

46. This is the general approach of social norm theory. See, for example, Richard H. McAdams & Eric B. Rasmusen, *Norms in Law and Economics*, in *HANDBOOK OF LAW AND ECONOMICS* (A. Mitchell Polinsky & Steven Shavell eds., forthcoming 2007) for a recent review. A current draft of the chapter is available at <http://www.rasmusen.org/papers/norms.pdf>.

depends on one's position in the network.<sup>47</sup> An individual's susceptibility to legal and social sanctions for socially undesirable behavior will be similarly dependent on his or her local social network.

Rather than thinking of an individual person or firm as being immersed in an average social ether of information and influences, it may sometimes be important to take into account the specific local relationships in which particular individuals are embedded. Dickerson, for example, discusses how the network structure within corporations and the distinctions between the connection environments of particular nodes has been implicitly, and might be explicitly, used to effect corporate change by targeting particular corporate actors.<sup>48</sup> The realization that a model in which an individual particle is subjected to the average influence of its fellows cannot always predict behavior was important in statistical physics—it led to the rejection of mean field theory and inaugurated the modern study of collective physical phenomena.<sup>49</sup> The statistical physics approach to network science stems from this background understanding, which may be quite important for legal analysis as well.<sup>50</sup>

### 3. *The Complicated Influence of Local Network Structure on Collective Behavior*

A corollary of the premise that individual behavior may be significantly affected by local network structure rather than simply by average social forces is that one cannot simply predict collective behaviors, such as social norms and behavioral regularities, the collective impact of legal

---

47. Similarly, statistical physics studies show that a randomly constructed network may make a transition from a set of disconnected clusters to a single connected network at a particular critical density of links. A scale-free network may or may not experience such a transition depending upon the precise form of the degree distribution and other factors. See Newman, *supra* note 2, at 225-28.

48. See Dickerson, *supra* note 6, at 556-67.

49. In statistical physics, the approximation that an individual atom reacts to the average influence of the other atoms is called "mean field theory." While mean field theory is very useful in some situations, it is particularly bad at explaining "phase transitions" (such as melting and freezing, evaporation, and magnetization), in which the collective behavior of a system changes from one macroscopic state to another. Part of the reason for the failings of mean field theory is its inability to handle the "scale-free" correlations between blocks of atoms of widely varying sizes that appear near many phase transitions. See, e.g., MA, *supra* note 43, at 34-39 (discussing mean field theory and its shortcomings in the phase transition context).

50. See Albert & Barabasi, *supra* note 2, at 91-92 (discussing the implications of local correlations for network structure); Newman, *supra* note 2, at 224-40; see also PHILIP BALL, CRITICAL MASS: HOW ONE THING LEADS TO ANOTHER (2004) (discussing, for a general audience, the application of statistical physics methods to social problems).

change, and flows of information, influence, and other goods by looking at local relationships and individual transactions.

Optimal collective behavior, for example, may depend not just on individual relationships but on the particular network structure in which those relationships are embedded. A simple example shown in Figure 5 illustrates the point. Assume that each dot represents an individual and that the associated + or - represents the individual's yes or no decision about some question. Also assume that each individual must make a decision and prefers, for some reason, to make the opposite decision from his or her immediate neighbors. (In physics this is known as the anti-ferromagnetic Ising model.<sup>51</sup>) If the network of relationships between individuals forms a square pattern, as shown in Figure 5 at the top, every pair can be made happy by making opposing decisions, and the global arrangement will reflect the optimization of the pairwise interactions. However, if the network of relationships between individuals forms a triangular pattern, as shown in the bottom panel of Figure 5, it is simply not possible for everyone to disagree with all of his or her neighbors. Even though there is a preferred arrangement for every pair of neighbors (disagreement), the network as a whole will be unable to get to a completely “disagreeable” state. The network will be, in the terminology used by statistical physicists, frustrated.<sup>52</sup> The best global arrangement (the one with the maximum number of local disagreements) is not predictable by scrutinizing the preferences of each pair in isolation. The collective behavior that results from the relationship pattern shown at the top of Figure 5 differs entirely from the collective behavior that results from the pattern shown at the bottom of Figure 5, even though the local interactions and preferences are identical.

Figure 5 is just an illustration of the more general point that one cannot always predict global behavior from interactions between pairs or in small groups. In the legal context, this insight suggests that focusing only on the efficiency of individual transactions between individual legal actors may, in some cases, lead to suboptimal legal policies because these actors are embedded in a network of relationships. Yet legal analysis and implementation frequently does focus on such individual transactions, either because legal rules are forged in the context of specific disputes or because, as a theoretical matter, generalizing beyond these individual transactions is difficult.<sup>53</sup>

---

51. SORNETTE, *supra* note 44, at 441-43.

52. See, for example, *id.* at 441-43, for a discussion of this example of frustration.

53. For example, applications of game theory to legal theory very frequently focus on analyzing a two-person or, at most, three-person game and then extrapolating to form predictions about collective behavior. See generally, DOUGLAS G. BAIRD ET AL., GAME

As another example of the importance of local network structure for collective behavior, diffusion of anything on a network—whether it be information, opinion, resources, or infection—may be a complicated function of the number and placement of network links and not simply of the average likelihood of transmittal from one node to its neighbors. Depending on the specifics of relationships, diffusion of information from one social group to another may be rapid or extremely slow.<sup>54</sup> Lior Jacob Strahilevitz, applying these ideas, argues that analysis of the privacy tort of “public disclosure of private facts” should rely on the network concepts of weak and strong ties to distinguish between information that likely would have spread without the challenged disclosure and information that likely would not.<sup>55</sup>

Merely recognizing these conceptual points may sometimes be enough to provide insight into an important legal question.<sup>56</sup> However, network science’s potential usefulness to legal analysis extends beyond highlighting these general concepts. Indeed, the observations that important social and economic quantities may be distributed in a highly skewed, scale-free manner and that collective behavior may not be accurately predictable by focusing either on average global influences or on individual transactions are not entirely new. Legal scholars and decision makers have employed the approximations of typicality, of average social forces, or of the generalizability of pairwise interactions in the past not because they were unaware that reality is more complicated but to reduce the complexity of the analysis.

Network science can contribute in that it not only highlights these conceptual issues, but also promises to provide ways to determine when these issues are likely to be important and tools to perform more accurate analyses in various contexts. These tools may be analytical, such as the introduction of explanatory concepts like “preferential attachment” and “small

---

THEORY AND THE LAW 33 (1994) and ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS (4th ed. 2003) for examples of applying game theory to legal issues. Probably the most famous example of this approach in the law is the classic Prisoner’s Dilemma game (cited in more than 2000 articles in the LEXIS law review database), which commentators often use to explain problems of collective action and free riding. *But see* Tom Ginsburg & Richard H. McAdams, 45 WM. & MARY L. REV. 1229 (2004) (arguing, in the context of international relations, that the Prisoner’s Dilemma game is overused in comparison to games involving motives for cooperation).

54. *See* Strahilevitz, *supra* note 6, at 988 (arguing for a view of privacy that takes into account, based on social theory, the likelihood that information will diffuse out of a limited group to which it has been revealed).

55. *Id.* at 948-73 (discussing studies of information propagation in social networks).

56. *See* Ayres & Baker, *supra* note 6, at 607-18; Strahilevitz, *supra* note 6, at 983.

worlds.” They may be mathematical, as illustrated by the analysis of patent citations in Part III of this Article. Or they may be computational, as exemplified by many social network theory software tools or by the computer simulation methods of statistical physics.<sup>57</sup>

Network science concepts may provide a basis for critiquing extant or proposed legal rules.<sup>58</sup> Network science methods may also provide, as we begin to demonstrate in Parts III and IV of this Article, means of empirical analysis that are applicable to legally significant systems in which heterogeneity and network structure are important. Network science techniques may also help to model the collective behavior that results from particular local interactions and preferences and to predict whether legal changes will lead to global behavioral change. For example, one may begin by postulating particular pairwise interactions, such as those embodied in many law and economics game theory models. Network science methods may then be employed to determine the collective behavior that results from such pairwise interactions.<sup>59</sup> Computer simulations will often be the most

---

57. See, e.g., Carrington, Scott & Wasserman, *supra* note 2 (describing social network methodologies); Harvey Gould, Jan Tobochnik & Wolfgang Christian, Simulations in Physics, <http://sip.clarku.edu> (last visited Nov. 12, 2006) (providing resources and links about simulation methods in physics); International Network for Social Network Analysis, <http://www.insna.org> (last visited Nov. 12, 2006) (providing descriptions of and links to numerous social network software packages).

58. See, e.g., Ayres & Baker, *supra* note 6, at 607-18; Matwyshyn, *Of Nodes and Power Laws*, *supra* note 6; Strahilevitz, *supra* note 6, at 988.

59. While the Prisoner's Dilemma paints a rather gloomy, non-cooperative view of the social world, simulations of iterated prisoners' dilemmas on networks of various structures with various strategic rules are beginning to offer a much more complex and varied view of the possible cooperative and non-cooperative outcomes of the game. See generally Guillermo Abramson & Marcelo Kuperman, *Social Games in a Social Network*, 63 PHYS. REV. E 030901 (2001); Petter Holme, Ala Trusina, Beom Jun Kim & Petter Minnhagen, *Prisoners' Dilemma in Real-World Acquaintance Networks: Spikes and Quasiequilibria Induced by the Interplay Between Structure and Dynamics*, 68 PHYS. REV. E 030901 (2003); Beom Jun Kim, Ala Trusina, Petter Holme, Petter Minnhagen, Jean S. Chung & M.Y. Choi, *Dynamic Instabilities Induced By Asymmetric Influence: Prisoners' Dilemma Game in Small-World Networks*, 66 PHYS. REV. E 021907 (2002); Y.F. Lim, Kan Chen & C. Jayaprakash, *Scale-Invariant Behavior in a Spatial Game of Prisoners' Dilemma*, 65 PHYS. REV. E 026134 (2002); Zhi-Xi Wu, Xin-Jian Xu, Yong Chen & Ying-Hai Wang, *Spatial Prisoner's Dilemma Game with Volunteering in Newman-Watts Small-World Networks*, 71 PHYS. REV. E 037103 (2005). Other strategic games are also being explored on networks using statistical physics approaches. See, e.g., J.C. González-Avella, M.G. Cosenza & K. Tucci, *Nonequilibrium Transition Induced by Mass Media in a Model for Social Influence* (May 22, 2005), [http://arxiv.org/PS\\_cache/nlin/pdf/0511/05111013.pdf](http://arxiv.org/PS_cache/nlin/pdf/0511/05111013.pdf); Ping-Ping Li, Da-Fang Zheng & P.M. Hui, *Dynamics of Opinion Formation in a Small-World Network* (Oct. 10, 2005), arXiv:physics/0510065

feasible approach to such questions and can in principle be devised to account for heterogeneous preferences and for various network structures.

Just as has been done with game theory models, it may well be possible to map some problems of legal significance onto models that physicists, sociologists, and other network scientists have already studied. Additionally, legal problems may eventually motivate network scientists to perform computer simulations or mathematical studies of models derived from those problems.

In sum, network science is an emerging discipline that holds great promise to provide insights, tools, and models that will make important contributions to legal analysis. In the next Part we provide a sample of the application of network science to law by investigating what the evolution of the patent citation network can tell us about how the standard of patentability is changing.

### III. NETWORK CLUES TO A DECREASING PATENTABILITY STANDARD

We now move from the lofty realms of possibility to a specific application of network science in the legal arena. We have performed a quantitative analysis of United States patents and their citations, treating the patents as nodes and citations from one patent to another as network links. Our analysis suggests that the “patent explosion” of recent years is not completely explained by either a rapidly increasing pace of technological advance or a broadened scope of patented technology. Something more has happened in the network—an increasing stratification of patent “citability,”<sup>60</sup> by which we mean the probability that a patent with particular characteristics will be cited. In this Part, we describe how our network science approach reveals this dynamic change and hypothesize that it may result from a decreasing patentability standard. We distinguish carefully between our empirical results—which any theory of the way in which patents and the citations between them have evolved will need to explain—and our proposed interpretation. A brief technical report focusing on the

---

v1; Sitabhra Sinha & Sudeshna Sinha, *Robust Emergent Activity in Dynamical Networks* (Oct. 22, 2005), [http://arxiv.org/PS\\_cache/condmat/pdf/0510/0510603.pdf](http://arxiv.org/PS_cache/condmat/pdf/0510/0510603.pdf).

60. Citability is defined above as the probability that a patent will be cited again depending on its current age and number of previous citations.

methodological contributions of this research to network science has been accepted for publication in the physics journal, *Physica A*.<sup>61</sup>

## A. Background

### 1. *The Patent System and Its Discontents*<sup>62</sup>

Recent years have seen a major upsurge in patenting, an expansion of the range of innovations which are eligible for patent protection, and a perception that the United States economy relies more and more heavily on knowledge and innovation for its success.<sup>63</sup> At the same time, developments in the law, including the establishment of a single appellate court—the U.S. Court of Appeals for the Federal Circuit—to hear the vast majority of patent appeals in the United States, have led to debate as to whether the legal system is becoming increasingly patent-friendly; whether patents are being issued for lower quality innovations; and whether the legal rights awarded to patentees are becoming stronger.<sup>64</sup>

---

61. Gábor Csárdi, Katherine Strandburg, László Zalányi, Jan Tobochnik & Péter Érdi, *Modeling Innovation by a Kinetic Description of the Patent Citation System*, PHYSICA A (forthcoming 2006).

62. The title of this Section is borrowed from a recent book by two well-known economists of innovation which exemplifies the kinds of criticisms of the patent system which have become prevalent in recent years. See JAFFE & LERNER, *supra* note 7.

63. See *infra* Appendix, Figure 6.

64. For a sample of the many discussions of these issues see, for example, FEDERAL TRADE COMMISSION, *supra* note 7; Dan L. Burk & Mark A. Lemley, *Designing Optimal Software Patents*, in INTELLECTUAL PROPERTY RIGHTS IN FRONTIER INDUSTRIES: SOFTWARE AND BIOTECHNOLOGY 81 (Robert Hahn, ed., 2005); JAFFE & LERNER, *supra* note 7; NATIONAL RESEARCH COUNCIL OF THE NAT'L ACADEMIES, PATENTS IN THE KNOWLEDGE-BASED ECONOMY, (Wesley M. Cohen & Stephen A. Merrill eds., 2003); Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in INNOVATION POLICY AND THE ECONOMY 119 (Adam Jaffe, Josh Lerner & Scott Stern eds., 2001); Merrill, Levin & Myers, *supra* note 7; David E. Adelman, *A Fallacy of the Commons in Biotech Patent Policy*, 20 BERKELEY TECH. L.J. 985 (2005); John R. Allison & Mark A. Lemley, *The Growing Complexity of the United States Patent System*, 82 B.U. L. REV. 77 (2002); Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575 (2003); Rochelle Cooper Dreyfuss, *The Federal Circuit: A Continuing Experiment in Specialization*, 54 CASE W. RES. L. REV. 769 (2004); Rochelle Cooper Dreyfuss, *The Federal Circuit: A Case Study in Specialized Courts*, 64 N.Y.U. L. REV. 1 (1989); Bronwyn H. Hall, *Exploring the Patent Explosion*, 30 J. TECH. TRANSFER 35 (2005); Heller & Eisenberg, *supra* note 7; Jay P. Kesan, *Carrots and Sticks to Create a Better Patent System*, 17 BERKELEY TECH. L.J. 763 (2002); William M. Landes & Richard A. Posner, *An Empirical Analysis of the Patent Court*, 71 U. CHI. L. REV. 111 (2004); Allan N. Littman, *Restoring the Balance of Our Patent System*, 37 IDEA 545 (1997); Glynn S. Lunney, Jr., *E-Obviousness*, 7 MICH. TELECOMM. & TECH. L. REV. 363 (2001); Ronald J. Mann, *Do Patents Facilitate Financing in the Software Industry?*, 83 TEX. L. REV. 961 (2005); Robert P. Merges, *As Many as Six Impossible Patents Before*

Empirical evidence increasingly raises questions as to the extent to which patents are needed to provide incentives for research and development.<sup>65</sup> These trends have converged to raise growing concerns among academics and policymakers about whether patent law and policy are adequately designed to “promote the Progress of . . . useful Arts.”<sup>66</sup> This discontent has gained the attention of members of Congress, who have proposed various patent reform bills,<sup>67</sup> of the Federal Trade Commission,<sup>68</sup> of the National Academies of Sciences,<sup>69</sup> and of the Supreme Court, which has granted review in six patent cases since 2005<sup>70</sup>—a level of Supreme Court interest

---

*Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577 (1999); Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law, 1900-2000*, 88 CALIF. L. REV. 2187 (2000) [hereinafter Merges, *One Hundred Years*]; Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839 (1990); Arti K. Rai, *Allocating Power Over Fact-Finding in the Patent System*, 19 BERKELEY TECH. L.J. 907 (2004); Arti K. Rai, *Engaging Facts and Policy: A Multi-Institutional Approach to Patent System Reform*, 103 COLUM. L. REV. 1035 (2003); John R. Thomas, *Formalism at the Federal Circuit*, 52 AM. U. L. REV. 771 (2003); Adelman & DeAngelis, *supra* note 17; Christopher A. Cotropia, *Patent Law Viewed Through an Evidentiary Lens: The “Suggestion Test” as a Rule of Evidence* (Tulane Public Law Research Paper No. 06-03, Mar. 2006), <http://ssrn.com/abstract=893965>; Gregory N. Mandel, *Patently Non-Obvious: Empirical Demonstration that the Hindsight Bias Renders Patent Decisions Irrational* (2005), available at <http://www.ssrn.com/abstract=871684>; John L. Turner, *In Defense of the Patent Friendly Court Hypothesis: Theory and Evidence* (2004), <http://www.terry.uga.edu/~jltturner/PatentFCH.pdf>.

65. See, e.g., James Bessen & Michael J. Meurer, *Lessons for Patent Policy from Empirical Research on Patent Litigation*, 9 LEWIS & CLARK L. REV. 1, 6-8 (2005) (reviewing empirical evidence for the relatively small size of the “patent premium”); Bronwyn Hall & Rosemarie Ziedonis, *The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995*, 32 RAND J. ECON. 101 (2001); Wesley M. Cohen et al., *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)* (Nat’l Bureau of Econ. Research, Working Paper No. 7552, 2000).

66. U.S. CONST. art. I, § 8, cl. 8.

67. See, e.g., Patent Reform Act of 2005, H.R. 2795, 109th Cong. (2005); Patent Reform Act of 2006, S. 3818, 109th Cong. (2006).

68. FEDERAL TRADE COMMISSION, *supra* note 7.

69. Merrill, Levin & Myers, *supra* note 7.

70. See *Ill. Tool Works, Inc. v. Indep. Ink, Inc.*, 126 S.Ct. 1281 (2006); *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 126 S.Ct. 2921, 2921 (2006) (Breyer, J., dissenting); *eBay, Inc. v. MercExchange, L.L.C.*, 126 S.Ct. 1837, 1842 (2006) (Kennedy, J., concurring); *KSR Int’l Co. v. Teleflex, Inc.*, No. 04-1350 (U.S. argued Nov. 28, 2006); *Microsoft Corp. v. AT&T Corp.*, 414 F.3d 1366, *cert. granted*, 127 S. Ct. 467 (2006) (No. 05-1056); *Merck KGaA v. Integra Lifesciences I, Ltd.*, 545 U.S. 193 (2005).

unheard of in at least 25 years. It has also provoked responses from defenders of the present system, who argue that criticisms are overblown.<sup>71</sup>

To accomplish its constitutionally mandated objective, patent protection must be carefully tailored to balance its benefits against its costs. The benefits may include providing incentives to invent, functioning as a signal of technical competence, and facilitating a market for intangible knowledge. However, because a patent provides exclusive rights to practice the patented technology, patents impose costs on society that may include not only supra-competitive pricing of patented products but also increased barriers to building upon existing technology. These barriers arise because improving upon a patented technology may require either using the patented technology during development or incorporating it into the improved result. In either case, United States law generally requires that the improver obtain a license, usually requiring the payment of royalties, from the holder of a patent on the foundational technology or technologies. Negotiating authorization to build upon a patented technology can be expensive, especially when there is disagreement as to the relative value of various contributions to the improved technology. In the extreme case, when the holders of the original patent and an improvement patent cannot agree on a licensing arrangement, this “blocking patent” situation can deprive the public of access to the improved technology altogether.

Part of the patent cost-benefit analysis is a requirement that the USPTO issue patents only for inventions that are novel and nonobvious<sup>72</sup>—that is, the inventions differ sufficiently from presently available technology to justify an award of legal exclusivity to the inventor. The legal standard of nonobviousness sets the height of the bar for “sufficient” difference. Along with other patentability requirements, it determines the tradeoff between social costs and benefits that results from the issuance of a patent. Because patents can impose substantial costs on consumers and subsequent innovators, one major objective of studies of the patent system is to determine whether the procedures and substantive standards that guide the issuance of patents are appropriately tuned.

Recent developments in the law, along with the issuance of a burgeoning number of patents, have led to widespread perceptions among patent

---

71. See Katherine J. Strandburg & R. Polk Wagner, Debate, *The Obviousness Requirement in the Patent Law*, 155 U. PA. L. REV. 100 (2006) (debating the question of obviousness) [hereinafter Strandburg & Wagner, Debate]; see also Adelman, *supra* note 64; Adelman & DeAngelis, *supra* note 17; W. Lesser & Travis Lybbert, *Do Patents Come Too Easy?*, 44 IDEA 381 (2004); Cotropia, *supra* note 64; Mandel, *supra* note 64.

72. See 35 U.S.C. §§ 101-103 (2000 & Supp. IV 2004).

law scholars and policymakers that the system is off balance.<sup>73</sup> Many argue, for example, that the legal standard of nonobviousness is insufficiently rigorous;<sup>74</sup> that patent examiners have insufficient access to potentially relevant prior art in some fields such as software or business methods;<sup>75</sup> and that patents issue on increasingly upstream technologies that form the basis for further advances.<sup>76</sup> There are dire predictions of a patent thicket,<sup>77</sup> in which technological progress is made increasingly difficult by the need to negotiate multiple levels of blocking patent rights on each of the many patented components which may be needed to produce a new commercial product.<sup>78</sup> One way to avoid a potential thicket is for competing patent holders to negotiate cross-licenses or patent pools. Such agreements between competitors raise concerns about collusion, however, and the societal ramifications depend upon the extent to which cross-licensing lowers barriers to the use of complementary technologies, as opposed to allowing competitors to avoid competition from substitute technologies.<sup>79</sup>

Understanding the interaction between innovation and the patent system is difficult for many reasons. Increased patenting, for example, can stem from various causes, including an increased pace of technological change, an increased range of patented technology due either to expansion of the scope of legally patentable subject matter or to the birth of new fields of technology, a growing perception of the usefulness of patents as business tools, or the issuance of lower quality patents. Empirical investigation of the patent system can play an important role in understanding

---

73. See, e.g., sources cited *supra* note 7.

74. See, e.g., FEDERAL TRADE COMMISSION, *supra* note 7; Brief for Strandburg et al. as Amici Curiae Supporting Petitioner, *KSR Int'l Co. v. Teleflex, Inc.*, 75 U.S.L.W. 3094 (2006); Merrill, Levin & Myers, *supra* note 7. *But see* Cotropia, *supra* note 64, at 48-60 (arguing applications of the Federal Circuit "suggestion test" are tailored to the complexity of the underlying technology); Mandel, *supra* note 64 (reporting empirical evidence that hindsight bias is likely to result in unwarranted findings of obviousness).

75. See, e.g., JAFFE & LERNER, *supra* note 7, at 198-202.

76. See, e.g., FEDERAL TRADE COMMISSION, *supra* note 7 and sources cited therein.

77. See, e.g., Shapiro, *supra* note 64; James Bessen, *Patent Thickets: Strategic Patenting of Complex Technologies* (Research on Innovation and Boston University School of Law, Working paper, last visited Nov. 11, 2006), available at <http://www.researchoninnovation.org/thicket.pdf>. *But see* Adelman, *supra* note 64; Adelman & DeAngelis, *supra* note 17; Mann, *supra* note 64.

78. See, e.g., Robert C. Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62 TENN. L. REV. 75 (1994).

79. See, e.g., Gilbert Goller, *Competing, Complementary and Blocking Patents: Their Role in Determining Antitrust Violations in the Areas of Cross-Licensing, Patent Pooling and Package Licensing*, 50 J. PAT. OFF. SOC'Y 723 (1968); see also FEDERAL TRADE COMMISSION, *supra* note 7.

how to maintain the appropriate balance. In what follows, we report the results of a detailed study of the evolution of the patent citation network and argue that the evolution suggests that the patentability standard has been decreasing. Before presenting our results, we detour to provide an overview of patent issuance and the meaning of patent citations and discuss how patent data has been used in previous studies. With that context in place, we then describe our study.

## 2. *Patent Issuance and Prior Art Citations*

The United States Patent and Trademark Office (USPTO) issues patents after applications are examined to determine, among other things, whether the patent claims meet the legal requirements of novelty and non-obviousness.<sup>80</sup> Patent claims are specific statements of the scope of the legal coverage of a patent. As noted above, the legal effect of a patent is to provide the patentee a right to exclude others from using the claimed technology without a license, as detailed in the infringement provisions of the patent statute.<sup>81</sup>

In the course of the examination of a patent application for novelty and nonobviousness, patent claims are compared against potential prior art, consisting in large part of prior patents and other publications in relevant technical fields. Applicants, their patent attorneys, and the official patent examiners all identify potential prior art. To qualify for a patent, the claimed invention must be novel, meaning there is no prior patent or other prior art that is identical to what is claimed. More importantly, the claimed invention must be nonobvious, meaning that at the time it was invented, the invention would not have been obvious to a person having ordinary skill in the art in the field of the invention. When an invention is deemed obvious, it is usually because it is an obvious combination of prior art technology. Determining whether a combination of prior technology would have been obvious is a tricky matter, since it requires both an exercise in hindsight and putting oneself in the shoes of the skilled practitioner of the patented technology.<sup>82</sup> The way in which obviousness is determined essentially sets the threshold of patentability.

---

80. See 35 U.S.C. §§ 102-103, 112 (2000) for the main statutory requirements for patentability. *See also* MPEP, *supra* note 15.

81. *See* 35 U.S.C. § 271 (2000).

82. *See, e.g.*, Rebecca S. Eisenberg, *Obvious to Whom? Evaluating Inventions from the Perspective of the PHOSITA*, 19 BERKELEY TECH. L. J. 885 (2004); Cyril A. Soans, *Some Absurd Presumptions in Patent Cases*, 10 IDEA 433, 438 (1966) (coining the name “Mr. PHOSITA”); Mandel, *supra* note 64.

To guide this endeavor, and to try to avoid hindsight bias, the Federal Circuit employs a controversial test. Under this “suggestion test,” a patent examiner may not reject a patent application as an obvious combination of prior art elements unless a “suggestion, teaching, or motivation to combine” the prior art elements appears “1) in the prior art references themselves; 2) in the knowledge of those of ordinary skill in the art that certain references . . . are of special interest or importance in the field; or 3) from the nature of the problem to be solved, ‘leading inventors to look to references relating to possible solutions to that problem.’”<sup>83</sup> The test has been widely, though certainly not universally, criticized as lowering the barrier to patenting of obvious combinations of or improvements on old technology<sup>84</sup> and is pending review by the Supreme Court.<sup>85</sup>

Seeking out prior art patents (and other sources of prior art) is key to determining both novelty and nonobviousness. The search for related prior patents is guided to a significant extent by an ad hoc classification scheme that has been developed by the USPTO over the years.<sup>86</sup> References are cited in the issued patent document if their technical relationship to the claimed technology is close enough that they are relevant to determining whether the claimed technology is new and nonobvious.<sup>87</sup> Recent studies show that patent examiners provide a large fraction of the cited references. During the 2001-2003 period, for example, examiners provided 67 per cent of all citations. Indeed, in 40 per cent of the patents granted, *all* citations were provided by examiners.<sup>88</sup> Because such a large fraction of ref-

---

83. *Teleflex, Inc. v. KSR Int’l Co.*, 119 Fed. Appx. 282, 285 (Fed. Cir. 2005), *cert. granted*, 126 S.Ct. 2965 (2006) (citing *Ruiz v. A.B. Chance Co.*, 234 F.3d 654, 665 (Fed. Cir. 2000)).

84. *See, e.g.*, Brief for the Petitioner, *supra* note 74; John H. Barton, *Non-Obviousness*, 43 IDEA 475 (2003); Lunney, *supra* note 64, at 374-89; Robert P. Merges, *Uncertainty and the Standard of Patentability*, 7 HIGH TECH. L. J. 1, 14 (1993). *But see, e.g.*, Cotropia, *supra* note 64 (arguing that the Federal Circuit has not applied a “narrow” version of the suggestion test, but has tailored the test to the complexity of the technology).

85. *See KSR Int’l Co. v. Teleflex, Inc.*, 126 S. Ct. 2965 (2006) (No. 04-1350); *see also* Dennis Crouch, *Supreme Court: Current Test of Obviousness is “Gobbledygook”*, Patently-O, Nov. 29, 2006, [http://www.patentlyo.com/patent/2006/11/supreme\\_court\\_c.html](http://www.patentlyo.com/patent/2006/11/supreme_court_c.html) and associated linked documents, including amicus briefs taking a variety of positions regarding the “suggestion test” issue.

86. *See* USPTO, MANUAL OF PATENT CLASSIFICATION, <http://www.uspto.gov/web/patents/classification> (last visited Oct. 19, 2006).

87. *See* MPEP, *supra* note 15, § 707.05.

88. Juan Alcacer & Michelle Gittelman, *How Do I Know What You Know? Patent Examiners and the Generation of Patent Citations*, REV. OF ECON. & STAT. (forthcoming 2006).

erences are provided by patent examiners, and another large group by patent attorneys, citations do not necessarily indicate a direct flow of knowledge and thus we treat them only as indications of technological relationships.

Patents and their citations form a directed network, meaning that citations go from later patents to earlier patents and not in the opposite direction, in which patents are the network nodes and citations are directed links. Citations convey valuable information about the relationships between the technologies covered by the citing and cited patents. One can thus view the patent citation network as a kind of map of the space of patented technology, indicating the relationships between various pieces of “property” in that space.<sup>89</sup> As discussed in Section III.B, the evolution of the network may help to illuminate whether the PTO is awarding patents for more trivial technological steps.

While the precise significance of a patent citation varies, a citation sometimes indicates that the claims of the cited patent encompass the claims of the citing patent and that a blocking patent situation exists. Consequently, one must obtain permission from both patent owners in order to use the invention claimed in the citing patent. As will be discussed in Part V, we believe it is likely that one can mine the structure of the patent citation “map” for signatures of patent thickets, in which there is a high density of overlapping patent claims, so as to test, for example, whether such thickets are increasingly prevalent in the patent system.

### *3. Relationship of this Study to Previous Uses of Patent Citation Data*

The United States patent system provides a historical record that encompasses much of the history of innovation in this country and, increasingly, abroad. Records of patent prosecution, patent citations, and patent litigation over many years are publicly available. The history provided by patent records is incomplete, of course. Many technical advances are either unpatented trade secrets or unpatentable technical know-how. Moreover, the patent record was historically limited to traditionally industrial innovations, though the scope of patentable subject matter has expanded to cover an increasingly broad range of innovations at an increasingly early stage of development. Whatever its limitations, the patent record has long been recognized as a rich source of data about innovation and innovation

---

89. Of course, this map is neither perfect nor complete. Examiners and applicants may miss relevant connections between patents, cite particular patents because they are familiar, and so forth. The analysis here assumes only that citations generally indicate technological relationships between citing and cited patents.

policy.<sup>90</sup> In the past, the form in which the data were available made quantitative analysis a difficult and painstaking process and limited the kinds of analyses that could be performed. Recent advances in computer technology, however, along with the efforts of empirical economists, have rendered a wealth of patent data suitable for large-scale quantitative analysis.<sup>91</sup>

A surge of attempts by economists, social scientists, and legal academics to capitalize on this new availability has resulted.<sup>92</sup> For the most part,

---

90. See, e.g., Zvi Griliches, *Patent Statistics as Economic Indicators: A Survey*, 28 J. ECON. LIT. 1661 (1990); Manuel Trajtenberg, *A Penny for your Quotes: Patent Citations and the Value of Innovations*, 21 RAND J. ECON. 172 (1990).

91. See Bronwyn H. Hall, Adam B. Jaffe & Manuel Trajtenberg, *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*, in PATENTS, CITATIONS, & INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY (Adam B. Jaffe & Manuel Trajtenberg, eds., 2003).

92. See, e.g., PATENTS, CITATIONS, & INNOVATIONS: A WINDOW ON THE KNOWLEDGE ECONOMY (Adam B. Jaffe & Manuel Trajtenberg, eds., 2003); Adelman & DeAngelis, *supra* note 17; Allison et al., *supra* note 13; Bessen & Meurer, *supra* note 65; Emmanuel Duguet & Megan MacGarvie, *How Well Do Patent Citations Measure Flows of Technology? Evidence from French Innovation Surveys*, 14 ECON. OF INNOVATION & NEW TECH. 375 (2005); Lee Fleming, *Recombinant Uncertainty in Technological Search*, 47 MGM'T SCI 117 (2001); Lee Fleming & Olav Sorenson, *Technology as a Complex Adaptive System: Evidence from Patent Data*, 30 RESEARCH POL'Y 1019 (2001); John Hagedoorn & Myriam Cloudt, *Measuring Innovative Performance: Is There an Advantage in Using Multiple Indicators?*, 32 RES. POL'Y 1365 (2003); Hall & Ziedonis, *supra* note 65; Hall, *supra* note 64; Dietmar Harhoff, Francis Narin, Frederic M. Scherer & Katrin Vopel, *Citation Frequency and the Value of Patented Inventions*, 81 REV. ECON. AND STAT. 511 (1998) [hereinafter Harhoff et al., *Citation Frequency*]; Dietmar Harhoff, Frederic M. Scherer, & Katrin Vopel, *Citations, Family Size, Opposition and the Value of Patent Rights*, 32 RES. POL'Y 1343 (2003) [hereinafter Harhoff et al., *Value of Patent Rights*]; Zan Huang, Hsinchun Chen, Alan Yip, Gavin Ng, Fei Guo, Zhi-Kai Chen & Mihail C. Roco, *Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field*, 5 J. NANOPARTICLE RES. 333 (2003); Jean O. Lanjouw & Mark Schankerman, *Patent Quality And Research Productivity: Measuring Innovation with Multiple Indicators*, 114 ECON. J. 441 (2004); Alan C. Marco, *The Option Value of Patent Litigation: Theory and Evidence*, 14 REV. FIN. ECON. 323 (2005); Alan C. Marco, *Dynamics of Patent Citations*, 94 ECON. LETTERS (forthcoming Feb. 2007), available at <http://irving.vassar.edu/VCEWP/VCEWP84.pdf>; Per Botolf Maurseth, *Lovely but Dangerous: The Impact of Patent Citations on Patent Renewal*, 14 ECON. OF INNOVATION & NEW TECH. 351 (2005); Kimberly A. Moore, *Worthless Patents*, 20 BERKELEY TECH. L.J. 1521 (2005); Iwan von Wartburg, Thorsten Teichert & Katja Rost, *Inventive Progress Measured by Multi-Stage Patent Citation Analysis*, 34 RES. POL'Y 1591 (2005); Arvids Ziedonis & Bhaven N. Sampat, *Patent Citations and the Economic Value of Patents: A Preliminary Assessment*, in HANDBOOK OF QUANTITATIVE SCIENCE AND TECHNOLOGY RESEARCH 277 (Henk Moed, Wolfgang Glänzel, & Ulrich Schmoch, eds., 2004); Lee Fleming, Adam Juda & Charles King III, *Small Worlds and Regional Innovation*, (Harvard Business School Working Paper Series, No. 04-008,

that work has applied statistical regression techniques to connect patent characteristics—such as the number of citations made and received, the number of patent claims and whether the patent was renewed; inventor characteristics—such as geographical location and employment context; and, most recently, patent examiner characteristics—such as length of experience—to financial, economic, and technological indicators. In this way, commentators have used patent data to investigate knowledge flows and spillovers;<sup>93</sup> to attempt to determine the characteristics of valuable patents;<sup>94</sup> and to try to understand the role innovation plays in the behavior of various types of institutions, from universities to small and large firms.<sup>95</sup>

Because of its heterogeneity and local structure, the network of patents and citations is a much richer source of information about the patent system and the associated technological development than is generally captured by these techniques. Network science analysis of the patent citation network has great potential both to complement existing econometric studies of patent citations and to advance the understanding of networks in general because of the fact that the patent citation network is one of the largest and most completely characterized networks available for study.

---

2006), available at <http://ssrn.com/abstract=892871>; Bhaven N. Sampat, *Examining Patent Examination: An Analysis of Examiner and Application Generated Prior Art* (Working Paper, Prepared for NBER Summer Institute, 2004); Gerald Silverberg & Bart Verspagen, *The Size Distribution of Innovations Revisited: An Application of Extreme Value Statistics to Citation and Value Measures of Patent Significance* (Eindhoven Ctr. for Innovation Stud. Working Paper 04.17, 2004); Olav Sorenson, Jan W. Rivkin, & Lee Fleming, *Complexity, Networks and Knowledge Flow* (Harvard Bus. Sch. Working Paper No. 04-027, 2003); Deborah Strumsky, Jose Lobo, & Lee Fleming, *Metropolitan Patenting, Inventor Agglomeration and Social Networks: A Tale of Two Effects*, (SFI Working Paper No. 05-02-004, 2005), available at <http://www.santafe.edu/research/publications/wpabstract/200502004>; Bhaven N. Sampat & Arvids A. Ziedonis, *Cite Seeing: Patent Citations and the Economic Value of Patents* (Nov. 1, 2002) (unpublished manuscript, on file with author); Alan C. Marco, *Patent Protection, Creative Destruction, and Generic Entry in Pharmaceuticals: Evidence from Patent and Pricing Data* (Oct. 28, 2005), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=945007](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=945007).

93. See, e.g., Jaffe & Trajtenberg, *supra* note 92; Duguet & MacGarvie, *supra* note 92; Fleming, Juda & King, *supra* note 92; Sorenson, Rivkin, & Fleming, *supra* note 92; Strumsky et al., *supra* note 92.

94. Jaffe & Trajtenberg, *supra* note 92; Allison et al., *supra* note 17; Bessen & Meurer, *supra* note 65; Hagedoorn & Cloudt, *supra* note 92; Harhoff et al., *Citation Frequency*, *supra* note 92; Harhoff et al., *Value of Patent Rights*, *supra* note 92; Lanjouw & Schankerman, *supra* note 92; Moore, *supra* note 92; Ziedonis & Sampat, *supra* note 92; Sampat & Ziedonis, *supra* note 92.

95. See, e.g., Jaffe & Trajtenberg, *supra* note 92; Hall & Ziedonis, *supra* note 65; Huang et al., *supra* note 92; Fleming, Juda & King, *supra* note 92.

Though legal scholars and economists have generally not employed a network approach,<sup>96</sup> a few studies of innovation by social scientists have applied social network analysis techniques to patent data. These studies, some of which we describe in a bit more detail in Part IV, view patents as footprints of innovation. They interpret the pattern of patent citations to indicate the combination of past technologies into new innovations and use the patent citation network to investigate theories of innovation as recombinant search.<sup>97</sup> These theories of innovation should be of interest to legal scholars since they move beyond a linear model of sequential or cumulative innovation and attempt to use network concepts and measures to incorporate some of the complexities of the innovative process. Dealing with inventive combinations is a critical and ill-understood aspect of patent law, as demonstrated by the difficulty in finding an effective approach to nonobviousness.

Our treatment of patent citations differs from that of a number of previous studies because we attempt to make minimal assumptions about the significance of individual citations. Many previous studies have assumed that when a patent cites another patent it is an indication of knowledge flow from the cited patent to the inventors of the citing patent.<sup>98</sup> This assumption is questionable in light of data, only recently available in electronic form, showing the extent to which citations are inserted, not by the patent applicant, or even by the applicant's attorneys, but by patent examiners long after the time of invention.<sup>99</sup> Our approach to patent citations is parsimonious and limited to an assumption that citations generally indicate significant technological relationships between patented technologies. We do not attempt to probe knowledge flows.

Finally, this study differs from many econometric studies because we make no assumptions about the distribution or functional form that describes the citation data. Econometric studies using patent data have been criticized due to their reliance on such assumptions because citation data are highly skewed and far from the "normal distributions" which are typi-

---

96. A notable exception is the work of Prof. Gavin Clarkson, who is employing a network density metric to investigate patent pools. *See* Clarkson, *supra* note 6.

97. Fleming, *supra* note 92, at 121-22; Fleming & Sorenson, *supra* note 92.

98. *See, e.g.,* Jaffe & Trajtenberg, *supra* note 92 (containing a collection of such studies).

99. *See, e.g.,* Alcacer & Gittelman, *supra* note 88, at 26; Paola Criscuolo & Bart Verspagen, *Does It Matter Where Patent Citations Come From? Inventor versus Examiner Citations in European Patents 3* (Eindhoven Centre for Innovation Studies, Working Paper No. 05.06, 2005); Sampat, *supra* note 92, at 9-10.

cally used in statistical analysis.<sup>100</sup> Thus, one must interpret those econometric analyses with particular care.<sup>101</sup> Network science approaches avoid making such assumptions and are tailored to highly heterogeneous systems, in which broad, skewed distributions are typical.

### **B. The Evolving Patent Citation Network: Is Patenting Getting Out of Hand?**

Commentators have widely remarked that patenting has burgeoned since the 1980s and almost equally widely alleged that patenting is getting out of hand—that patents are increasingly of low quality, providing the transaction costs of a divided and proprietary knowledge base without the benefit of spurred innovative progress. In fact, the number of patents issued by the USPTO has increased more or less exponentially since the patent system was inaugurated in 1790, and it is true that the rate of increase, which had been more or less constant since 1870, sharpened noticeably in the early 1980s.<sup>102</sup> An increase in patenting is an ambiguous signal, however. The early 1980s was a time of both technological ferment and changes in patent law. Both the computer revolution and the biotechnology industry got their starts at around this time. Moreover, the Supreme Court in 1980 and 1981 issued its opinions in *Diamond v. Chakrabarty*<sup>103</sup> and *Diamond v. Diehr*,<sup>104</sup> putting its stamp of approval on the patentability of biological materials and computer software, respectively. In 1983, the Federal Circuit published its first patent opinion,<sup>105</sup> inaugurating what many have argued has been an era of patent-friendly legal review after a period of purported judicial hostility to patentees.<sup>106</sup>

Any evaluation of the significance of the patenting boom must thus attempt to distinguish among potential causes, which encompass at least three possibilities. Increased patenting might stem from a faster pace of technological change, from a broader range of patented technology (which may have resulted both from the extension of the scope of legally patentable subject matter and from technological advances that have broken new ground), or from a weakened legal patentability standard. In this Section we use a network approach to separate these effects. Our results indicate that not only have the pace of innovation and breadth of patented

---

100. See, e.g., Adelman & DeAngelis, *supra* note 17, at 33-34 (critiquing various statistical studies of patent citations).

101. See *id.* at 33, 57-58.

102. See *infra* Appendix, Figure 6.

103. *Diamond v. Chakrabarty*, 447 U.S. 303 (1980).

104. *Diamond v. Diehr*, 450 U.S. 175 (1981).

105. *In re Heck*, 699 F.2d 1331 (Fed. Cir. 1983).

106. See, e.g., Merges, *One Hundred Years*, *supra* note 64, at 2223-24.

technology changed in recent years, but the pattern of patent citations has also changed. The change in citation patterns indicates an increasingly skewed distribution of citability for patents issued beginning in the late 1980s. If citability reflects technical value, then the increasing gap between the most and least citable patents reflects an increasing gap between most and least valuable patents.

While a definitive interpretation of this increasing gap is not possible without further investigation, we argue that this change is not likely to be due entirely to an increased pace and scope of patentable innovations. Instead, the increasing gap between most and least citable patents suggests that the USPTO may be issuing patents on comparatively more trivial advances. The timing of the increase in citability stratification is suggestive of an association with the Federal Circuit's increasing reliance on the "motivation or suggestion to combine" test for nonobviousness, which arguably has lowered the legal standard of patentability. It is also possible, of course, that patents are being issued on more trivial advances because inventors are applying for patents more readily. If the perceived business value of patents increases generally, the cost-benefit balance might shift toward patenting of less significant inventions. In other words, an effectively lower patenting threshold might reflect increased propensity to *apply* for patents rather than weakened standards for issuing patents. The legal patentability standard places a lower limit on the extent to which businesses can simply "choose" to patent more trivial inventions, however.

In this Section we describe the results of our study of the evolution of the pattern of citations in the patent citation network. In analyzing the patent citation network, we have used the Hall, Jaffe, and Trajtenberg dataset,<sup>107</sup> which includes the approximately 16 million citations made by the more than two million patents issued by the USPTO between 1975 and 1999, as well as updated citation data from the USPTO that extends to 2006.<sup>108</sup> We begin by discussing the effects of innovative pace and breadth on the network's evolution and explain why these effects are not sufficient to explain what is happening in the network. We then turn to the heart of our analysis, in which we demonstrate that since the late 1980s patents have become increasingly stratified in their citability. We interpret this increasing stratification as evidence of an increasing gap in technical value between the most and least valuable patents. We then argue that this in-

---

107. Hall, Jaffe & Trajtenberg et al., *supra* note 91.

108. Our analysis includes all of the patents and citations in the database. We do not randomly sample the data.

creasing value gap is likely to be due to a lowering patentability threshold and that changes in the nonobviousness standard may be responsible.

*1. Concerning the Pace and Breadth of Innovative Progress*

Before describing our analysis of the changing pattern of citations, which we argue suggests a lowering of the threshold of patentability, we consider how to separate out the effects of either a faster pace of technological advance or a broader landscape of patented technologies. We certainly do not deny that the pace of technological progress may be increasing and that the breadth of patented technology has expanded. However, we do not aim to investigate those effects here. Rather, we want to find out whether additional important changes in patenting may be occurring.

For that reason, in our analysis we measure time and patent age in units of patent numbers, rather than in units of months and years. We do this to mask out any effects of a mere speed-up of technological progress. If innovation is simply occurring at a faster pace without any change in the innovative step between patents or the character of the relationships between patents, then measuring age and time in terms of patent numbers should cancel out the effects of the increasing pace. After canceling out this source of time-dependence, remaining changes that we observe must be due to some more fundamental change in the way in which the citation network is evolving.

Another way in which the patent citation network has undoubtedly changed in recent years is through a broadening of the technologies covered by patents. To understand this issue, imagine a “landscape” of patented technology. As more patents are issued, they might overlap the existing “landscape” by claiming a somewhat improved or different version of a well-established technology—such as coffee filters or disposable diapers, for example. Alternatively, they might strike out into new territory by, for example, being the first patent to claim a practical application of a scientific advance, such as stem cell research. While this is surely an oversimplified view of the complexities of technological advancement, it is perhaps sufficient to illustrate that there is a distinction between the size of the innovative step represented by a new patent (how much better is the claimed disposable diaper, for example) and the extent to which it represents a broadening of the technological landscape. We need to consider, therefore, whether the changes in the citation network that we observe can be explained by the broadening range of patented technology rather than by a lower patentability threshold.

We have measured two quantities that are relevant to the question of to what degree the “space” of patented technologies is “spreading out” as

opposed to “filling in.” The first is the average number of citations *made* per patent. As shown in the top graph in Figure 7, the average number of citations made per patent has been steadily growing (in fact, nearly proportional to the number of patents, which is our measure of time) throughout the period of our data. This increasing number of citations made per patent seems unlikely to have been caused by mere expansion of the technological frontier. Patents that are breaking new ground by heading off in uncharted technological directions would, if anything, seem likely to encounter fewer material prior art patents to cite, rather than more. And the addition of patents in a new technical field seems unlikely to mean that patents in older fields will have to cite more patents. It seems unlikely, for example, that the addition of the field of biotechnology patents would require that patents on mechanical devices cite more prior art.

On the other hand, the increase in citations made per patent is consistent with an increasingly dense space of narrow patents, with new patents having an increasing number of patents to cite because more and more patents are material. It is similarly consistent with patenting of smaller and smaller incremental advances, such that inventors need to cite more patents to reach back to all those which are relevant to the claims of the new patent. It would also be consistent with an increased issuance of patents that are based on combinations of older technology. Thus, the increasing number of citations made suggests that something is going on in addition to the broadening of the landscape of patented technology.

Some criticisms of the recent evolution of the patent system, particularly those in the popular press, have highlighted another way in which the breadth of patented technologies might be expanding by providing examples of “outlier” patents, often of a somewhat humorous character, such as the much-discussed peanut-butter-and-jelly sandwich patents.<sup>109</sup> If the addition of such “outlier” patents were significantly expanding the landscape of patented technology, one would expect that the average citability of brand new patents would decline since such patents would be less likely to spawn improvements and follow-on innovation than patents on more mainstream technology. In fact, our analysis of the citation network’s evolution finds that the average likelihood that a new patent will be cited by the next patent that issues has *increased* somewhat in recent years due to

---

109. See, e.g., Gary L. Reback, *Patently Absurd: Too Many Patents are Just as Bad for Society as Too Few*, FORBES ASAP, Jun. 24, 2002, at 44; Hal R. Varian, *Patent Protection Gone Awry*, N.Y. TIMES, Oct. 21, 2004, at C2; see also Daniel Wright, *Patently Silly*, <http://www.patentlysilly.com> (last visited Nov. 2, 2006) (maintaining an archive of questionable patents).

the increasing number of citations made by each patent,<sup>110</sup> suggesting that true technological “outlier” patents, while making for entertaining rhetoric, are not a major part of the patenting explosion. On the whole, patent examiners and applicants have deemed it necessary to cite more patents, suggesting that the density of “mainstream” patents is increasing despite the undeniable increase in the breadth of patented technology.

Having accounted for a possible speed-up in innovative progress through our choice to measure time and age in units of patent numbers and having concluded that recent citation patterns are not likely to be explained solely in terms of an expansion of the breadth of patented technology, we now turn to the question of whether something else is going on in the patent system beyond an overall “speeding up” or “spreading out.”

## 2. *The Evolving Citability Distribution*

Our approach is motivated by statistical physics studies of a diverse range of other growing networks. We describe the evolution of the patent citation network in terms of the probability that a patent with given characteristics will be cited, which we call “citability.” To get a result that will provide insight into the general features of the network evolution, we attempt to describe the probability of citation using only those characteristics that strongly affect the likelihood that a given patent will be cited. Because a case-by-case evaluation of the underlying reasons that one patent might cite another is impossible for a large network of citations, we look for objective characteristics that we expect, on average, to be good proxies for the underlying citation process. In our study, we hypothesize that we will be able to describe the evolution of the patent network to a very good approximation by assuming that the probability that a particular patent will be cited at a given time depends primarily on its age, which we will call  $l$ , and on the number of times it has already been cited, which we will call  $k$ .<sup>111</sup> Our results bear out the hypothesis that, on average, these two characteristics are highly determinative of the likelihood that a patent will be cited.

Our assumption that the probability that a patent will be cited depends on its age requires little explanation—technology tends to become obsolete. Our expectation that the probability that a patent will be cited de-

---

110. See *infra* Appendix, Figure 7, bottom graph.

111. A more complete description of our data analysis procedure is available in our technical publication, Gábor Csárdi, Katherine Strandburg, László Zolányi, Jan To-bochnik & Péter Érdi, *Modeling Innovation by a Kinetic Description of the Patent Citation System*, PHYSICA A (forthcoming 2006), available at <http://www.arxiv.org/abs/physics/0508132> (last visited Nov. 13, 2006).

depends on how many citations it has already received derives both from our intuitions about patents and from experience with other evolving networks. As we discussed in Part II, one of the most interesting discoveries of network science has been that it is very common for nodes that already have many links to acquire additional links more quickly than nodes with fewer links. One might reasonably expect this preferential attachment or, colloquially, “rich get richer” phenomenon to occur in the patent citation network.

There is considerable statistical evidence that highly cited patents are more valuable and may be of greater technological merit.<sup>112</sup> Moreover, technology has its own “popular crowd,” depending on what field is “hot” at a particular time. In our analysis, we thus assume that the number of previous citations to a patent, which we call  $k$ , will be relevant to its likelihood of being cited again. Our computations confirm that the likelihood of being cited increases as the number of previous citations increases, as we would have guessed, but also permit us to extract empirically the detailed structure of that increase, as we explain below.

Of course, the fact that we extract a citation “probability” is not meant to suggest that the particular citation choices made by patent examiners or applicants are actually random. Our results mean only that, cumulatively, those individual citation decisions result in a likelihood of citation which depends in interesting ways on  $k$  and  $l$ .<sup>113</sup> Patents presumably have inherent “quality” that affects whether they are cited, just as individuals have personalities that make them more or less popular and websites have content that make them more or less useful. The higher citability of highly cited patents presumably depends on some combination of inherent quality and the extent to which a particular patent is already known to the relevant examiner. Our present analysis says nothing directly about the inherent patent quality of any particular patent, but does allow us to make reasonable inferences about the patent system as a whole based on the observed citation distribution and the way in which it has changed in recent years.

---

112. See, e.g., Allison et al., *supra* note 13 and references therein (discussing studies of the relationship between patent citations and patent value).

113. There is also an overall time-dependent scale factor,  $S(t)$  (where  $t$  is measured by patent numbers). More specifically, we find that the probability function  $P(l,k,t)$  can be written to a good first approximation as the ratio of a time-independent function,  $A(k,l)$ , and a time-dependent scale factor,  $S(t)$ . The scale factor  $S(t)$  is just the sum of  $A(k,l)$  over all existing patents at time  $t$ . Thus,  $S(t)$  changes over time only because the number of patents of age  $l$  and connectedness  $k$  changes. See our technical paper, Csárdi et al., *supra* note 111, for a detailed explanation of this factor.

To find the likelihood of citation for given  $k$  and  $l$  we developed a novel iterative technique, which we used to extract from the patent citation data the average likelihood that a patent with  $k$  previous citations of age  $l$  will be cited again. We do not assume a particular functional form for this probability—we derive the functional dependence on citations previously received and patent age directly from the data. The likelihood that a patent will be cited depends on both its age and on the number of citations it has previously received. Figure 9 shows the dependence of citability on age. Very young patents have low citability ( $A_l(l)$  is small); the likelihood of being cited grows rapidly in the first couple of years after a patent issues, peaks, and then it falls slowly thereafter. As it turns out, however, even though the probability of being cited depends on age, the way in which the probability of being cited depends on the number of citations previously received is more or less the same for patents of any age.<sup>114</sup> No matter the age of the patent, the ratio of the likelihood that a patent with ten previous citations will be cited to the likelihood that one with three previous citations will be cited is nearly the same. So if, for example, patents that are 10,000 patent numbers “old” are five times more likely to be cited if they have ten previous citations than if they have three, roughly the same will be true for patents that are 100,000 patent numbers “old.”

This means that we can look separately at the age dependence of the citability and the dependence on number of citations already received.<sup>115</sup> In this Section, we focus on the dependence on citations received,  $k$ . In Part IV, we discuss some of our results for the age dependence of citability.

Figure 8 shows the way in which the likelihood of being cited depends on the number of citations already received, which we call  $A_k(k)$ . Figure 8 also displays the results extracted from the data for patents of several different ages. The more often a patent has already been cited (the higher its value of  $k$ ) the more likely it is to be cited again (the higher its value of  $A_k(k)$ )—the signature of preferential attachment.

This demonstration of preferential attachment in the patent citation network is not especially surprising—preferential attachment is a common property of growing networks and is intuitively sensible in the patent system. Preferential attachment is cumulative—highly cited patents are more likely to be cited, hence becoming even more highly cited and even more

---

114. See *infra* Appendix, Figure 8.

115. Mathematically, this means that the likelihood of being cited can be written approximately as a product of two functions, i.e.,  $A(k,l) = A_l(l)A_k(k)$ .  $A_l(l)$  depends only on age while  $A_k(k)$  depends only on the number of citations previously received.

likely to be cited, and so forth. The pattern of citability that Figure 8 shows thus eventually leads to an extremely skewed distribution of citations eventually received by patents. Most patents are hardly cited at all, while a few patents become citation “billionaires” (well, “hundredaires,” really).<sup>116</sup> If we make the reasonable (and empirically supported) assumption that the number of times a patent is cited signals its technological value,<sup>117</sup> we can infer that the very skewed and broad citation distribution in Figure 1 reflects a very stratified distribution of patent values, with a few superstars and a vastly larger number of patents that go nowhere.

This general picture of a highly skewed patent value is well known by now.<sup>118</sup> Our network analysis allows us to get behind this general observation, however, to ask just how stratified patent value is and how the extent of stratification has evolved over time. If more patents are issuing simply as a result of faster or broader technological progress, we would expect the degree of stratification to remain about the same over time. On the other hand, if the patentability standard decreases, there would be not only more patents issued, but a higher proportion of them would be less important; the degree to which highly citable patents dominate trivial patents should increase.

The citability function,  $A_k(k)$ , gives us a quantitative handle on the degree of stratification among issued patents. We arrive at this function after factoring out the effects of patent aging due to obsolescence (which, as noted, turn out to be roughly the same no matter how many citations a patent has previously received). The function thus gives a relatively simple and unobscured view of the underlying citability of patents. Looking at Figure 8 we see that the citability of a patent on average increases somewhat more than proportionally to the number of citations already received. We can quantify this “somewhat more” by noting that  $A_k(k)$  is closely fit by the form  $A_k(k) \sim k^\alpha$ . The parameter  $\alpha$  is a measure of the extent to which highly cited patents are preferred. When averaged over all of the data from 1975 on,  $\alpha \cong 1.19$  (with an estimated error of less than .01).

To give us a feel for the interpretation of the values of  $\alpha$ , we can compare the patent system’s  $\alpha$  value with known  $\alpha$  values of other network studies. If there were no aging or obsolescence, nothing would inhibit old patents from acquiring extremely large numbers of citations. In that situa-

---

116. See *infra* Appendix, Figure 1 (showing the proportion of patents that had received 0, 1, 2, . . . citations).

117. See, e.g., Allison et al., *supra* note 13, at 449.

118. See, e.g., Bessen & Meurer, *supra* note 65, at 8; F.M. Scherer & Dietmar Harhoff, *Technology Policy for a World of Skew-Distributed Outcomes*, 29 RES. POL’Y 559 (2000).

tion, theoretical network models have shown that when  $\alpha$  is larger than 1 the network “condenses” to a highly unequal situation in which nearly all nodes in the system have very low connectivity. These “unimportant” nodes are all connected to a very small (finite even in an infinite network) number of very highly connected nodes<sup>119</sup>—almost all nodes are “peons” connected to a very few “royal” nodes and there is no “middle class” of moderately connected nodes. In the patent system, such an extreme stratification does not occur because highly cited patents eventually become obsolete and generally cease to be cited. However, the fact that  $\alpha > 1$  in the patent network still suggests that citations are highly concentrated. There is a highly unequal distribution of patent citability—and hence, most likely, of patent value.

### 3. *The Increasing Stratification of Patent Citation Patterns*

The network analysis not only shows us that patent citability is stratified, but also gives us a means to investigate whether the degree of stratification has varied over time. We find that patent citability has been becoming increasingly stratified since the late 1980s. We investigate this phenomenon by calculating  $\alpha$  using only the patents within a 500,000-patent sliding time window and calculating a value for  $\alpha$  after every 100,000 patents. The value of  $\alpha$ —and hence the degree of stratification of patent citability—has varied in an interesting way. As Figure 10 shows, the stratification of citability, as reflected in the value of  $\alpha$ , began to rise in the late 1980s and has continued to rise throughout the period of our study. The rise followed a period of slightly decreasing stratification. (We do not know what was happening before 1982 because we do not have sufficient earlier data.) Figure 6 shows that the number of patents issued annually has been rising essentially since the inauguration of the patent system (and rising very rapidly since the early 1980s). Thus, during a period throughout which the absolute number of patents issued was rising rapidly, the relationships between those patents were also changing—but not in a way that simply reflects increasing numbers. Patent citability became first less and then increasingly stratified. The increase in  $\alpha$ —corresponding to increasingly stratified citation patterns—began nearly 10 years later than the beginning of the recent rise in patent issuance and after the major change in the patent system represented by the inauguration of the Federal Circuit.

---

119. P.L. Krapivsky, S. Redner & F. Leyvraz, *Connectivity of Growing Random Networks*, 85 PHYS. REV. LETTERS 4629 (2000).

How should we interpret the increasing stratification of patent citability since the late 1980s? There are several possibilities. One possibility is that the legal patentability standard has decreased, resulting in the issuance of a larger fraction of more trivial—and hence less citable—patents. This explanation is consistent with growing societal concern with “low quality” patents.<sup>120</sup> A second possibility is that the proportion of highly citable pioneer patents has increased. A third possibility is that the change in stratification parameter reflects changes in the subject matter mix of patented technologies. Finally, it is possible that the change in citability stratification reflects some kind of change in citation practices, rather than a change in inherent patent characteristics.

The increasing stratification of patent citability is correlated with increasing concerns with patent quality. Anecdotal and survey evidence suggests that patent quality has been decreasing in recent years, resulting in the issuance of a larger fraction of more trivial—and hence less citable—patents. One possible reason for the issuance of lower quality patents is a lower legal threshold of nonobviousness. At around the time that patent citability stratification began to increase, the Federal Circuit (followed of course by the USPTO and, presumably, by patent applicants themselves in making their filing decisions) increasingly adopted the “teaching, suggestion, or motivation to combine” test for nonobviousness.<sup>121</sup> While the effect of the “suggestion test” on patent issuance is a subject of debate<sup>122</sup> (and the test is, as already mentioned, the subject of a pending Supreme Court case<sup>123</sup>), there are doctrinal reasons to believe that it is a systematically lower standard of patentability than the previous legal standard. It is thus plausible that the increasing stratification reflects a lower threshold of nonobviousness resulting from increasing application of the “suggestion test.”

While we believe that a weakening patentability standard is the most likely explanation for the increasing stratification of patent citability—and, inferentially, the increased disparity in patents’ technical worth—there are other possible explanations. Perhaps the increasing stratification is due not to an increasing issuance of trivial patents, but to an increasing issuance of more highly citable pioneer patents. An increase in pioneer patents could result from increased patenting of upstream research results, such as has been occurring, for example, in the field of biotechnology.

---

120. See, e.g., *supra* note 7.

121. See *infra* Appendix, Figure 11 (showing the Federal Circuit’s increasing use of the “suggestion test”).

122. See, e.g., Strandburg & Wagner, Debate, *supra* note 71.

123. *KSR Int’l Co. v. Teleflex, Inc.*, 75 U.S.L.W. 3094 (2006).

A third possibility, which we plan to explore in future studies, is that the average stratification parameter  $\alpha$  is changing because of changes in the subject matter mix of patented technologies. If patent “importance” is inherently more stratified in one field of technology than in another (because of a difference in the importance of “pioneer” patents, for example), then an increasing prevalence of patents in that field could change the average degree of stratification that we observe. Preliminary studies of a rough division of patents into six technological categories did not turn up any significant variations in  $\alpha$ , but a more in-depth study of differences between technological fields is under way.

A final possibility is that the change in citability over the years reflects a change in citation practice, rather than a change in inherent patent characteristics. This possibility is especially interesting because the timing of the increased stratification in the late 1980s corresponds to the time at which computerized searching became increasingly prevalent. We are inclined to reject this possibility at present because most of the trends in patent citation practice that we can think of—most notably the increased ease of computerized searching for prior art—seem unlikely to have changed direction from decreasing  $\alpha$  to increasing  $\alpha$  during the 1980s. Computerized searching seems likely to have had a one-way influence. One way to check for the influence of search technology would be to compare the behavior of the United States patent citation network with other citation networks, such as the European patent citation network or the network of citations in scientific journals.

To summarize, in this Part we have shown empirically that the distribution of patent citability has been changing in recent years. Prior to the late 1980s, citability was becoming slightly more egalitarian—the difference between the citability of the most highly cited patents and that of less cited patents was decreasing. From the late 1980s on, that trend has been reversed. Citability has become more stratified, with highly cited patents becoming more and more citable compared to less cited patents. In line with our intuition and with earlier studies of patent value, we interpret citability as a reflection (on average) of technological importance. We thus conclude that the distribution of patent importance has also become more stratified since the late 1980s. There are several possible explanations of this trend. We hypothesize, consistent with more general evidence of decreasing patent quality, that the increased stratification may result from a lowered patentability threshold, possibly reflected in the “teaching, suggestion, or motivation to combine” test of nonobviousness, and results in the issuance of more trivial patents.

#### IV. OF SLEEPER PATENTS AND INNOVATIVE PROGRESS: HOW INNOVATION BUILDS ON PRIOR TECHNOLOGY

We now turn to our results for the age dependence of the likelihood that a patent will be cited. Our study of how patent citability depends on patent age contributes to understanding the way in which innovation proceeds by building on prior technology. Most studies of the process of innovation reflected in the legal and economic literature have, to render the analysis tractable, taken a simplified, linear view in which innovation is pictured as sequential or cumulative, each new advance building on a previous innovative step. In fact, innovation is a much more complicated process involving new advances, combinations of old technologies, and potentially reaching back into the past to make new uses of technologies that acquire new relevance in light of some new advance.<sup>124</sup> Our results for how citability depends on age provide a window into this process.

As already mentioned, in the course of our analysis of patent citability, we determined how patent citability depends on a patent's age (in patent numbers). We noted that the way in which a patent's citability varies with the number of citations it has already received is roughly the same for patents of all ages. Correspondingly, the way in which patent age affects the likelihood of being cited is nearly the same no matter how many times a patent has been cited before. Patents that have rarely been cited have a lower overall chance of being cited than highly cited patents, but the "age profile" of their chance of being cited is about the same. Mathematically, the dependence of citability on age, which we call  $A_l(l)$ , is relatively independent of  $k$ , especially at large  $l$  and  $k$ , as shown in Figure 9 for several values of  $k$ .

The observed dependence of citation probability on patent age provides insight into the complicated dynamics of technical innovation. Not surprisingly, citability peaks at a relatively young age (small  $l$ ). More interestingly, citability decays unexpectedly slowly for older patents.<sup>125</sup> For large  $l$ ,  $A_l(l) \sim l^{-\beta}$ , where  $\beta \cong 1.6$ . This power law form signifies a long, slow decay—very old patents are still being cited. The peak in citations relatively soon after issuance likely corresponds to "typical," incremental improvements, such as those reflected in linear models of cumulative in-

---

124. "Percolation" and recombination models of innovation attempt to incorporate the non-sequential character of innovation. *See, e.g.*, Fleming, *supra* note 92; Fleming & Sorenson, *supra* note 92; Gerald Silverberg & Bart Verspagen, *A Percolation Model of Innovation in Complex Technology Spaces*, 29 J. ECON. DYNAMICS & CONTROL 225 (2005).

125. Previous studies of citation lags have also noted that old patents continue to be cited. *See, e.g.*, Hall, Jaffe & Trajtenberg, *supra* note 91, at 421-24, 448-51.

novation. Such incremental advances are likely to occur relatively soon after a patent issues. The tail of citations that occur long after issuance no doubt includes “pioneer”-type patents that continue to influence innovation (and hence continue to be cited) over long periods of time. However, not all of these late citations are to broad, pioneer patents. The form of  $A_i(t)$ —including the tail of citability long after issuance—is the same even for patents that have rarely been cited in the past. This means that even patents that have been cited very rarely or never in the past are sometimes cited long after issuance.

Thus, there is no age at which patents can be pronounced “dead.” While most patents that have not been cited shortly after they are issued will never be cited, there are “sleeper” patents that go without citation for long periods of time, only to reawaken at a later time. If one makes the reasonable assumption that receiving a citation is some indication of social value, this observation suggests that inventive progress is not simply a matter of steady accumulation of incremental progress, but a complicated process in which old patents may gain new significance in light of later advances. The long tail of inventive relevance also underscores the difficulty of predicting the social value of innovations in advance.

A network approach permits us to take into account the multi-dimensional and combinatorial nature of technological progress. Our age profile is just one simple way to use citation network analysis to probe the innovative process. Social scientists have also begun to use network measures of patent citations to explore the way in which technological change occurs.<sup>126</sup> From this perspective, one may view citations as indicators of the ways in which prior technologies have been combined to produce a new invention. In this vein, Fleming, Sorenson, and their collaborators<sup>127</sup> used patent citations in combination with the USPTO classification system to explore the way in which innovation proceeds as a process of search and recombination of prior technologies. They argue that successful innovation is a balance between re-using familiar components—an approach that is more certain to succeed—and combining elements that have rarely been used together—an approach that is more likely to fail entirely, but also more likely to result in radical improvements.

---

126. See, e.g., Fleming, *supra* note 92; Fleming & Sorenson, *supra* note 92; Joel M. Podolny, Toby E. Stuart & Michael T. Hannan, *Networks, Knowledge, and Niches: Competition in the Worldwide Semiconductor Industry, 1984-1991*, 102 AM. J. SOC. 659 (1996); Joel M. Podolny & Toby E. Stuart, *A Role-Based Ecology of Technological Change*, 100 AM. J. SOC. 1224 (1995).

127. Fleming, *supra* note 92; Fleming & Sorenson, *supra* note 92.

Podolny and Stuart and collaborators also made explicit use of social network analysis by using a basic network measure of local network structure, the transitivity, to study the innovative process.<sup>128</sup> Transitivity measures the likelihood that two nodes that are connected to a specific third node are also connected to one another. Figure 12 illustrates this concept. In the patent context, transitivity measures the likelihood that two patents that cite or are cited by the same patent also cite one another.

Podolny and collaborators used transitivity-type quantities to define local measures of competitive intensity and competitive crowding based on indirect patent ties.<sup>129</sup> A basic insight of their work is that if many innovations are building on the same technological antecedents, and thus contributing very similar technological outputs, the likelihood of a new entrant into the associated niche is lowered. They used network transitivity measures to observe how a technological niche can become popular, but then overly crowded and “exhausted” as a result of a flurry of inventive activity in the niche. They point out that counting the numbers of citations made or received is not sufficient to identify crowded niches—an innovation that provides a technological foundation for a variety of unrelated advances will be highly cited without crowding. Transitivity measures thus show promise as means to distinguish patents in hot fields from broad, important patents.

Studying network metrics and their evolution over time thus allows us to delve into the way in which inventions are combined, recombined, and re-used to produce new innovations. In the future, we can build on our work and the social science studies discussed above by combining the approaches. For example, we might investigate the extent to which delayed citations in the power law tail of the  $A_i(l)$  function target patents with high “originality” (low transitivity) which connect disparate technologies.

## V. FUTURE DIRECTIONS AND SOME PRELIMINARY RESULTS

The network approach has the potential to elucidate other questions of patent and innovation policy, some of which we discuss briefly here.

---

128. See Podolny, Stuart & Hannan, *supra* note 126; Podolny & Stuart, *supra* note 126.

129. See Podolny, Stuart & Hannan, *supra* note 126; Podolny & Stuart, *supra* note 126.

### A. Patent Classification and the Meaning of Analogous Arts

The categorization of patents by technological field of endeavor is of both practical and analytical significance. Patent classification plays an important role in prior art searching. Moreover, under the patent doctrine of analogous arts, the technological field of endeavor determines which prior art patents one must consider in determining whether a claimed invention is nonobvious.<sup>130</sup> Though the law requires that novelty be absolute, courts judge nonobviousness only with respect to prior technology in analogous arts of which a person having ordinary skill in the art would reasonably be apprised. To apply this standard, it is necessary, of course, to determine which fields of technology are analogous to the field of the patented invention and which patents are in a relevant art. Moreover, any study of technological innovation, or of the functioning of the patent system, which relies on patent data and seeks to inquire into potential differences between fields of technology must find some means of categorizing patents.

At present, a USPTO examiner categorizes each patent application according to a scheme of classes and subclasses that has developed in an ad hoc manner over the years. The classification assists examiners and patent applicants in searching for relevant prior art. Researchers have also used USPTO patent classifications to assess such things as the generality of a patent (evaluated by the extent to which it is cited by patents from different classes), the originality of a patent (evaluated by the extent to which it cites patents from different classes), the technological closeness of firms involved in patent litigation (evaluated by the extent to which the patents assigned to those firms are in the same classes), and the way in which innovations recombine components from the prior art.<sup>131</sup> The usefulness of these measures is limited, however, by the fact that the classification scheme is ad hoc and does not provide a well-defined way to measure the degree of distinctiveness between inventions in different subclasses or classes. A more objective and quantitative measure of patent “closeness” is desirable.

Network analysis provides an alternative way to measure technological distance that is more fine-grained and quantitative than comparing USPTO classifications and less ad hoc than the classifications themselves. The

---

130. MPEP, *supra* note 15, § 904.01(c); 2 CHISUM, *supra* note 15, § 5.03.

131. Jaffe & Trajtenberg, *supra* note 92; Fleming, *supra* note 92; Fleming & Sorenson, *supra* note 92; James E. Bessen & Michael J. Meurer, *The Patent Litigation Explosion* (Boston Univ. School of Law Working Paper No. 05-18, 2005), available at <http://ssrn.com/abstract=831685>.

number of citation “hops” that it takes to get from one patent to another is one way of measuring how “far apart” two patented technologies are. Essentially, patented technology is likely to be related most closely to the technology of the patents it cites or is cited by directly, less closely to the technology of the patent that is two steps away, and so on. Path length measures may provide a more fine-grained (and complementary) measure of technological closeness than metrics that use the number of overlapping subclasses in the USPTO classification system.

One may also use the network path length to evaluate and explore existing patent classification schemes. One may compare the distances between patents in the same USPTO category with distances between patents in different categories to evaluate the accuracy of the classification system. Moreover, one can compute the average distances between different categories to get a more qualitative measure of the relationships between different categories.

## B. The Shrinking Patent Citation Network

In a preliminary study, we have measured the overall “size” of the patent citation network by looking at the shortest path lengths between patent nodes. As discussed in Part II, one of the most interesting results of network science studies is that many networks have what has come to be called the “small world” property.<sup>132</sup> This property, made famous by the phrase “six degrees of separation”<sup>133</sup> and by the “Kevin Bacon game,”<sup>134</sup> measures how closely knit a network is. To determine whether a network has the small world property, one counts how many “hops” from node to node along the network one must make to travel between any two nodes. The number of hops along the shortest path between two nodes is sometimes called the “geodesic distance.” The maximum geodesic distance between any two nodes in the network is the “diameter” of the network. If the maximum geodesic distance between any two nodes in a network re-

---

132. See WATTS, *supra* note 2, Albert & Barabasi, *supra* note 2, Newman, *supra* note 2, and Watts & Strogatz, *supra* note 31, for discussions of the basic properties of small world networks.

133. See WATTS, *supra* note 2; Stanley Milgram, *The Small World Problem*, 2 PSYCHOL. TODAY 61, 64-67 (1967) (reporting Milgram’s famous experiment which involved counting the number of transfers necessary to get a letter delivered from an arbitrary individual in Omaha, Nebraska to a specified individual in Sharon, Massachusetts).

134. The Kevin Bacon Game, which may be played at <http://www.cs.virginia.edu/oracle/>, counts any two actors as linked if they have had roles in the same movie. The “Oracle of Kevin Bacon” shows that any two of the 800,000 or so actors in the Internet Movie Database are linked to Kevin Bacon in an average of fewer than 3 steps.

mains relatively small,<sup>135</sup> the network has the “small world property.” The diameter of many real world networks is surprisingly small.<sup>136</sup>

We have determined so far that the patent citation network displays the small world property discussed in Section II.A.1.b. The patent citation network is not completely connected. There are a few isolated patents and clusters. However, the network contains a “giant” component which contains the vast majority of patents. Because citations are directed links—a particular citation goes from one patent to another and not vice versa—one can measure the small world property in two different ways. We have measured the small world property using directed paths, meaning that we permit only hops from citing to cited patents. The longest directed geodesic distance between nodes in the “giant” component (of nearly 4 million patents) is only 24 steps,<sup>137</sup> while the average is 6.9 steps. It is also possible to investigate the small world property of the undirected citation network, in which hops are permitted either from citing to cited patents or from cited to citing patents. An exact calculation of the diameter of the undirected network is computationally difficult because of the large size of the patent citation network. An approximate calculation on the giant component of the undirected patent citation network has been performed by Leskovec et al.<sup>138</sup> They determined, using a sampling method, that the average undirected geodesic distance is about 10 steps. These two calculations measure somewhat different quantities, but the qualitative conclusion is the same—the patent citation network is very closely connected.<sup>139</sup>

Leskovec et al. also determined that the “size” of the patent citation network has been shrinking over time, even though the number of patents has increased dramatically. This shrinking diameter is not typical of small

---

135. Technically, if it increases no more rapidly than the logarithm of the number of nodes.

136. See, e.g., BARABASI, *supra* note 2; WATTS, *supra* note 2; Newman, *supra* note 2.

137. The giant component contains more than two million patents because it includes cited patents from before 1975.

138. Jure Leskovec, Jon Kleinberg & Christos Faloutsos, *Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations*, 2005 PROC. OF THE ELEVENTH ACM SIGKDD INT’L CONF. ON KNOWLEDGE DISCOVERY IN DATA MINING 177. This paper won the Best Research Paper Award.

139. The fact that the average undirected path length is longer than the average directed path length is somewhat counterintuitive. This occurs because there are pairs of patents which are not connected at all by a directed path. The directed path length between such patents is effectively infinite, but these infinite path lengths are not included in the average of 6.9 steps which we report. Once undirected hops are permitted, some of these nodes are connected, but the path lengths required to connect them may be long enough to raise the average. There are also difficult questions about how best to sample a network to determine average properties.

world networks, which generally have very slowly *increasing* diameters, and may indicate increasing interdisciplinarity (interdisciplinary patents will cite patents in disparate fields, thus providing a network “shortcut” between disciplines) or the increasing importance of “bridge” technologies which are used in a number of fields. We plan to study the distribution of directed and undirected path lengths in more detail and, if the shrinking “size” is confirmed, will combine path length measures with the USPTO classification scheme to understand the source of the closer connectedness of the network.

The small world property of the network also suggests the intriguing possibility that a relatively small number of “hub” or “bridge” patents are integral to many apparently separate fields (and thus provide “shortcuts” between patents in disparate fields). In future work, we hope to investigate that possibility and to study qualitatively any “hub” or “bridge” patents that we can identify.

### C. Patent Thickets and Potential Anti-Competitive Licensing Practices

The proliferation of patents results in a situation in which many commercial products require the use of more than one—and in some cases hundreds of—patented technologies. This situation raises a number of sometimes conflicting concerns. On the one hand, there is the fear of a patent thicket in which the transaction costs associated with obtaining the necessary patent licenses to do something of practical usefulness become so high as to undermine the social value of the patents.<sup>140</sup> There can even be a hold-up problem in which the owners of contributing technologies cannot come to an agreement at all, as a result of differing assessments of the relative value of various contributions to a commercial product.<sup>141</sup>

There are two distinct ways in which such a transaction costs nightmare could arise. It might be that a commercial application requires many distinct and complementary patented technologies. This could happen either as a result of increasing complexity in the technology itself or as a result of patenting more and “smaller” pieces of technology space. A need for many patent licenses might also be due to patent blocking. Patent blocking occurs when a patent improves on the invention claimed in another patent in such a way that the claims of the earlier patent still cover the improved invention. Where patent blocking occurs, improvers may require licenses to a string of “nested” improvement patents to use one

---

140. See, e.g., Shapiro, *supra* note 64; Bessen, *supra* note 77.

141. See, e.g., Merges, *supra* note 78, at 84-91.

particular technology. In either case, one way to handle the necessary cross-licensing and cut down on transaction costs is for the owners of the relevant patents (who are often commercial entities in the same industry) to form patent pools.<sup>142</sup>

A high density of patents in a particular technological “niche” need not always indicate a patent thicket, however. Closely related patented technologies may be potential substitutes for one another. This situation creates something more like patent supermarkets, offering many nearly interchangeable options, than patent thickets. If these patents are separately owned, competition between patent holders will reduce licensing fees and the issue of hold-up will not arise. In situations like this, in contrast to the patent thicket situation, patent pools and cross-licensing agreements between industry actors may be anti-competitive. Rather than serving to reduce transaction costs and avoid hold-up, such agreements may permit firms to avoid competing for licensing revenues and drive up royalties.

Distinguishing between these two kinds of relationships between patents is of great theoretical and practical interest. Citation network structural measures may be useful in addressing questions related to the patent thicket issue. Prof. Gavin Clarkson has suggested that a high density of citations (meaning the fraction of possible citations that are actually made) in a particular patent neighborhood may suggest where patent pools might legitimately form. Setting aside some technical issues about how to compare citation densities from different sized network samples, citation density is a useful way to assess the level of technological interrelatedness in a particular field.<sup>143</sup> Density alone cannot distinguish between patent thickets and patent supermarkets, in which closely related technologies compete for licensing customers, however. The mere existence of a citation from one patent to another cannot tell us which scenario is most likely. Determining the meaning of a particular citation is an extremely labor-intensive process requiring understanding of the legal and technical relationship between the citing and cited patents. To investigate the existence of patent thickets and the potential for related antitrust problems, some structural metric that is sensitive to the character of a citation is highly desirable. It may be possible to design such a metric based on transitivity concepts.

Podolny and Stuart note that a technological tie (and hence a likelihood of a citation) exists between two patents if “the contribution of [the second] incorporates, builds on, or is bounded by a technological contribu-

---

142. See, e.g., Shapiro, *supra* note 64, at 123; Goller, *supra* note 79, at 727.

143. Clarkson, *supra* note 6.

tion of [the first.]”<sup>144</sup> The distinction between these relationship types is critical to understanding the extent to which increased patenting in a particular area signifies patent blocking and high-transaction-cost thickets or desirable designing around and competition.

As we alluded to above, patents may cite an earlier patent because they build on its technology (what might be termed “lovely” citations) or because they replace or distinguish its technology (“dangerous” citations).<sup>145</sup> Depending on the scope of a patent’s claims, a lovely citation may mean the owners of both patents must authorize that use of the later technology. A dangerous citation, on the other hand, may mean that a prospective user of the related technologies can play them off against one another in bargaining for a license. The existence of lovely and dangerous citations is thus related to the potential for patent thickets and for antitrust problems when industry competitors sign cross-licenses or form patent pools.

We suggest that a patent C that cites a patent B which makes a lovely citation to an earlier patent A is relatively likely to cite the earlier patent as well.<sup>146</sup> This is because for a lovely citation, the claims of A are likely to be broader in scope than the claims of B, and patent C must therefore be distinguished from both A and B to be patentable. If the citation is dangerous, on the other hand, one may not need to distinguish patent C from patent A. Thus, on average, we would expect a group of patents in a high-transaction cost thicket to have a higher value of the variant of transitivity illustrated in Figure 12c. We plan to investigate whether the Figure 12c transitivity will provide insight into the extent to which patents in a given technical field tend to be competing substitutes or blocking patent complements. If we are successful in identifying a structural measure of this kind, it may be both of analytical use in understanding the prevalence of blocking patents in particular technical fields and of practical use in evaluating the potential anti-competitive effects of patent pools and cross-licensing agreements.

## VI. CONCLUSIONS

This Article began by arguing that network science is poised to begin making important contributions to legal scholarship by offering relevant concepts, empirical methods, and modeling techniques that will be better able than current approaches to account for the importance of heterogeneity and local network structure in determining collective behavior. Be-

---

144. Podolny & Stuart, *supra* note 126, at 1229.

145. Maurseth, *supra* note 92.

146. *See infra* Appendix, Figure 12c.

cause networks are ubiquitous in the social problems to which the law addresses itself, network science, we argued, may make a significant contribution to legal analysis. Network science is particularly promising for dealing with situations in which local relationships are important and heterogeneous. It provides means to illuminate the complicated relationship between local relationship patterns and structure and interactions and global, collective behavior.

In the second half of this Article, we illustrated how one can apply the network approach to study the patent citation network. In Part III we showed empirically that, not only has the number of patents been increasing rapidly in recent years, but citation patterns have been changing as well. The average citability of a patent increases very rapidly with the number of times it has already been cited, demonstrating the preferential attachment, or “rich get richer,” phenomenon observed in many complex networks. Moreover, the extent to which highly cited patents are more “citable” than less cited patents has changed over time. Citability has become more stratified since the late 1980s. Since citability is likely related to a patent’s technical importance, the increasing stratification suggests that the patentability standard may have decreased, resulting in the issuance of a larger fraction of more trivial patents. Neither a general increase in the pace of technological progress, nor a general broadening of patented technology seems to explain the increasing stratification of patent citability. In fact, the average number of citations that a patent makes has increased over time, suggesting that patents inhabit a more and more densely crowded technological space on average.

The increasing stratification began in the late 1980s, a period which has no obvious connection with legal changes related to biotechnology, software, or business methods patenting, but is around the time that the “suggestion or motivation to combine” test for determining whether a claimed invention is obvious became increasingly established as the definitive Federal Circuit approach.<sup>147</sup> One possible hypothesis, based on our study so far, is that a weakening of the nonobviousness requirement, per-

---

147. See, e.g., *In re Raynes*, 7 F.3d 1037, 1039 (Fed. Cir. 1993) (“teaching, suggestion or motivation”); *In re Oetiker*, 977 F.2d 1443, 1447 (Fed. Cir. 1992) (“reason, suggestion, or motivation”); *In re Geiger*, 815 F.2d 686, 688 (Fed. Cir. 1987) (“teaching, suggestion or incentive”); *ACS Hosp. Sys., Inc. v. Montefiore Hosp.*, 732 F.2d 1572, 1577 (Fed. Cir. 1984) (“suggestion or incentive”). While the “suggestion test” was first applied by the Federal Circuit around 1983 when it began issuing opinions, the number of cases in which the test was applied increased dramatically during the 1990s. An informal survey of Federal Circuit case law finds more than three times as many cases citing the suggestion test in 2000 as in 1990. See *infra* Appendix, Figure 11.

haps associated with the suggestion test, has given rise to a proliferation of patents on minor improvements over the prior art. These improvements typically have little impact on the development of the technology and are thus cited only for a brief period after issuance. We cannot rule out other explanations for the change in the pattern of citation network evolution, however.

In Part IV we discussed how network-based analysis provides insights into the innovative process. For example, consistent with earlier statistical studies of citation lags, we found that the probability that a patent will be cited peaks at a relatively young patent age. But the citation probability also has a long, slow power law decay at older ages, suggesting that some patents retain their influence over very long times.<sup>148</sup> Surprisingly, we see this persistent vitality of some older patents even for patents which have received few or no citations—thus even “unpopular” patents have a significant probability of being revived after a long period of dormancy. This complicated dependence on patent age is evidence of the highly nonlinear nature of innovation and suggests that there may be two different innovation types. First, there is incremental innovation, for which the standard models of patents as incentives to investment may be relevant. Second, there is unpredictable innovation, which may be less susceptible to a simple incentive theory and more in line with “percolation” models of invention, in which incremental progress is coupled with complicated linkages back to previous technology.<sup>149</sup> Social scientists have begun to apply social-network-based approaches, looking at local network structure, to validate theories of innovation based on concepts of search and recombination. It should be possible to incorporate and build upon these studies in the analysis of innovation policy from a legal perspective.

In Part V we reported some preliminary results of other network-based approaches to understanding the patent system. We suggested that the concept of network path length or “distance” may be a fruitful means to explore the connections between different technical fields. One could use a network measure of “technological distance” to evaluate the USPTO classification scheme and as an alternative and more quantitative way to classify patents and to evaluate concepts such as the patent law concept of analogous arts that identifies the prior art that courts and examiners must consider in assessing nonobviousness.

---

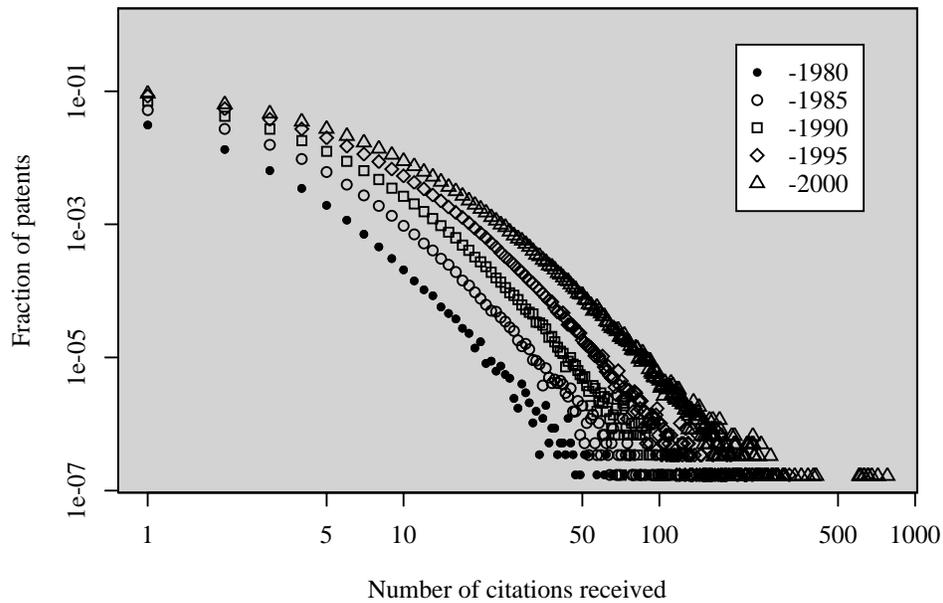
148. This long tail in the probability of citation as a function of patent age is masked to some extent by the rapidly increasing number of younger patents in earlier studies which simply compute average distributions of citation lags. Our dynamic analysis disentangles these two effects.

149. Silverberg & Verspagen, *supra* note 124.

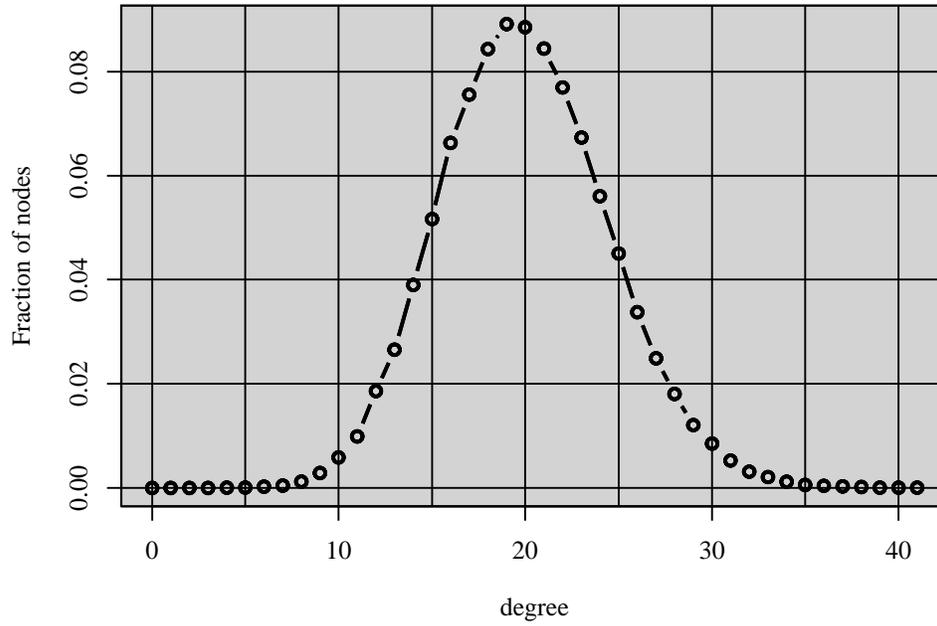
Another area in which a network approach shows promise is the evaluation of the extent to which increased patenting relates to the emergence of patent thickets and blocking patents. Network-based measures of local structure may be able to distinguish on average between patent thickets and areas of patenting of competing technologies. This approach may even help to distinguish between socially valuable patent pools and anti-competitive cross-licensing arrangements.

In this Article we have provided only a small sample of the application of network science to legal problems. Much remains to be done in applying network science to the patent system and to other legal issues.

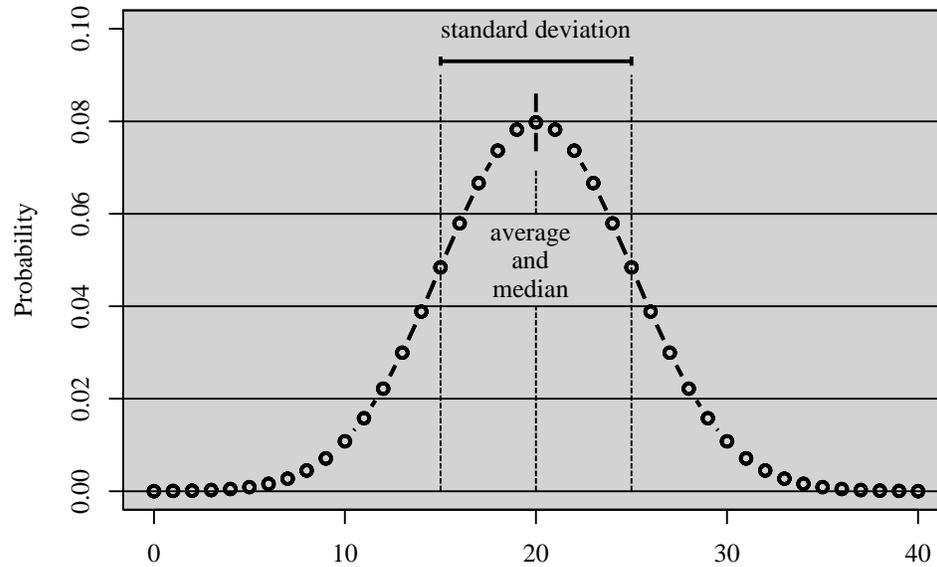
## APPENDIX



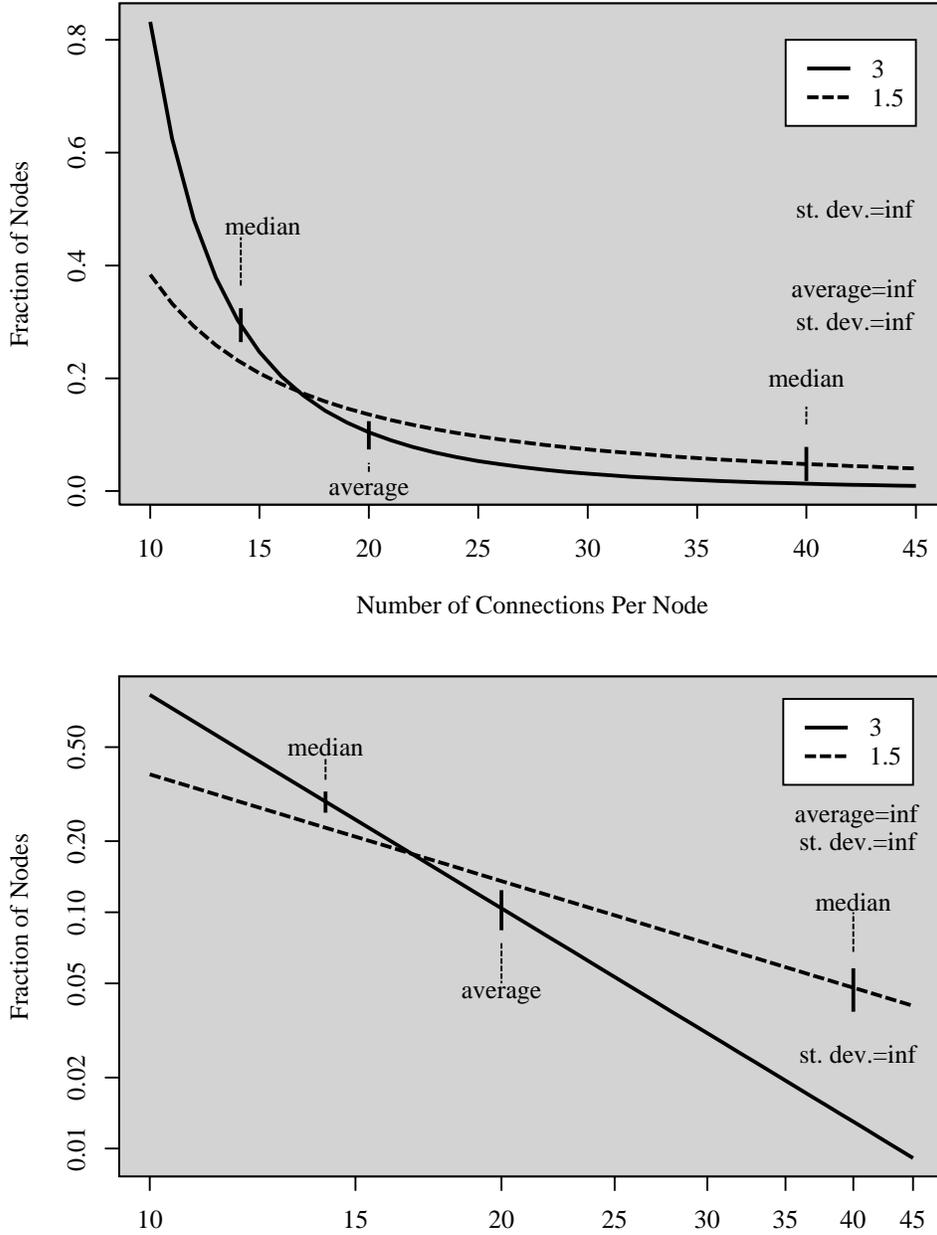
**Figure 1:** Distributions of number of citations received for the patent citation network at various times, shown on a log-log plot. A scale-free, power law distribution would be a straight line on this plot. Note that here the curvature increases as the network grows larger, inconsistent with the predictions of pure preferential attachment and consistent with the effects of patent aging.



**Figure 2:** The degree distribution for a random network. The shape is a Poisson distribution.

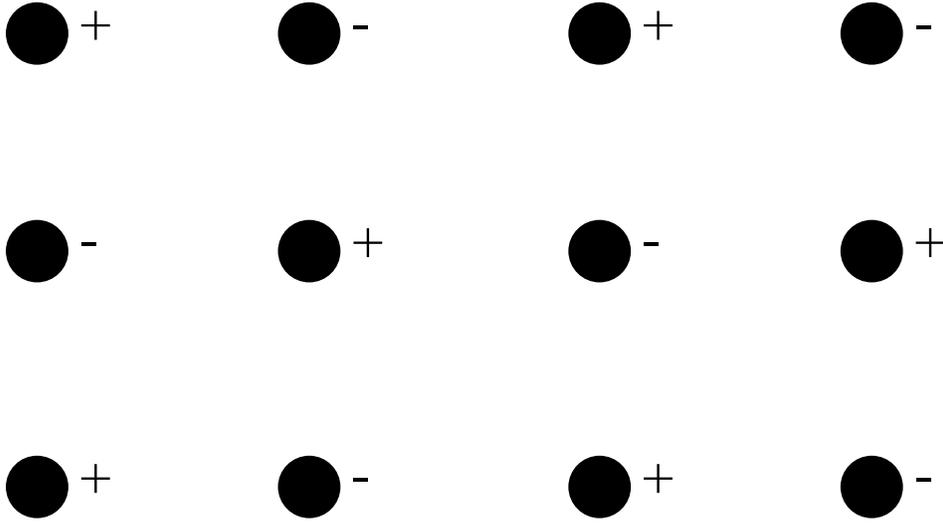


**Figure 3:** This figure illustrates the "normal" distribution or "Bell curve" that characterizes many probability distributions. It is symmetric, strongly peaked, and has a well-defined mean and median which are equal to one another. The probability outside of a few standard deviations from the mean is very small.

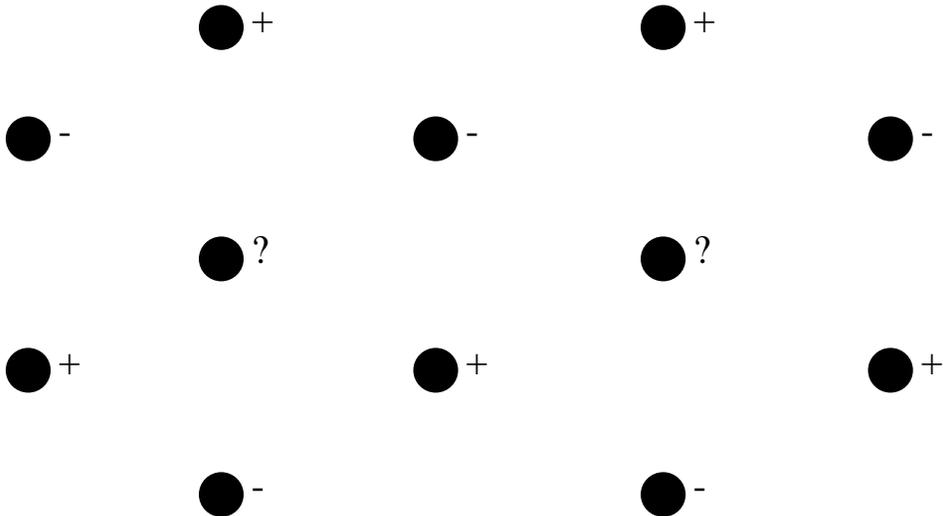


**Figure 4:** Top: This figure illustrates power law or “scale-free” probability distributions of node degree for two values of the decay exponent. Unlike the normal distribution, these distributions are highly skewed and cannot be meaningfully characterized by a “typical” value. Depending upon the decay exponent, the mean and median values can be quite far from one another. Indeed, if the decay exponent is small enough, the mean (or average) value is infinite even though the most likely value is zero. The standard deviation is infinite for both of the exponent values shown. Bottom: The same distributions are shown on a “log-log” plot, in which power law distributions are straight lines.

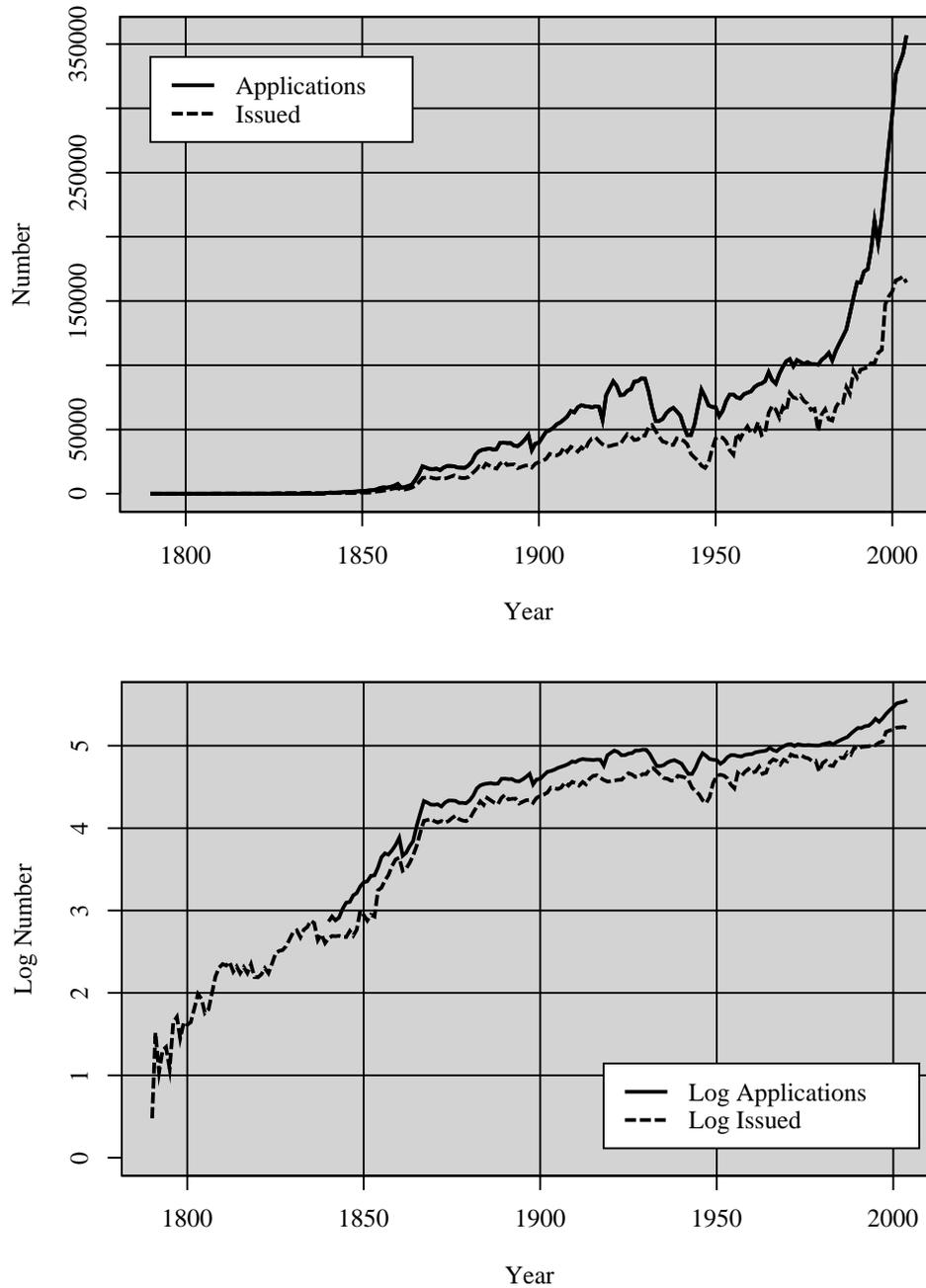
**Anti-Ferromagnetic Ising Model on a Square Lattice—All Neighbor Pairs Can Disagree**



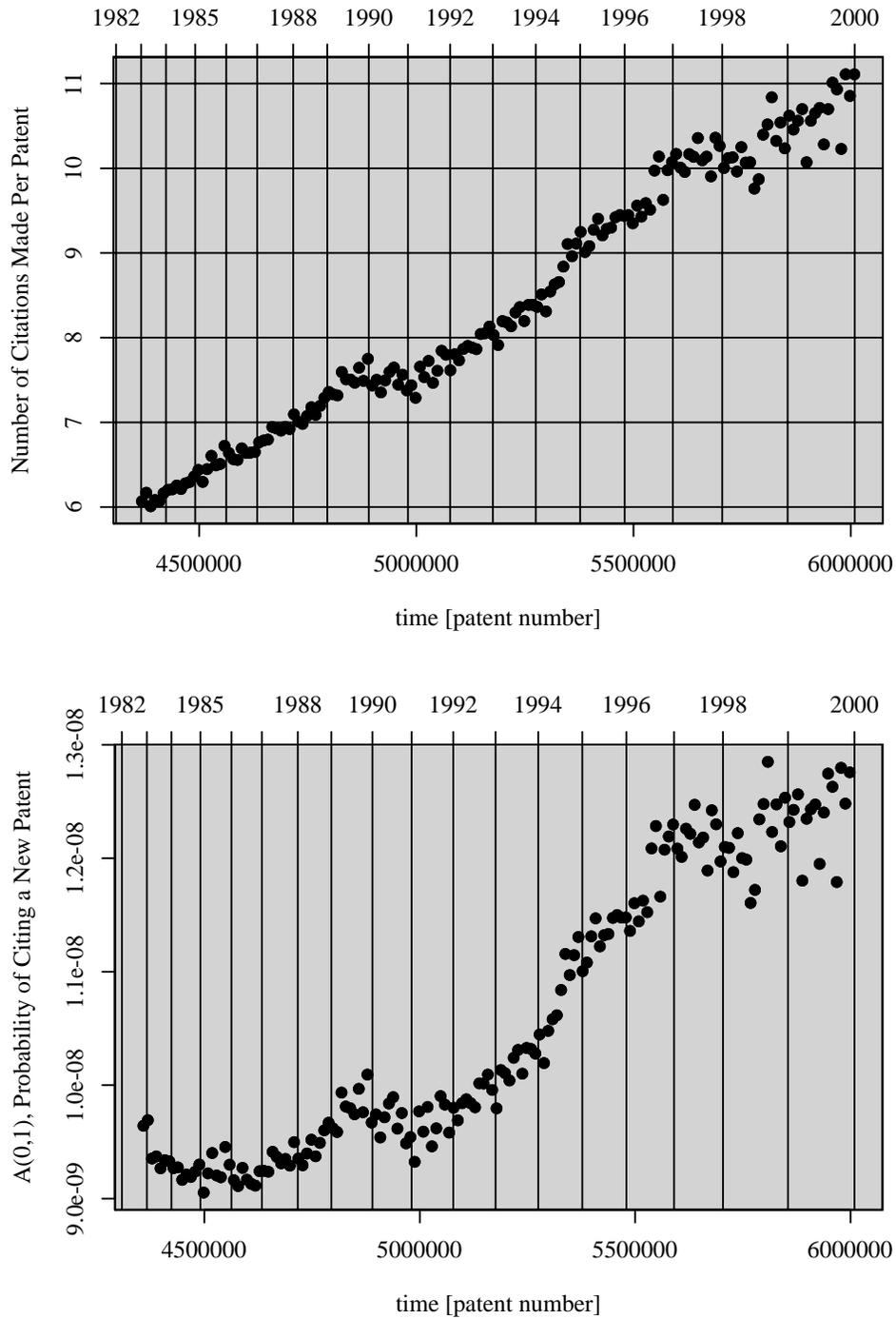
**Anti-Ferromagnetic Ising Model on Triangular Lattice—Some Neighbor Pairs Must Agree**



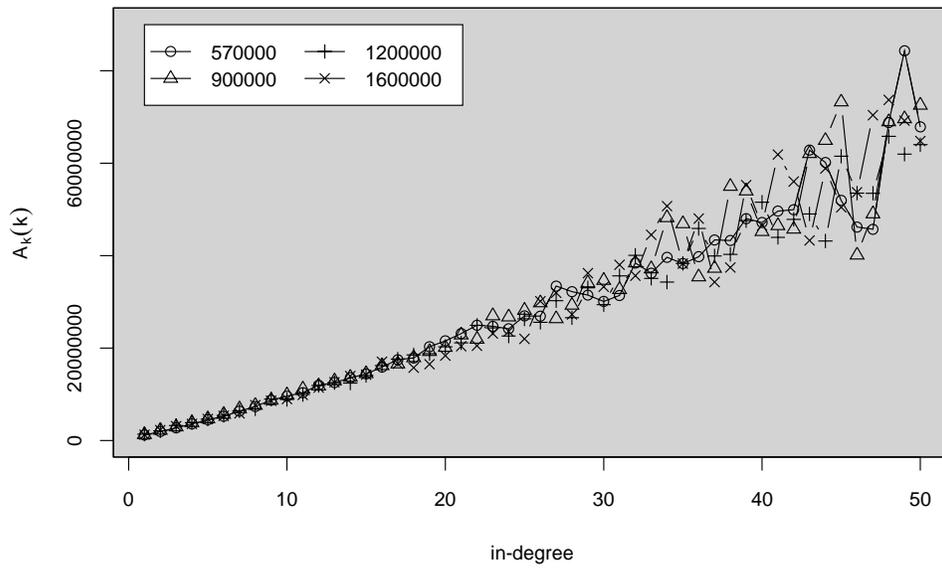
**Figure 5:** Illustration of how local structure may affect optimal collective behavior even when the preferred relationships between local pairs are the same.



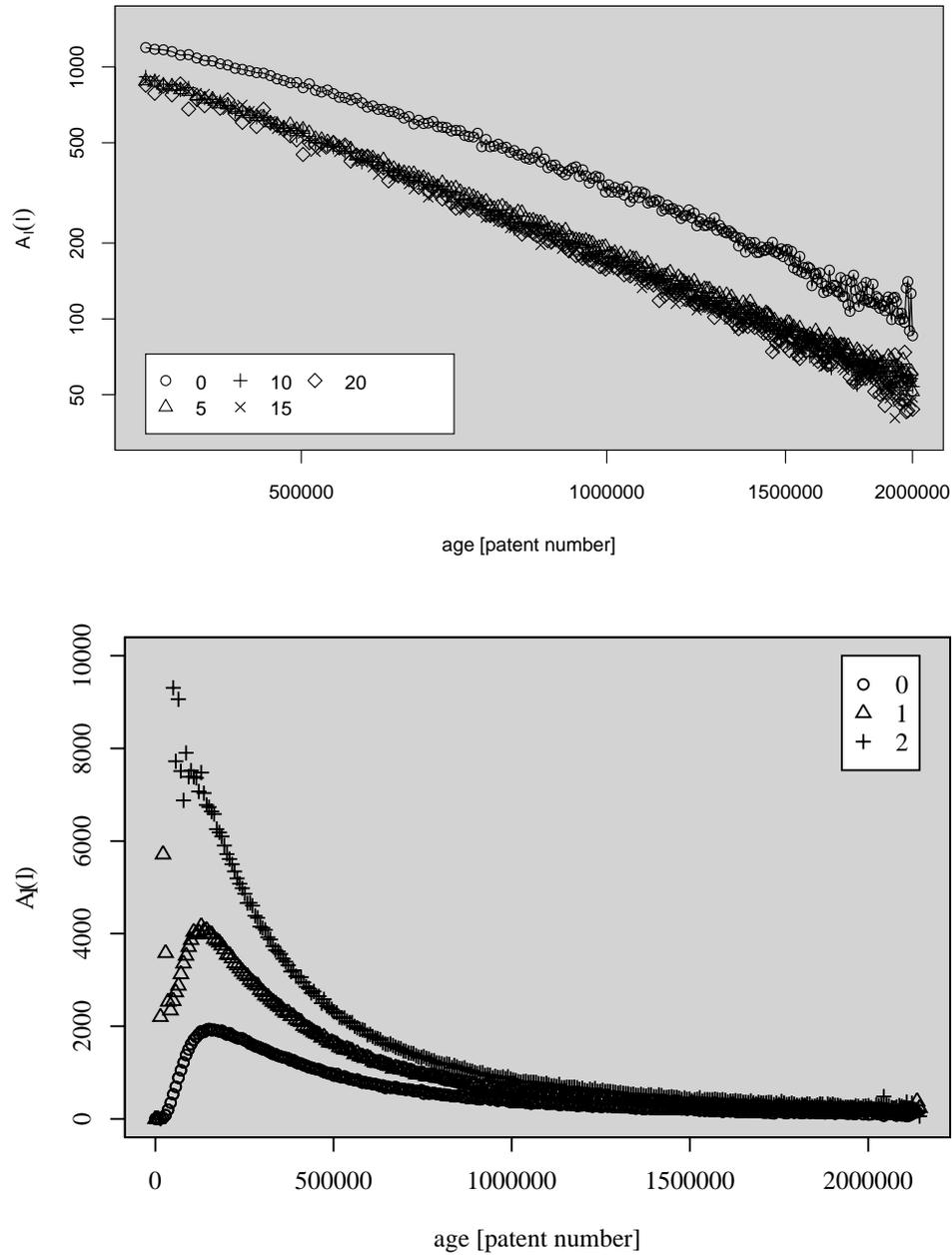
**Figure 6:** Top: Number of patents issued and applications filed per year as a function of time on a linear scale. Bottom: Number of patents issued and applications filed per year on a semi-log scale where linear portions correspond to exponential growth. (Note that patents were issued without examination from 1793-1836.)



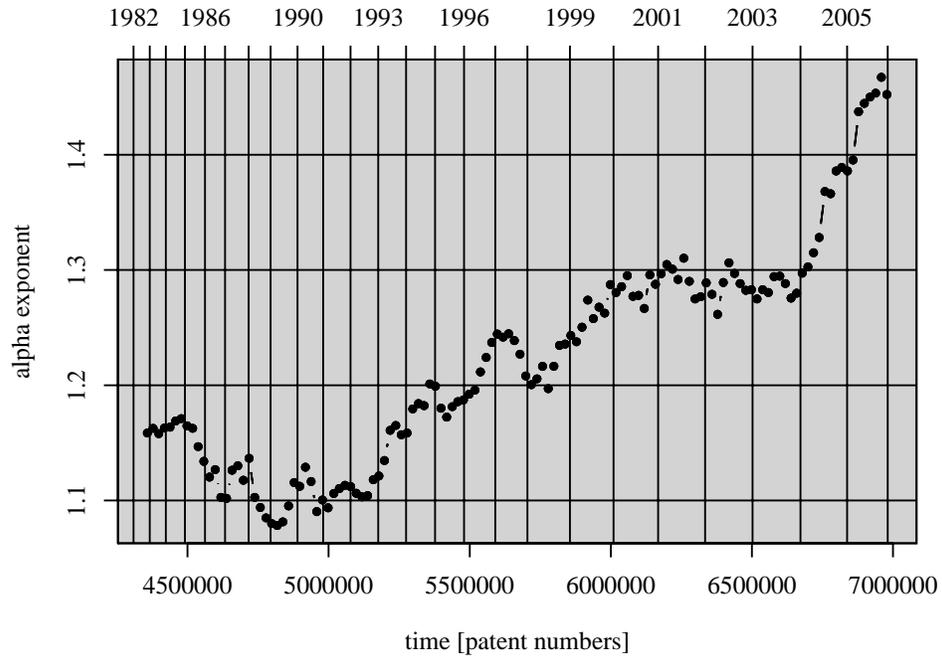
**Figure 7:** Top: Average number of citations made per patent as a function of time in patent number units. Bottom: Probability that a new patent will be cited as a function of time in patent number units.



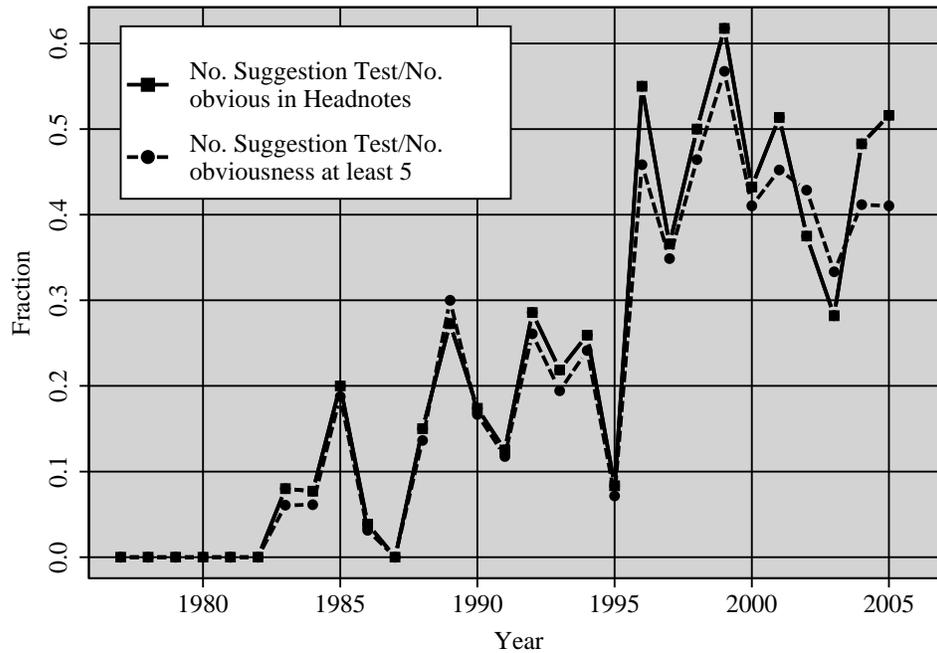
**Figure 8:**  $A_k(k)$  as a function of number of citations received,  $k$ , for various values of patient age,  $l$ .



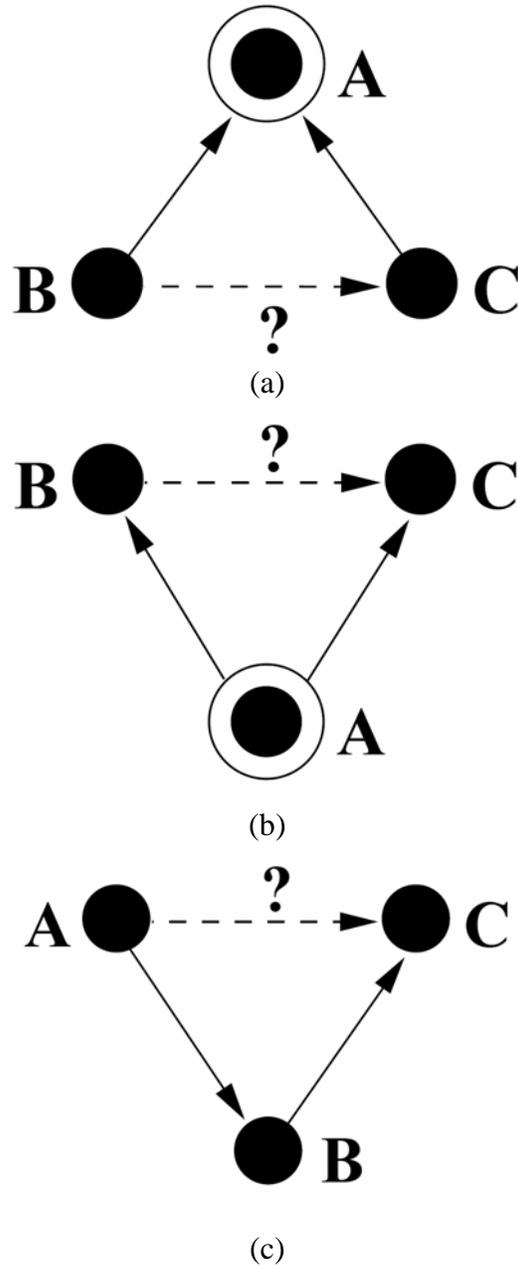
**Figure 9:** Top:  $A(k,l)$  as a function of patent age  $l$ , in units of patent number, for various values of in-degree,  $k$ . Bottom: The decaying portion of  $A_l(l)$  as a function of patent age  $l$ , for various values of  $k$ .



**Figure 10:** Stratification parameter,  $\alpha$ , as a function of time.



**Figure 11:** Fraction of Federal Circuit cases involving obviousness that referred to the “suggestion, teaching, or motivation to combine” test as a function of time. These numbers were obtained using LEXIS searches for “(suggestion or motivation or teaching) w/s combine” (to count references to the suggestion test) and either a reference to “obvious” in the headnotes (dashes) or at least 5 uses of the word “obvious” in the case (diamonds). These two methods of counting yield the same qualitative results showing an increase in use of the suggestion test throughout the 1990s.



**Figure 12:** Illustrations of definitions of transitivity relevant to (a) generality; (b) originality; (c) whether citation from B to C likely indicates a “design around” (“dangerous” citation) or a blocking patent (“lovely” citation).

# THE SECRET LIFE OF LEGAL DOCTRINE: THE DIVERGENT EVOLUTION OF SECONDARY LIABILITY IN TRADEMARK AND COPYRIGHT LAW

By Mark Bartholomen<sup>†</sup> & John Tehranian<sup>‡</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1364
II. THE DIVERGENCE OF SECONDARY LIABILITY THEORIES IN TRADEMARK AND COPYRIGHT LAW .....	1366
A. TRACING THE ORIGINS OF SECONDARY LIABILITY .....	1366
B. COMPARING VICARIOUS LIABILITY IN TRADEMARK AND COPYRIGHT LAW .....	1369
1. <i>The Nature of the Relationship</i> .....	1370
a) Principal-Agent Requirements in Trademark Law.....	1370
b) The Right and Ability to Supervise in Copyright Law.....	1373
2. <i>The Notion of Financial Benefit</i> .....	1374
a) Direct Financial Benefit in Trademark Law .....	1374
b) Expanding Notions of Financial Benefit in Copyright Law	1375
3. <i>Differences in Vicarious Liability Doctrine in Practice</i> .....	1377
C. COMPARING CONTRIBUTORY LIABILITY IN TRADEMARK AND COPYRIGHT LAW .....	1378
1. <i>Actual and Imputed Knowledge</i> .....	1379
a) The Scope of Imputation in Trademark Law.....	1379
b) Imputed Knowledge, Active Inducement, and the <i>Sony</i> Safe Harbor in Copyright Law .....	1382
2. <i>Material Contribution: Relationships Suitable for Contributory             Liability</i> .....	1388
a) The Direct Control Requirement in Trademark Law .....	1389
b) The Attenuated Notion of Control in Copyright Law .....	1391

---

© 2006 Mark Bartholomew and John Tehranian

<sup>†</sup> Visiting Associate Professor of Law, State University of New York, University at Buffalo Law School, bartholo@buffalo.edu.

<sup>‡</sup> Associate Professor of Law, University of Utah, S.J. Quinney College of Law, tehranianj@law.utah.edu. The authors would like to thank the organizers of and participants in the 2006 Intellectual Property Scholars Conference, co-hosted by the Berkeley Center for Law & Technology, Boalt Hall School of Law, and the Stanford Program in Law, Science and Technology, Stanford Law School, for their insightful comments and suggestions. John Tehranian would like to thank the University of Utah College of Law Research Fund for its support for this project.

III. UNDERSTANDING THE DIVERGENT EVOLUTION OF SECONDARY TRADEMARK AND COPYRIGHT LIABILITY.....	1394
A. DIFFERENCES IN THE TRADEMARK AND COPYRIGHT PROPERTY BUNDLES.....	1397
1. <i>Sources of Origin</i> .....	1397
2. <i>Differences in Scopes of Protection</i> .....	1400
B. CONCERNS WITH CHILLING BEHAVIOR OF INDIRECT PARTICIPANTS .....	1402
C. COPYRIGHT PANIC .....	1403
1. <i>Early Judicial Responses to the Digital Era: Copyright vs.             Trademark</i> .....	1405
2. <i>“The Unlawful Objective Was Unmistakable”: Peer-to-Peer             File Sharing, Grokster, and the Fundamental Transformation             of the Secondary Copyright Regime</i> .....	1407
3. <i>The Dangers of Panic</i> .....	1412
a) <i>The Problematic Implications of Grokster</i> .....	1412
b) <i>The Potential Expansion of Secondary Trademark                 Liability: Panic or Sound Policy?</i> .....	1414
IV. TOWARDS A BETTER SECONDARY LIABILITY REGIME .....	1417

## I. INTRODUCTION

Today’s intellectual property owners face unprecedented rates of copyright and trademark infringement. The widespread availability of digital technology and broadband internet access has enabled individuals in the most remote regions of the world to violate intellectual property rights. To shield themselves from liability, infringers have exploited the shortcomings of the post-Westphalian international legal regime and have relied on shadowy shell corporations, anonymizing technologies, and the impracticality and high cost of litigation.

As a result, stakeholders have had to fine-tune their litigation tactics to enforce their rights. Instead of pursuing the direct infringers on peer-to-peer file sharing networks, the recording industry set its sights on “secondary” infringers—operators of the networks and the distributors of the software that enabled users to reproduce copyrighted materials without authorization.<sup>1</sup> As intellectual property owners have increasingly turned to secondary liability theories, the courts have responded by enunciating substantial reinterpretations of extant principles, thereby precipitating a veritable secondary liability revolution. Numerous commentators have be-

---

1. See, e.g., *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2764, 2770 (2005); *In re Aimster Copyright Litig.*, 334 F.3d 643, 651 (7th Cir. 2003); *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1011, 1027 (9th Cir. 2001).

moaned this trend, contending that judicial recasting of liability rules has dramatically expanded intellectual property rights beyond their intended scope, resulting in an overprotective regime that stifles innovation.

Yet one of the most striking aspects of the secondary liability revolution has been all but ignored in the literature. While recent years have witnessed a dramatic broadening of the scope of secondary liability principles with respect to copyright law, no such move has occurred in the trademark arena. This divergence between trademark and copyright law is unusual for two reasons. First, secondary theories of liability in both trademark and copyright law share the same origins—the common law of tort and agency. Second, digital technology appears to pose just as much of a threat to trademark holders as to copyright interests because digital technology eases the reproduction of marks and facilitates the global distribution of infringing products. Nevertheless, the courts continue to police vigorously the metes and bounds of secondary trademark liability, even narrowing it at times, while simultaneously broadening the ambit of secondary copyright liability. This flux has created tremendous legal uncertainty that threatens investment in new technologies.

This Article takes a critical first step in clearing the murky waters of secondary infringement by setting forth and analyzing the divergence between the secondary trademark and copyright liability regimes. Part II disaggregates the various theories of secondary liability by analyzing the current law of contributory and vicarious trademark and copyright infringement. Despite common origins, trademark and copyright law have diverged over the years. Although many courts have recognized this divergence, we argue that they have not carefully parsed out the differences, blindly accepting them without serious scrutiny or rationalization.

Part III attempts to explain why the courts have created a two-tier system of secondary liability. In so doing, it examines what the divergent path of secondary trademark and copyright liability principles says about the law-making process, the evolution of legal doctrine, and the choices being made between two complementary systems of intellectual property protection. Our analysis reveals that neither fundamental differences in the nature or origin of trademark and copyright, rational balancing of economic risk-bearing, nor notions of romantic authorship have precipitated this bifurcation. Rather, a panic over copyright infringement in the digital age has beset the courts, causing the injudicious and often uncritical expansion of secondary liability principles in the copyright arena.

Part IV assesses how the law of secondary trademark and copyright liability fails to lay a reasonable template for resolving complex issues of technological change. We conclude the Article by discussing how future

scholarship may shed light on appropriate reforms to the secondary liability regime.

## II. THE DIVERGENCE OF SECONDARY LIABILITY THEORIES IN TRADEMARK AND COPYRIGHT LAW

### A. Tracing the Origins of Secondary Liability

Secondary liability—the imposition of liability on a defendant who did not directly commit the violation at issue<sup>2</sup>—originates in tort law.<sup>3</sup> A, for example, encourages direct participant B to throw rocks during a riot. B throws a rock that injures victim C. Even though A does not throw any rocks himself, A is still subject to liability to C as a contributory tortfeasor.<sup>4</sup> Contributory liability springs from the principle that certain parties should be held responsible for harms even if they are not the direct cause of the harm. Courts often rationalize secondary liability on economic efficiency grounds, viewing it as a means to shift injury costs to those who are in a position to prevent future injuries.<sup>5</sup> Others justify secondary liability on a moral basis: those who intentionally act to bring about tortious conduct should be held accountable, even if their actions are not the direct cause of harm to the victim.<sup>6</sup>

Secondary liability comes in two forms: vicarious liability and contributory liability. Vicarious liability does not require knowledge of the tortious act. Rather, the defendant is liable strictly because of his or her relationship with the direct tortfeasor. Unlike contributory liability, vicarious liability does not expand tort law to proscribe forms of conduct outside of the tort at issue. In fact, the conduct of the accused tortfeasor is not at issue in assessing vicarious liability. Instead, courts broaden liability for

---

2. WILLIAM L. PROSSER, LAW OF TORTS § 69 (5th ed. 1984).

3. See 4 J. THOMAS MCCARTHY, MCCARTHY ON TRADEMARKS AND UNFAIR COMPETITION § 25:23 (4th ed. 2005); Sverker K. Högberg, Note, *The Search for Intent-Based Doctrines of Secondary Liability in Copyright Law*, 106 COLUM. L. REV. 909, 914 (2006).

4. RESTATEMENT (SECOND) OF TORTS § 876, cmt. b, illus. 4 (1979).

5. THOMAS H. KOENIG & MICHAEL L. RUSTAD, IN DEFENSE OF TORT LAW 22 (2001) (explaining that at early common law, masters were strictly liable for their servants' torts because "the master was in the best position to prevent his servants' wrongdoing by proper supervision, training, and discipline"). See generally GUIDO CALABRESI, THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS (1970) (describing a theory of general deterrence, whereby potential tortfeasors factor in the cost of their accident-producing behavior when choosing which activities to undertake).

6. RESTATEMENT (SECOND) OF AGENCY § 212, cmt. a (1958) (stating the "general rule . . . that one causing and intending an act or result is as responsible as if he had personally performed the act or produced the result").

the original tort by imposing a penalty on an additional, albeit innocent, defendant.<sup>7</sup>

The most common test used to determine vicarious liability is control or the right to control the direct tortfeasor. The master-servant relationship is one example of a relationship between a principal and agent under the law of agency.<sup>8</sup> The law typically holds the master liable for the tortious acts of her servant if the servant acted within the scope of his employment.<sup>9</sup> Unlike individual employees, for whom one tort liability verdict might financially crush, employers can distribute tort losses by raising prices or by securing liability insurance.<sup>10</sup>

Under the doctrine of contributory liability, parties other than the direct tortfeasor may be held jointly and severally liable if they acted in concert with or provided assistance or encouragement to the direct tortfeasor.<sup>11</sup> The indirect participant's assistance must be "substantial." This means that there must be evidence that the contributory tortfeasor's actions helped cause the tortious act.<sup>12</sup> In addition, knowledge is required for contributory liability: the contributory tortfeasor must purposefully assist the performance of a tortious act.<sup>13</sup> Thus, the contributory tortfeasor must recognize that the direct tortfeasor's conduct constituted a breach of duty.<sup>14</sup>

Courts have recognized the availability of both common law theories of secondary liability—contributory and vicarious—in assisting content creators and trademark holders in their legal battles against facilitators of intellectual property infringement.<sup>15</sup> Both secondary liability theories re-

7. See PROSSER, *supra* note 2, § 69; AT&T Co. v. Winback & Conserve Program, Inc., 42 F.3d 1421, 1430-31 (3d Cir. 1994).

8. Lewis A. Kornhauser, *An Economic Analysis of the Choice Between Enterprise and Personal Liability for Accidents*, 70 CALIF. L. REV. 1345, 1346 (1982); RESTATEMENT (SECOND) OF AGENCY § 220(1) (1958).

9. 5 JAMES GRAY HARPER, THE LAW OF TORTS § 26.3 (2d ed. 1986).

10. *Id.* §§ 26.1, 26.5; *see also* Harris v. Trojan Fireworks Co., 120 Cal. App. 3d 157, 162 (1981).

11. Hilmes v. Stroebel, 17 N.W. 539, 539 (Wis. 1883) ("But any encouragement or aid given the principal actor, any concert of action in the execution of the unlawful design, will amount to a guilty participation in the trespass.").

12. PROSSER, *supra* note 2, § 41 at 240.

13. 3 HARPER, *supra* note 9, § 10.1.

14. 1 STUART M. SPEISER ET. AL, THE AMERICAN LAW OF TORTS § 3:4, at 401 (1983).

15. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844 (1982), represents the seminal case in secondary trademark liability jurisprudence. In *Ives*, the Supreme Court confirmed the application of secondary liability principles to trademark law by holding that a trademark owner could hold the manufacturer of a generic drug contributorily liable for

quire an underlying act of direct infringement. Contributory liability then attaches where there also exists (1) the defendant's knowledge of the infringement; and (2) the defendant's material contribution to the infringement.<sup>16</sup> Vicarious liability, as an outgrowth of the respondeat superior doctrine, requires (1) the right and ability of the defendant to control the actions of the infringer; and (2) a direct financial benefit to the defendant from the infringement.<sup>17</sup>

Despite their common genesis and shared language, copyright and trademark theories of secondary liability increasingly encompass divergent activities. The Supreme Court has explicitly refused to apply the standard it set for trademark contributory liability to cases of secondary copyright infringement,<sup>18</sup> arguing that trademark law has "little or no analogy" to copyright,<sup>19</sup> and that "fundamental differences" exist between the two bodies of law.<sup>20</sup> The lower courts have heeded these words, emphasizing the need to evaluate liability under two different standards depending on whether a copyright or trademark is at issue.<sup>21</sup>

---

the actions of pharmacists. *Id.* at 853-54. While not elaborating on the justification for importing tort principles into the federal trademark regime, the Court affirmed that liability for trademark infringement can extend past those who actually "use" a protected mark by imposing indirect liability on Inwood. *Id.* Similarly, in *Kalem Co. v. Harper Bros.*, 222 U.S. 55 (1911), the Supreme Court affirmed the application of secondary liability doctrines to copyright infringement. *Id.* at 63. The Court held that the producer of an unauthorized film dramatization of the copyrighted book *Ben Hur* was liable for his sale of the film to middlemen who arranged for the film's commercial exhibition. The Court explained that although the producer did not take part in the final act of infringement—the exhibition of the infringing film to paying customers—his contribution was sufficient to make him secondarily liable. *Id.* Although *Ives* and *Kalem Co.* involved contributory liability claims, the decisions imply that both types of secondary liability theories—contributory and vicarious—are available to copyright and trademark plaintiffs.

16. *See Gershwin Publ'g Corp. v. Columbia Artists Mgmt., Inc.*, 443 F.2d 1159, 1162 (2d Cir. 1971).

17. *See id.*

18. *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 439 n.19 (1984).

19. *Id.* (quoting *United Drug Co. v. Rectanus Co.*, 248 U.S. 90, 97 (1918)).

20. *Id.*

21. *E.g.*, *Hard Rock Café Licensing Corp. v. Concession Servs., Inc.*, 955 F.2d 1143, 1150 (7th Cir. 1992) ("[T]he Supreme Court tells us that secondary liability for trademark infringement should, in any event, be more narrowly drawn than secondary liability for copyright infringement."); *United States v. Wash. Mint, LLC*, 115 F. Supp. 2d 1089, 1107 (D. Minn. 2000) (referring to "the more narrow standards applicable to [indirect] trademark infringement claims"); *Lockheed Martin Corp. v. Network Solutions, Inc.*, 985 F. Supp. 949, 965 (C.D. Cal. 1997) ("Because the property right protected by trademark law is narrower than that protected by copyright law, liability for contributory infringement of a trademark is narrower than liability for contributory infringement of a copyright."); *Banff Ltd. v. Limited, Inc.*, 869 F. Supp. 1103, 1111 (S.D.N.Y. 1994)

At the same time, however, the courts have failed to identify with any care the specific divergences between secondary trademark and copyright liability. They have repeatedly conflated the contributory and vicarious doctrines in general.<sup>22</sup> Although concerned with the proper application of these doctrines, commentators have recognized the inconsistent standards for secondary liability in copyright and trademark law and have largely yielded to these doctrinal distinctions without qualm or scrutiny.<sup>23</sup> Moreover, neither the courts nor scholars have fully explored the underlying justifications for the bifurcation. As we will demonstrate, the legal standards for secondary trademark infringement differ markedly from those applied to vicarious and contributory copyright infringers. A careful disaggregation of the secondary liability doctrines suggests that, while the courts continue to ground trademark law in the traditional doctrine of common law secondary liability, they have abandoned copyright's common law moorings and reshaped copyright law to encompass a wider range of activities than those covered by traditional tort principles.

### **B. Comparing Vicarious Liability in Trademark and Copyright Law**

To succeed in a claim for vicarious liability, a plaintiff must demonstrate that the defendant has the right and ability to control the direct infringer and that the infringement translated into a direct financial benefit for the defendant. As discussed below, courts have interpreted these standards in a relaxed manner in copyright cases, but not in trademark cases. Copyright plaintiffs have succeeded by merely alleging an ability to supervise the direct tortfeasor. For trademark plaintiffs, though, the courts

---

("While it might be tempting to apply the standard articulated for copyright infringement, the Supreme Court has made it clear that liability for non-direct infringers under trademark law is narrower than liability under the copyright laws.").

22. *E.g.*, *Monsanto Co. v. Campuzano*, 206 F. Supp. 2d 1271 (S.D. Fla. 2002) (conflating the contributory and vicarious liability doctrines in general); *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 846, 854 (1982) (using the terms "vicariously" and "contributorily" interchangeably).

23. *E.g.*, Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777, 812 (2004) (referring to the doctrines of contributory infringement in patent and copyright law as "distant cousins" of contributory trademark infringement); *id.* at 829 ("Unlike patent and copyright law, the doctrine of contributory trademark infringement is narrowly drawn."); Ian C. Ballon, *Pinning the Blame in Cyberspace: Towards a Coherent Theory for Imposing Vicarious Copyright, Trademark and Tort Liability for Conduct Occurring Over the Internet*, 18 HASTINGS COMM. & ENT. L.J. 729, 750 (1996) ("As under copyright law, online service providers may be held contributorily or vicariously liable for trademark, service mark, or trade dress infringement, although the grounds for imposing indirect trademark liability are more narrow.").

demand evidence of a specific principal-agent relationship for vicarious trademark liability. And while courts require proof of a direct financial benefit from the tortious conduct in trademark cases, courts in copyright cases have virtually read the word “direct” out of the “direct financial benefit” requirement, permitting liability based on hypothetical future returns to the defendant.

1. *The Nature of the Relationship*

a) Principal-Agent Requirements in Trademark Law

For vicarious liability under either copyright law or trademark law, a sufficient link between the defendant and the alleged infringer must exist. But courts have increasingly required a stronger connection for vicarious trademark liability. Vicarious trademark liability relies on traditional tort and agency law principles to determine if a defendant should be held responsible for someone else’s direct infringement of a mark.<sup>24</sup> Vicarious liability results only when an agent acts on a principal’s behalf in committing trademark infringement.<sup>25</sup> A principal-agent relationship exists only if “the defendant and the direct infringer have an apparent or actual partnership, have authority to bind one another in transactions with third parties, or exercise joint ownership or control over the infringing product.”<sup>26</sup>

Vicarious trademark liability therefore has strict limits. Absent a principal-agent relationship between the defendant and the direct infringer, the defendant cannot face exposure to vicarious liability.<sup>27</sup> Other relationships will not give rise to a claim. Unlike in copyright, “courts do not recognize vicarious liability in the trademark context based on ability to supervise in combination with a financial interest.”<sup>28</sup>

---

24. *Fare Deals, Ltd. v. World Choice Travel.com, Inc.*, 180 F. Supp. 2d 678, 684 (D. Md. 2001). According to some, vicarious liability can also involve joint tortfeasors, i.e., parties who act “in concert” to commit a tort and are held jointly liable for all harm caused to the victim. See John T. Cross, *Contributory and Vicarious Liability for Trademark Dilution*, 80 OR. L. REV. 625, 650, 660 (2001).

25. 3 JEROME GILSON, *TRADEMARK PROTECTION AND PRACTICE* § 11.02(2)(h) (1974).

26. *Hard Rock*, 955 F.2d at 1150. A principal will even risk liability for its agent’s misrepresentations “upon matters which the principal might reasonably expect would be the subject of representations, provided the other party has no notice that the representations are unauthorized.” RESTATEMENT (SECOND) OF AGENCY § 258 (1958).

27. *Fare Deals*, 180 F. Supp. 2d at 686.

28. *United States v. Wash. Mint, LLC*, 115 F. Supp. 2d 1089, 1106 (D. Minn. 2000) (citing *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 439 (1984)).

A recent case from the Tenth Circuit colorfully illustrates this point.<sup>29</sup> Randy L. Haugen, a distributor of Amway products, had widely disseminated defamatory statements on Amway's e-mail distribution list. According to the urban folklore recited by Haugen, Amway's competitor, consumer products manufacturer Proctor & Gamble, was an agent of Satan. As Haugen asserted, Proctor & Gamble diverted a large portion of its profits to the Church of Satan and the company's logo, a ram's horn, formed a 666—Satan's fabled digits. Haugen even claimed that Proctor & Gamble's president had "come out of the closet"<sup>30</sup> about his association with the Church of Satan on an episode of the *Phil Donahue Show*. When asked if the revelations would hurt business, Haugen claimed that Proctor & Gamble's president had nonchalantly demurred, opining that "there are not enough Christians in the United States to make a difference."<sup>31</sup> Doubtlessly concerned with potential litigation, Amway asked Haugen to recant and he did. But Proctor & Gamble still sued both Haugen and Amway, claiming that the alleged association with Lucifer violated, *inter alia*, the Lanham Act because it constituted a "false or misleading representation of fact which . . . in commercial advertising or promotion misrepresents the nature, characteristics, [or] qualities . . . of . . . another person's goods, services or commercial activities."<sup>32</sup> In particular, Proctor & Gamble wanted Amway held vicariously liable for the actions of its distributor.

The Tenth Circuit reinstated Proctor & Gamble's Lanham Act claim against Haugen after the district court had dismissed it based on a narrow construction of the Lanham Act, but refused to reinstate a claim for vicarious liability against Amway.<sup>33</sup> Despite the fact that Amway supervised its distributors in a number of ways, including setting the parameters within which its distributors functioned and dedicating company resources to create uniform standards of behavior, the court found that the plaintiff had failed to demonstrate an employment or principal-agent relationship be-

---

29. *Proctor & Gamble Co. v. Haugen*, 317 F.3d 1121 (10th Cir. 2003). The summary of the facts in this and the following paragraph are drawn from the Tenth Circuit's previous opinion in the case, *Proctor & Gamble Co. v. Haugen*, 222 F.3d 1262 (10th Cir. 2000), and the district court opinion on remand, *Proctor & Gamble Co. v. Haugen*, 158 F. Supp. 2d 1286 (D. Utah 2001), *aff'd*, *Proctor & Gamble Co. v. Haugen*, 317 F.3d 1121 (10th Cir. 2003).

30. *Proctor & Gamble Co. v. Haugen*, 222 F.3d 1262, 1268 (10th Cir. 2000).

31. *Id.*

32. *See Proctor & Gamble Co. v. Haugen*, 317 F.3d 1121, 1124 n.3 (10th Cir. 2003) (quoting 15 U.S.C. § 1125(a)(1)(B)).

33. *See Proctor & Gamble Co. v. Haugen*, 222 F.3d 1262, 1276, 1278 (10th Cir. 2000).

tween Amway and Haugen.<sup>34</sup> Since Haugen's violating conduct was not naturally and ordinarily incident to Amway's business, the court refused to find Amway vicariously liable for Haugen's actions.<sup>35</sup> Thus, contractual relationships such as that between licensor and licensee or franchisor and franchisee are not sufficient to give rise to vicarious liability in trademark law.<sup>36</sup>

Of course, ambiguity exists in determining when a defendant has the necessary degree of control or authority over a direct infringer to establish a principal-agent relationship and to trigger vicarious liability. In *Haugen*, the court held that Amway was not liable for the acts of its distributors since there was no principal-agent relationship. Amway had not vested its distributors with the authority to act on its behalf.<sup>37</sup> However, in another recent case, a court found that an internet search engine could be vicariously liable for the infringing acts of its advertisers.<sup>38</sup> The advertisers purchased the marks of other companies as keyword search terms for their own products. The court denied the search engine's motion to dismiss, holding that an allegation that the search engine "exercise[d] significant control over the content of advertisements" was enough to state a claim for vicarious liability.<sup>39</sup> Thus, the amount of control necessary to make a defendant vicariously liable is imprecise, and subject to the interpretation of different courts. Nevertheless, the mere right and ability to supervise does not create vicarious liability in the trademark context.<sup>40</sup> More importantly, in trademark law, courts continue to couch the threshold relationship for vicarious liability as one of principal-agent. Consequently, they require proof of "significantly greater involvement with the infringement by the

---

34. *See id.* at 1278.

35. *Id.*

36. *See Mini Maid Servs. Co. v. Maid Brigade Sys., Inc.*, 967 F.2d 1516, 1519 (11th Cir. 1992) ("The law imposes no duty upon a franchisor to diligently prevent the independent acts of trademark infringement that may be committed by a single franchisee."); *Oberlin v. Marlin Am. Corp.*, 596 F.2d 1322, 1327 (7th Cir. 1979) (explaining that a licensor's duty to supervise its licensee's use of its trademark does not establish principal-agent relationship under state law). Such relationships may, however, make the defendant liable for *contributory* trademark infringement. *See infra* Section II.C.2.a.

37. *Procter & Gamble Co. v. Haugen*, 317 F.2d at 1127-28.

38. *Gov't Employees Ins. Co. v. Google, Inc.*, 330 F. Supp. 2d 700, 705 (E.D. Va. 2004).

39. *Id.* at 704.

40. *See, e.g., Wesley v. Don Stein Buick, Inc.*, 996 F. Supp. 1312, 1316 (D. Kan. 1998) (holding that a car manufacturer's ability to supervise a car dealership was insufficient to support a finding of vicarious trademark liability).

party against whom vicarious liability is sought than is required under the copyright laws.”<sup>41</sup>

b) The Right and Ability to Supervise in Copyright Law

Copyright law is markedly different. Courts do not require a principal-agent relationship to find vicarious liability.<sup>42</sup> Instead, as courts have repeatedly held, “one may be vicariously liable [for copyright infringement] if he has the right and ability to supervise the infringing activity and also has a direct financial interest in such activities.”<sup>43</sup> In copyright, courts do not require that the direct infringer be an agent of the defendant or that the defendant cause others to believe that the direct infringer is acting under his authority in order to expose a defendant to vicarious liability. A defendant may be guilty of vicarious copyright infringement even in the absence of an actual agency relationship.<sup>44</sup>

Thus, a restaurant owner faced vicarious liability when his hired musician violated copyright law, even though the musician served as an independent contractor, rather than as an employee.<sup>45</sup> Similarly, the owner of a racetrack suffered vicarious liability when a company hired to supply music over the track’s public address system violated copyright law.<sup>46</sup> In the decision, the court appealed to public policy to rationalize rejection of the racetrack owner’s “independent contractor” defense. The court stated, “The proprietor of a public establishment operated for a profit could oth-

41. *Banff Ltd. v. Limited, Inc.*, 869 F. Supp. 1103, 1111 (S.D.N.Y. 1994).

42. *Lowry’s Reports, Inc. v. Legg Mason, Inc.*, 271 F. Supp. 2d 737, 745 (D. Md. 2003) (“Vicarious copyright liability stems from the common law doctrine of respondeat superior. Unlike that doctrine, however, it does not depend on the existence of a master-servant or employer-employee relationship. Vicarious copyright liability extends more broadly.” (citations omitted)); Craig A. Grossman, *From Sony to Grokster: The Failure of the Copyright Doctrines of Contributory Infringement and Vicarious Liability to Resolve the War Between Content and Destructive Technologies*, 53 *BUFF. L. REV.* 141, 147 (2005).

43. *Gershwin Publ’g Corp. v. Columbia Artists Mgmt., Inc.*, 443 F.2d 1159, 1162 (2d Cir. 1971); *see also Pinkham v. Sara Lee Corp.*, 983 F.2d 824, 834 (8th Cir. 1992); 3 MELVILLE B. NIMMER & DAVID NIMMER, *NIMMER ON COPYRIGHT* § 12.04(A)(1) (2004) [hereinafter *NIMMER ON COPYRIGHT*].

44. *See AT&T v. Winback & Conserve Program, Inc.*, 42 F.3d 1421, 1439-40 (3d Cir. 1994).

45. *Warner Bros., Inc. v. Lobster Pot, Inc.*, 582 F. Supp. 478, 482 (N.D. Ohio 1984); *see also Realsongs v. Gulf Broad. Corp.*, 824 F. Supp. 89, 92 (M.D. La. 1993) (holding radio station owners vicariously liable for actions of minister disc jockeys who purchased airtime and played copyrighted songs over airwaves even though owners had instructed ministers not to play copyrighted materials).

46. *Famous Music Corp. v. Bay State Harness Horse Racing & Breeding Ass’n*, 554 F.2d 1213, 1214-15 (1st Cir. 1977).

erwise reap the benefits of countless violations by orchestras, itinerant or otherwise, by merely claiming ignorance that any violation would take place."<sup>47</sup> Yet such a defense absolves a similarly situated defendant from vicarious trademark liability.<sup>48</sup>

## 2. *The Notion of Financial Benefit*

At the same time that the courts have expanded the nature of the relationship necessary to trigger vicarious copyright liability, they have radically reworked the critical element of financial benefit, creating a divergence in its treatment between vicarious trademark and copyright doctrine. A comparison of two cases involving secondary liability claims against flea market owners for vendor infringements (trademark in the first case, copyright in the second) provides an ideal illustration of this point. Despite similar facts, the courts drew opposite conclusions on the issue of liability, largely due to unscrutinized differences between secondary trademark and copyright doctrine.

### a) Direct Financial Benefit in Trademark Law

In *Hard Rock Café Licensing Corp. v. Concession Servs.*,<sup>49</sup> the owner of the Hard Rock trademark sued a flea market owner, CSI, for both contributory and vicarious liability, contending that the owner was responsible for trademark infringement committed by a t-shirt vendor, Parvez. Parvez sold counterfeit Hard Rock t-shirts on the premises. Although not foreclosing the possibility of vicarious liability, the Seventh Circuit issued guidance to the lower court, noting that it was "inclined to favor [defendant] CSI's side of the dispute. CSI neither hired Parvez to entertain its customers . . . nor did it take a percentage of his sales."<sup>50</sup> While one could argue that the sale of infringing t-shirts brought more customers to the flea

---

47. *Id.* at 1215.

48. For example, in *Oberlin v. Marlin American Corp.*, 596 F.2d 1322, 1326-27 (7th Cir. 1979), the court held that no agency relationship existed between the defendant and an independent contractor that used the defendant's trademark. Plaintiff's argument that the Lanham Act created an agency relationship between the mark owner and the contractor was rejected. The court wrote:

The purpose of the Lanham Act, however, is to ensure the integrity of registered trademarks, not to create a federal law of agency. Furthermore, the scope of the duty of supervision associated with a registered trademark is commensurate with this narrow purpose. . . . [This duty] does not automatically saddle the licensor with the responsibilities . . . of a principal for his agent.

*Id.* at 1327.

49. 955 F.2d 1143 (7th Cir. 1992).

50. *Id.* at 1150 n.4 (citations omitted).

market (thereby increasing parking and admission fee revenues), boosted Parvez's profits (thereby enabling him to afford the vendor fee he paid to CSI), and ultimately inured to the financial benefit of the flea market owner, the court declined an opportunity to seize upon such an attenuated link between infringement and financial benefit for liability purposes. The court suggested that only an actual profit-sharing regime between the owner and the vendor or use of the direct infringer for customer/client entertainment purposes would create a sufficient nexus between acts of infringement and an owner's revenue stream to warrant vicarious liability.<sup>51</sup> In short, courts require obvious and direct financial benefit before they impose vicarious liability for trademark infringement.

b) Expanding Notions of Financial Benefit in Copyright Law

By sharp contrast, in *Fonovisa, Inc. v. Cherry Auction, Inc.*,<sup>52</sup> a copyright holder sued the operator of a flea market for the activities of one of its vendors who sold infringing recordings. The defendant flea market operator, Cherry Auctions, reaped substantial revenues from the concession stand fees it generated from third-party vendors and from the parking and admissions fees it collected from the public. The court found Cherry Auctions vicariously liable because the infringing activities of the record seller "enhance[d] the attractiveness of the venue to potential customers," luring them to the flea market grounds and driving up Cherry Auctions's revenues.<sup>53</sup>

As Lemley and Reese point out, the *Fonovisa* decision represents a startling expansion in the definition of financial benefit since the flea market did not directly profit from the sales of infringing recordings and received no percentage of the vendor's business. As they observe, "the existence of infringing activity is assumed to draw customers in greater numbers than noninfringing activity, and any money those customers pay to the defendant appears to count as revenue 'directly' related to the infringing activity for purposes of vicarious liability."<sup>54</sup> This causal chain linking infringement and profit is not only unsubstantiated but starkly different than the established precedent in trademark cases such as *Hard Rock Café*.

Moreover, the schism between trademark and copyright law on this point is widening. *Fonovisa* and its progeny constitute a significant departure from prior copyright doctrine, as earlier cases embraced a much more

---

51. *See id.*

52. 76 F.3d 259 (9th Cir. 1996).

53. *Id.* at 263.

54. Mark Lemley & R. Anthony Reese, *Reducing Digital Copyright Infringement Without Restricting Innovation*, 56 STAN. L. REV. 1345, 1368 (2004).

demure definition of financial benefit. In 1938, for example, Judge Augustus Hand immunized a landlord from vicarious copyright liability claims based on a tenant's actions: "Something more than the mere relation of landlord and tenant must exist to give rise to a cause of action by plaintiffs against these defendants for infringement of their copyright on the demised premises."<sup>55</sup> The court also pronounced its circumspect vision of financial benefit: "[The landlords] received nothing, and were not entitled to receive anything through [the tenant's] acts of infringement."<sup>56</sup> A quarter-century later, the Second Circuit reiterated this concept, arguing that the defendant must enjoy "an obvious and direct financial interest in the exploitation of copyrighted materials" before courts will impose vicarious liability.<sup>57</sup>

But in recent years, courts have significantly transformed the financial benefit component of the vicarious liability regime in copyright law. *Fonovisa* first subverted the constrained notion of financial benefit by imputing it from the mere draw of an audience to a site in which a defendant has an economic interest. In the wake of *Fonovisa*, the requirements for financial benefit in the copyright context have slackened even further. In *Napster*, the Ninth Circuit presented only a cursory analysis of the issue of financial benefit, arguing in a brief paragraph that Napster received financial benefit from the availability of infringing materials on its peer-to-peer file sharing network.<sup>58</sup> The court summarily concluded financial benefit based on its simple observation that the infringing materials served as a draw for customers: "Ample evidence supports the district court's finding that Napster's *future* revenue is directly dependent upon 'increases in user-base.' More users register with the Napster system as the 'quality and quantity of available music increases.'"<sup>59</sup> Remarkably, the Ninth Circuit failed to acknowledge that Napster had earned no revenue and had never charged its customers any fees.<sup>60</sup> Thus, on the purely hypothetical notion

---

55. *Deustch v. Arnold*, 98 F.2d 686, 688 (2d Cir. 1938).

56. *Id.*

57. *Shapiro, Bernstein & Co. v. H.L. Green Co.*, 316 F.2d 304, 307 (2d Cir. 1963) (imposing vicarious liability on an individual who received a share of the gross receipts from an infringer's sale of bootleg records).

58. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1023 (9th Cir. 2001). Peer-to-peer networks allow users—the direct infringers—to share copyrighted digital works from their home computers. *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2764, 2770 (2005).

59. *Id.* (emphasis added).

60. On motion for preliminary injunction, the lower court acknowledged that at the time of suit, Napster had not earned any revenue. See *A&M Records, Inc. v. Napster, Inc.*, 114 F. Supp. 2d 896, 902 (N.D. Cal. 2000).

of profitability—including the eventual monetization of its user base through e-mail, advertising, linking, and direct marketing—the court found financial benefit.<sup>61</sup> *Napster* therefore expands *Fonovisa*, imputing financial benefit from infringing activity that lures an audience to a virtual site, even one from which defendant does not draw revenue.

### 3. *Differences in Vicarious Liability Doctrine in Practice*

A recent case, *United States v. Washington Mint, LLC*,<sup>62</sup> illustrates how the different standards for secondary liability in copyright and trademark law work in practice. The United States government sued the Washington Mint, a private mint, for direct trademark and copyright infringement for manufacturing and selling replicas of the Sacagawea dollar.<sup>63</sup> The government also sued the company's marketing and advertising agency, Novus, as well as certain corporate officers, under theories of vicarious and contributory copyright and trademark infringement.

With respect to the vicarious liability claims, the court found that Novus exercised supervisory control over the Washington Mint and financially benefited from its infringing activities—the threshold requirements for vicarious copyright and trademark liability. Control existed because Novus employed a large number of the Mint's employees, including its Chief Executive Officer. The court inferred a financial benefit from the infringement because Novus served as the exclusive advertising space for the Mint's Sacagawea dollar.<sup>64</sup>

More significantly, the court explicitly bifurcated the issue of vicarious liability against the corporate officers, finding sufficient evidence of vicarious copyright liability but not vicarious trademark liability.<sup>65</sup> The corporate officers held roles as co-CEOs and co-presidents of Novus and were limited partners in another company that was the Mint's controlling shareholder.<sup>66</sup> Based on this information, the court concluded that there was sufficient evidence to defeat a summary judgment motion filed by the co-CEOs on the issue of vicarious copyright infringement. Given their role as corporate officers, the court inferred that they had supervisory authority

---

61. *Napster*, 239 F.3d at 1023.

62. 115 F. Supp. 2d 1089 (D. Minn. 2000).

63. *Id.* at 1091.

64. *Id.* at 1107.

65. *Id.* at 1106. It should be noted that the court—like many others—conflated the issue of vicarious and contributory liability, thereby failing to carefully parse out the distinctions between the two doctrines. As a result, some interpolation of its decision was needed for this analysis.

66. *Id.* at 1107.

over the employees of the direct copyright infringer.<sup>67</sup> Given their status as limited partners in the controlling shareholder, they had a direct financial interest in any revenues the direct copyright infringer received.<sup>68</sup>

However, the court found the same evidence insufficient to support a finding of vicarious trademark infringement because of “the more narrow standards applicable to trademark infringement claims.”<sup>69</sup> Vicarious trademark infringement requires greater proof of the defendant’s intent than vicarious copyright infringement, the court explained, and the government had produced no evidence demonstrating that the defendants knew or should have known about the manufacturer’s infringement.<sup>70</sup>

Thus, a financial interest and limited supervisory authority over the direct infringer is not enough to establish vicarious trademark infringement but it is enough to support a finding of vicarious copyright infringement. Although it appeared that the corporate officers’ financial interests were intertwined with the infringing manufacturer, the court apparently felt that the government had not demonstrated the requisite principal-agent relationship to warrant the imposition of vicarious trademark liability.<sup>71</sup>

### C. Comparing Contributory Liability in Trademark and Copyright Law

The divergence of trademark and copyright from a common source is even starker in the context of contributory liability. The basic contributory infringement doctrine—that both trademark and copyright law rhetorically share—finds liability where a defendant knows or should know of a third-party’s infringing activity and materially contributes to it. These common elements of knowledge and material contribution have taken on strikingly different meanings depending on whether trademark or copyright protection is at stake.

---

67. *Id.*

68. *Id.*

69. *Id.*

70. *Id.* at 1106-07. The court may have confused vicarious trademark infringement with contributory trademark infringement. Knowledge is not a requirement for vicarious trademark liability.

71. *See also* Banff Ltd. v. Limited, Inc., 869 F. Supp. 1103, 1111 (S.D.N.Y. 1994) (holding that because “the required showing of involvement with the [trademark] infringement would need to be the same or greater than the showing required by copyright law,” an analysis of the plaintiff’s vicarious trademark infringement claim was unnecessary once the court found in defendant’s favor regarding vicarious copyright infringement).

1. *Actual and Imputed Knowledge*

a) The Scope of Imputation in Trademark Law

As with vicarious liability, a defendant may be held liable as a contributory infringer even though the defendant has not taken any direct action to infringe on a trademark.<sup>72</sup> The seminal *Coca-Cola Co. v. Snow Crest Beverages* decision firmly establishes the parameters of contributory liability in trademark law.<sup>73</sup> In the case, both Snow Crest and Coca-Cola supplied cola-flavored soft drinks to bars. Coca-Cola sued Snow Crest for contributory infringement, contending that the court should hold Snow Crest indirectly liable for the infringing acts of the bars, who served drinks made with the defendant's "Polar Cola" when bar customers asked for "Coke."

In concluding that Snow Crest could not be held liable for the actions of the bar owners, the court tried to pinpoint the boundaries of contributory liability based on its scrutiny of both common law principles and the Lanham Act.<sup>74</sup> It cited the Restatement of Torts to explain that Snow Crest was under a duty to avoid intentionally inducing bars to market its Polar Cola product as "Coca Cola."<sup>75</sup> Snow Crest also had a duty to avoid knowingly aiding bars that purchased its products from engaging in infringing conduct.<sup>76</sup> Most importantly, Snow Crest was under an obligation to take precautionary measures if it knew or could reasonably be expected to know that the bars were using its product as a substitute when customers ordered Coca-Cola.<sup>77</sup> Thus, according to the *Snow Crest* court, knowledge—actual or constructive—of the direct infringer's infringing behavior is a required ingredient for contributory trademark infringement.

But the *Snow Crest* court also stressed that these obligations mark the outer limits of a manufacturer's duties with regard to policing the infringing acts of its customers: "There is no broader legal principle that always makes the defendant his brother's or his customer's keeper."<sup>78</sup> Instead, liability turns on whether "a reasonable person in the defendant's position" would realize that she had created a situation likely to result in infringe-

---

72. *Power Test Petroleum Distribs. v. Manhattan & Queens Fuel Corp.*, 556 F. Supp. 392, 395 (E.D.N.Y. 1982).

73. 64 F. Supp. 980 (D. Mass. 1946). *Snow Crest* was relied on by the Supreme Court in the *Ives* decision. See *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 854 (1982).

74. *Snow Crest*, 64 F. Supp. at 985.

75. *Id.* at 989 (citing RESTATEMENT OF TORTS § 713 (1938)).

76. *Id.* (citing RESTATEMENT OF TORTS § 738 (1938)).

77. *Id.*

78. *Id.*

ment or was transacting with a customer that she should know would be particularly likely to use her product wrongfully.<sup>79</sup> After all, as Judge Wyzanski observed in the case, “any man of common sense knows that in any line of business . . . there are some unscrupulous persons, who, when it is to their financial advantage to do so, will palm off on customers a different product from that ordered by the customer.”<sup>80</sup> In other words, *Snow Crest* sets out a reasonable person standard for imputing the knowledge necessary for contributory infringement, only permitting liability when a defendant knew or reasonably should have known that her actions would result in infringement by another.

Since 1946, this firm limitation on knowledge imputation has dominated the law of contributory liability in trademark infringement cases, even receiving the Supreme Court’s blessing. As Justice White wrote in his concurring opinion in *Inwood Laboratories v. Ives Laboratories*, “The mere fact that a generic drug company can anticipate that some illegal substitution will occur to some unspecified extent, and by some unknown pharmacists, should not by itself be a predicate for contributory liability.”<sup>81</sup> Thus, even a guarantee of trademark infringement somewhere down the stream of commerce is not enough to support a finding of contributory liability. Today, courts continue to follow the limitations of the *Snow Crest* decision, imposing no affirmative duty to investigate or take precautions against trademark infringement by a third party, barring some specialized knowledge of the infringement at issue.<sup>82</sup>

The actual or constructive knowledge standard in *Snow Crest* is common in other branches of tort law.<sup>83</sup> In the typical intentional inducement case, knowledge of the direct infringement is readily apparent because there is evidence that the defendant specifically requested that the direct

---

79. *Id.* (citing RESTATEMENT OF TORTS § 302 (1938)).

80. *Id.* at 988-89.

81. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 861 (1982) (White, J., concurring); see also 2 MCCARTHY, TRADEMARKS AND UNFAIR COMPETITION § 25:2, at 242 (1973) (“[T]he supplier’s duty does not go so far as to require him to refuse to sell to dealers who merely *might* pass off its goods.”).

82. See Dogan & Lemley, *supra* note 23, at 830; Cross, *supra* note 24, at 653. It is important to emphasize that contributory liability for trademark infringement is not a negligence standard. See 3 GILSON, *supra* note 25, § 11.02(2)(h)(i)(c). A mere failure to take reasonable precautions is not enough to make a defendant liable. *Hard Rock Café Licensing Corp. v. Concession Servs., Inc.*, 955 F.2d 1143, 1149 (7th Cir. 1992).

83. See, e.g., *Nitsche v. CEO of Osago Valley Elec. Co-op*, 446 F.3d 841, 844-45 (8th Cir. 2006) (hostile work environment sexual harassment); *In re Nokia Oyaj Sec. Litig.*, 423 F. Supp. 2d 364, 403 (S.D.N.Y. 2006) (securities fraud); *Valentine v. LaBow*, 897 A.2d 624, 633 (Conn. App. Ct. 2006) (intentional infliction of emotional distress).

infringer violate another's trademark.<sup>84</sup> For example, the finding of contributory liability was relatively routine in a case where a sales representative told others that he had received a "royal screwing" and that he was going to "even the score" with a manufacturer that used to supply him with the product.<sup>85</sup> The plaintiff in that case submitted proof that the sales representative contacted two other manufacturers and asked them to produce lamps nearly identical to those the plaintiff produced.<sup>86</sup>

Cases that involve the supply of a product without actual evidence of a specific request to infringe are more difficult. In determining whether the defendant had sufficient knowledge of infringement to be contributorily liable, the standard is that the defendant "understand what a reasonably prudent person would understand."<sup>87</sup> This is a "high burden" for a plaintiff,<sup>88</sup> and the most difficult element for a plaintiff to prove.<sup>89</sup> Even a demand letter from the plaintiff trademark owner to the defendant is not sufficient to create the amount of knowledge needed for a contributory infringement claim.<sup>90</sup> It is only when the reasonably prudent person would expect wrongdoing that contributory liability may attach. A reasonably prudent person would not assume infringement without real evidence of same. Mere awareness of a potential for infringement is not enough.<sup>91</sup>

On the other hand, under the *Snow Crest* standard, a defendant cannot purposely avoid evidence of infringement in order to immunize itself from contributory liability. If a defendant expects wrongdoing yet fails to investigate, such "willful blindness" will subject the defendant to trademark infringement liability.<sup>92</sup> For example, in the case involving allegations of a flea market vendor's direct infringement of the Hard Rock Café mark and the flea market owner's contributory infringement, the Seventh Circuit held that the owner could be contributorily liable even if he did not actu-

---

84. See, e.g., *William R. Warner & Co. v. Eli Lilly & Co.*, 265 U.S. 526, 530 (1924); *Sealy, Inc. v. Easy Living, Inc.*, 743 F.2d 1378, 1382 (9th Cir. 1984); *Transdermal Prods., Inc. v. Performance Contract Packaging, Inc.*, 943 F. Supp. 551, 553-54 (E.D. Pa. 1996).

85. *Bauer Lamp Co. v. Shaffer*, 941 F.2d 1165 (11th Cir. 1991).

86. *Id.* at 1169, 1171.

87. *Hard Rock*, 955 F.2d at 1149 (citing RESTATEMENT (SECOND) OF TORTS § 12(1) cmt. a (1965)).

88. *Gucci Am., Inc. v. Hall & Assocs.*, 135 F. Supp. 2d 409, 420 (S.D.N.Y. 2001).

89. Cross, *supra* note 24, at 653.

90. See *Gucci*, 135 F. Supp. 2d at 420; *Lockheed Martin Corp. v. Network Solutions, Inc.*, 985 F. Supp. 949, 964 (C.D. Cal. 1997).

91. *Monsanto Co. v. Campuzano*, 206 F. Supp. 2d 1271, 1278 (S.D. Fla. 2002) (holding that a distributor's awareness of a similar scheme involving someone different than the direct infringer is not enough to find distributor contributorily liable).

92. *Hard Rock*, 955 F.2d at 1149.

ally know that the vendor was selling fake Hard Rock t-shirts on his property.<sup>93</sup> The court would impute knowledge to the owner if the owner suspected, or had reason to suspect, wrongdoing and did nothing about it.<sup>94</sup> The *Hard Rock* court cautioned, though, that it was not converting the knowledge requirement for contributory trademark liability into a negligence standard. Like the *Snow Crest* court, it stressed that the flea market owner had “no affirmative duty to take precautions against the sale of counterfeits.”<sup>95</sup> And like the *Snow Crest* court, it borrowed from the Restatement of Torts and traditional common law tort doctrine to explain that, although the knowledge requirement for contributory trademark liability requires an owner “to understand what a reasonably prudent person would understand, it does not impose any duty to seek out and prevent violations.”<sup>96</sup>

b) Imputed Knowledge, Active Inducement, and the *Sony Safe Harbor* in Copyright Law

As with trademark infringement, a party “who, with knowledge of the infringing activity, induces, causes or materially contributes to the infringing conduct of another” will be liable for contributory copyright infringement.<sup>97</sup> Similarly, knowledge can be either actual or constructive<sup>98</sup>—the requisite knowledge exists if the defendant knew or had reason to know of the infringing activity.<sup>99</sup> As with trademark law, a defendant’s willful blindness of user infringement will satisfy the knowledge element in copyright.<sup>100</sup>

However, several additional factors have radically altered the knowledge requirement in the copyright context. First, the Supreme Court created the *Sony* safe harbor in its seminal *Sony Corporation v. Universal*

---

93. *Id.*

94. *Id.*

95. *Id.*

96. *Id.*

97. *Cable/Home Commc’n Corp. v. Network Prods., Inc.*, 902 F.2d 829, 845 (11th Cir. 1990) (citations omitted).

98. *In re Aimster Copyright Litig.*, 334 F.3d 643, 650 (7th Cir. 2003).

99. *Cable/Home*, 902 F.2d at 845. Although there is little guidance in the case law regarding the specificity of the knowledge required, a general understanding or belief that the infringement alleged is likely taking place usually suffices for a finding of contributory liability. Grossman, *supra* note 42, at 151; *see also* *UMG Recordings, Inc. v. Sinott*, 300 F. Supp. 2d 993, 998 (E.D. Cal. 2004) (stating that “actual knowledge of specific instances of infringement” is not required to satisfy the knowledge prong of contributory liability).

100. *Aimster*, 334 F.3d at 650.

*City Studios* decision.<sup>101</sup> In that case, the major motion picture studios filed suit against Sony for contributory and vicarious copyright infringement stemming from its development of the Betamax technology. Warning of the potential demise of Hollywood at the hands of video recording technology, the studios argued that the advent of the Betamax (and, ultimately, its more popular counterpart, the VHS) would dramatically reduce audiences for television programming. The studios argued that consumers would simply record programs and watch them at a later date.<sup>102</sup> This would devastate both the television and motion picture industries by decreasing the film and broadcast television audiences. The studios also contended that the recording features of the Betamax would annihilate the potential market for film rentals because consumers could create their own libraries of recorded movies from television.<sup>103</sup> The Supreme Court, however, disagreed.

The Court found that the existence of potential infringing uses for a technology should not render that technology illegal *per se*. Specifically, the *Sony* Court barred contributory liability based on imputed intent to cause infringement where a “staple article of commerce” used in infringement possessed “substantial noninfringing uses.”<sup>104</sup> The Court concluded that the VCR possessed significant noninfringing uses. Consequently, Sony could not be held liable for acts that its Betamax technology facilitated.<sup>105</sup> Under the *Sony* decision, when a product is capable of both substantial infringing and noninfringing use, without more, defendant’s mere knowledge of the product’s infringing capabilities is insufficient for a finding of contributory copyright infringement.<sup>106</sup> The *Sony* safe harbor therefore prevents courts from imputing knowledge of infringement to manufacturers of technologies having “commercially significant” or “substantial noninfringing uses.”<sup>107</sup>

Nonetheless, the *Sony* safe harbor is limited in two critical ways. First, it is riddled with ambiguity, making it difficult to rely on *ex ante*. This fact is especially problematic for developers of cutting-edge technologies with both infringing and noninfringing uses, as they risk millions of dollars in

---

101. *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984).

102. *Universal City Studios, Inc. v. Sony Corp.*, 480 F. Supp. 429, 466 (C.D. Cal. 1979).

103. *Id.* at 467.

104. *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 442 (1984).

105. *Id.* at 456. The Court also held that consumers are entitled under the fair use doctrine to engage in time-shifting, i.e., the recording of a televised program for personal and private viewing at a different time. *Id.* at 455.

106. *See id.* at 442.

107. *See id.*

potential secondary infringement liability. Moreover, it is unclear whether noninfringing uses must be actual or probable to qualify for the defense. Similarly unknown is the amount of time courts should grant technologies to develop the substantiality of their noninfringing uses. The interplay of the respective magnitudes of infringing and noninfringing uses also remains in doubt.<sup>108</sup>

Second, the Supreme Court's recent *Grokster* ruling explicitly and significantly limited the scope of the *Sony* safe harbor. In *Grokster*, movie and sound recording copyright holders brought suit against peer-to-peer software distributors for secondary copyright infringement.<sup>109</sup> The Court held that even though the software had substantial lawful uses, clear evidence that *Grokster* took steps to foster infringement obviated the *Sony* defense.<sup>110</sup>

In *Grokster*, the Ninth Circuit had construed *Sony* as immunizing from *all* contributory liability any technology capable of substantial or commercially viable noninfringing use, unless the distributor of that technology had actual knowledge of specific instances of infringement and failed to act upon that knowledge.<sup>111</sup> In a unanimous reversal, the Supreme Court disagreed with this broad interpretation of *Sony*. The Court clarified that while the *Sony* safe harbor prevents a court from imputing knowledge to a defendant distributing a product with substantial or commercially significant noninfringing uses, a finding that the defendant has "actively in-

---

108. On these points, *Grokster*'s two concurring opinions—the first written by Justice Ginsburg and joined by Chief Justice Rehnquist and Justice Kennedy, and the second written by Justice Breyer and joined by Justices Stevens and O'Connor—likely epitomize the locus of future litigation. To Justice Ginsburg, the sheer volume of infringing uses on peer-to-peer networks suggests no "reasonable prospect that substantial or commercially significant noninfringing uses were likely to develop over time." *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2764, 2786 (2005). Thus, Justice Ginsburg implicitly rejects the applicability of a *Sony* defense for *Grokster* and *StreamCast* if they are somehow able to avert active inducement liability on remand. By contrast, Justice Breyer views the *Sony* holding as precluding liability against *Grokster* or *StreamCast* on any theory beyond active inducement. Specifically, Justice Breyer contends that the *Sony* safe harbor applies to a technology unless it was clear that it would "be used almost exclusively to infringe copyrights." *Id.* at 2791. Thus, under Justice Breyer's interpretation, since peer-to-peer networks invariably possess noninfringing uses, *Grokster* and *StreamCast* would qualify for the *Sony* defense so long as they were not guilty of active inducement.

109. *Id.* at 2771.

110. *Id.* at 2779-80.

111. *MGM Studios, Inc. v. Grokster, Ltd. (Grokster I)*, 380 F.3d 1154, 1160 (9th Cir. 2004).

duced” infringement trumps the *Sony* defense.<sup>112</sup> Active inducement liability attaches if the defendant distributes that product “with the object of promoting its use to infringe copyright,” regardless of whether the product is capable of commercially significant noninfringing uses.<sup>113</sup>

*Grokster* therefore represents the holding that the affirmative *Sony* defense applies when the defendant has knowledge that its product can be used to infringe but the product is also capable of substantial lawful use.<sup>114</sup> Yet evidence that goes beyond the product’s design and demonstrates intent to infringe will trump the *Sony* defense and satisfy the knowledge element for contributory liability.<sup>115</sup> The Court emphasized that “direct evidence of unlawful purpose” was the key to overriding a *Sony* affirmative defense.<sup>116</sup> The Court explained that what it was looking for was evidence of “clear expression or other affirmative steps taken to foster infringement.”<sup>117</sup> The Court quoted Prosser and Keeton’s tort law treatise to justify placing this premium on direct evidence, indicating that higher penalties should apply to those with actual knowledge of illegal behavior.<sup>118</sup> Thus, *Grokster* unequivocally deems intent critical to the contributory liability calculus. Regardless of the availability of substantial noninfringing uses, active inducement can warrant a finding of secondary infringement.

---

112. *Grokster II*, 125 S. Ct. at 2779-80. The Supreme Court did not clarify whether active inducement constituted a third, independent, form of secondary liability or a subspecies of contributory liability. *Id.* at 2783 (Ginsburg, J., concurring) (stating that “active inducement” and contributory liability “overlap” but “capture different culpable behavior”); Amicus Curiae Brief of the American Intellectual Property Law Ass’n in Support of Vacatur and Remand at 6, *Grokster II*, 125 S. Ct. 2764 (No. 04-480) (“[P]atent law . . . substantiates the idea that ‘active inducement’ is a form of ‘contributory infringement.’”). For the purposes of clarity, we assume that active inducement is one means of imputing knowledge to meet the required elements for contributory liability.

113. *Grokster II*, 125 S. Ct. at 2780.

114. *Id.* at 2777-78.

115. *Id.* at 2779-80.

116. *Id.* at 2779. The Court stated:

The rule on inducement of infringement as developed in the early cases is no different today. Evidence of “active steps . . . taken to encourage direct infringement,” such as advertising an infringing use or instructing how to engage in an infringing use, show an affirmative intent that the product be used to infringe, and a showing that infringement was encouraged overcomes the law’s reluctance to find liability when a defendant merely sells a commercial product suitable for some lawful use.

*Id.* (citations omitted).

117. *Id.* at 2780.

118. *Id.* (“There is a definite tendency to impose greater responsibility upon a defendant whose conduct was intended to do harm, or was morally wrong.” (quoting W. KEETON, ET AL., PROSSER AND KEETON ON LAW OF TORTS 37 (5th ed. 1984))).

In articulating this active inducement standard, the Supreme Court dramatically increased the types of evidence considered relevant for a court's assessment of the knowledge requirement in copyright law. As the scope of the possible evidence of contribution increases, so increases the likelihood of imposing liability for contributory infringement. Specifically, the court can look toward any manifestation of intent to foster infringement in order to meet the threshold state of knowledge for contributory liability. Three factual considerations evidenced the *Grokster* defendants' clear intent to promote their products for infringing uses, apparently forcing the Court to conclude that "the unlawful objective is unmistakable."<sup>119</sup> These three factual considerations showcase types of evidence that, while relevant to the *Grokster* decision regarding liability for copyright infringement, have historically not been germane to assessing contributory trademark liability.

First, in their advertisements and solicitations, the creators of the peer-to-peer software at issue "voiced the objective that recipients use [their programs] to download copyrighted works, and each took active steps to encourage infringement."<sup>120</sup> Specifically, the defendants held themselves out as Napster substitutes, thereby trying to capture users of a known source of prior copyright infringement.

Second, albeit in language steeped in caution, the *Grokster* Court drew on the defendants' failure to develop filtering tools to bolster its finding of inducement.<sup>121</sup> *Grokster* establishes the critical importance of network architecture decisions to the secondary liability inquiry, even if such decisions are not outcome determinative.<sup>122</sup> The decision makes the failure to take affirmative precautions to prevent infringement a relevant factor for imputing knowledge of infringement for copyright contributory liability, but not for trademark contributory liability.

---

119. *Id.* at 2782.

120. *Id.* at 2772.

121. *Id.* at 2774, 2781.

122. The Court warned that "in the absence of other evidence of intent, a court would be unable to find contributory infringement liability merely based on a failure to take affirmative steps to prevent infringement." *Id.* at 2781 n.12. The Court therefore appeared to reject both the Ninth Circuit's position in *Grokster*—that an inability to control content on a network, even if that inability stems from a technology provider's willful desire to divest itself of such control, is entirely irrelevant to the liability calculus—and the Seventh Circuit's position in *Aimster*—that technology providers categorically cannot turn a blind eye towards infringing activities on their networks. See *MGM Studios, Inc. v. Grokster, Ltd. (Grokster I)*, 380 F.3d 1154, 1166 (9th Cir. 2004); *In re Aimster Copyright Litig. (Aimster)*, 334 F.3d 643, 649 (7th Cir. 2003).

Third, the Court highlighted the *Grokster* business model, importing the financial benefit calculus from *vicarious* infringement into its determination of *contributory* infringement. As the Court pointed out, defendants made money through advertising. As the number of users in their network increased, so would their advertising revenues.<sup>123</sup> Since close to ninety percent of volume on the network involved the unlawful exchange of copyrighted works, the Court concluded that the defendants' business model thrived on infringement.

The Court's emphasis on financial benefit as evidence of inducement and knowledge further transforms the contributory liability regime in copyright law. Specifically, it seizes upon the recent body of case law on financial benefit from the vicarious liability context in such cases as *Fonovisa* and *Napster* to broaden the scope of contributory liability through the inducement/knowledge factor. This trend further conflates vicarious and contributory liability which, as discussed earlier, courts have historically failed to parse out with deserved precision. Additionally, this trend introduces an imprecise financial metric to infer intent, thereby slackening contributory liability standards significantly. Traditional tort law and the law of contributory trademark infringement employ the standard of whether a reasonable person knew or should have known of the infringement. *Grokster*, on the other hand, allows a court assessing liability for contributory copyright infringement to impute knowledge based merely on financial motive.

All told, even when the *Sony* defense is taken into consideration, it does little to rectify the imbalance between trademark and copyright in imputing knowledge for contributory liability. The imprecise nature of the *Sony* defense together with the *Grokster* decision's emphasis on evidence of financial benefit and failure to take precautionary measures limit the utility of the *Sony* safe harbor for accused copyright infringers. As the law currently stands, the *Grokster* Court relied on inducement evidence that would not satisfy the standards for liability in a trademark infringement case. Neither a financial interest in the infringement nor a failure to take remedial measures would meet the "high burden" required to establish knowledge in a trademark case.<sup>124</sup> Instead, to warrant a finding of contributory trademark liability, courts only accept specific evidence of intentional inducement to infringe or proof that would lead a reasonably prudent person to conclude that infringement is taking place.

---

123. *Grokster II*, 125 S. Ct. at 2782.

124. *See, e.g.*, *Gucci Am., Inc. v. Hall & Assocs.*, 135 F. Supp. 2d 409, 419 (S.D.N.Y. 2001); *Coca-Cola Co. v. Snow Crest Beverages, Inc.*, 64 F. Supp. 980, 989 (D. Mass. 1946).

Particularly in the wake of *Grokster*, the more relaxed knowledge standard for contributory copyright infringement could implicate a wider range of defendants than contributory trademark infringement principles do.<sup>125</sup> Trademark doctrine imposes liability only when the defendant should have known that her actions would result in infringement by another. Contributory trademark defendants have no duty to investigate or adopt precautionary measures against third party infringement. In contrast, the Supreme Court's new "active inducement" standard for contributory copyright infringement awards liability based on a watered-down concept of knowledge of infringement. Under *Grokster*, the court may infer knowledge of the infringing conduct based on a failure to take steps to prevent infringement (for example, neglecting to develop tools to filter out infringing content) or through evidence of a financial benefit to be gained from the infringement. Thus, *Grokster* permits courts to infer knowledge of the infringement from evidence that would not satisfy trademark law's reasonably prudent person standard.

## 2. *Material Contribution: Relationships Suitable for Contributory Liability*

As with vicarious liability, courts are more willing to infer a relationship sufficient to trigger contributory liability when the plaintiff is a copyright, rather than trademark, holder. Contributory liability does not attach to every party who has knowledge of infringing activity. In addition to determining whether a "reasonably prudent person" would have perceived infringement, a court must also assess the nature of the relationship between the defendant and the direct infringer.<sup>126</sup> These requirements are interrelated. Whether or not it is reasonable for a defendant to perceive infringement depends on the defendant's interaction with the direct infringer. Contributory trademark liability requires direct control and monitoring of the means of infringement. The broader concept of contributory copyright liability has been stretched to include situations where the de-

---

125. Cf. Deborah J. Peckham, *The Internet Auction House and Secondary Liability—Will eBay Have to Answer to Grokster?*, 95 TRADEMARK REP. 977, 1004-05 (2005) (discussing the relevance of *Grokster*'s inducement standard to secondary trademark liability); see also *United States v. Wash. Mint, LLC*, 115 F. Supp. 2d 1089, 1107 (D. Minn. 2000) (referring to "the more narrow standards" applicable to trademark infringement claims to deny liability against corporate officers even though liability was found against officers for copyright infringement).

126. Dogan & Lemley, *supra* note 23, at 812 (explaining that contributory infringement requires "both an act of direct infringement . . . and a special, narrowly defined relationship between the defendant and that infringement").

defendant exercised no control over the direct infringer and merely helped produce an opportunity to infringe.

a) The Direct Control Requirement in Trademark Law

In trademark law, a court looking to assess contributory liability against a defendant that is not a manufacturer or distributor of the infringing product must sail into somewhat uncharted waters. “[I]t is not clear how the doctrine [set out in *Ives*] applies to people who do not actually manufacture or distribute the good that is ultimately palmed off as made by someone else.”<sup>127</sup> Relatively few cases have extended contributory trademark liability past manufacturers and distributors of products, as the courts have been hesitant to move past the relationship that was at issue in the *Ives* decision.<sup>128</sup> As one court explained, “[e]ach extension of contributory liability doctrine beyond defendants who manufacture or distribute a mislabeled product has required careful examination of the circumstances to determine whether knowledge of infringement should be imputed to the alleged contributory infringer.”<sup>129</sup> The courts encounter little difficulty in finding the knowledge required to support a finding of contributory infringement when the defendant has passed the product along the distributive chain. Imputing knowledge becomes trickier, though, when the defendant has not built or issued a misleading product.<sup>130</sup>

Despite this judicial reluctance, contributory trademark liability has broadened in recent years to cover more than just manufacturers and distributors.<sup>131</sup> For example, the Eleventh Circuit has opined that a franchisor could be held contributorily liable for its franchisee’s direct trademark in-

---

127. *Hard Rock Café Licensing Corp. v. Concession Servs., Inc.*, 955 F.2d 1143, 1148 (7th Cir. 1992); *cf.* *Stabilisierungsfonds Fur Wein v. Kaiser Stuhl Wine Distribs. Pty. Ltd.*, 647 F.2d 200, 207 (D.C. Cir. 1981) (“Courts have long held that in patent, trademark, literary property, and copyright infringement cases, any member of the distribution chain can be sued as an alleged joint tortfeasor.”).

128. *See, e.g., Acad. of Motion Picture Arts & Scis. v. Network Solutions, Inc.*, 989 F. Supp. 1276 (C.D. Cal. 1997); *see also* *Dogan & Lemley*, *supra* note 23, at 829 (“[C]ontributory liability for the provision of a service is extremely rare in trademark law . . .”).

129. *Lockheed Martin Corp. v. Network Solutions, Inc.*, 985 F. Supp. 949, 961 (C.D. Cal. 1997), *aff’d*, 194 F.3d 980, 981 (9th Cir. 1999).

130. *Power Test Petroleum Distribs., Inc. v. Manhattan & Queens Fuel Corp.*, 556 F. Supp. 392, 394 (E.D.N.Y. 1982).

131. *See, e.g., Procter & Gamble Co. v. Haugen*, 317 F.3d 1121, 1128 (10th Cir. 2003) (stating that action may extend to “licensors, franchisors, or similarly situated third parties”).

fringement.<sup>132</sup> Additionally, the Seventh Circuit held that a flea market owner could face contributory liability for the infringing actions of vendors on its property if it responded with “willful blindness” to the vendors’ infringement.<sup>133</sup>

Nevertheless, in contrast to the unprecedented expansion of contributory copyright liability in recent years, courts have resisted reconsideration of the standards for contributory trademark infringement, even in cases that involve new technologies and provide no easily applicable precedent in common law tort. For example, in *Lockheed Martin Corporation v. Network Solutions, Inc.*, a domain name registrar was sued for contributory infringement.<sup>134</sup> The plaintiff contended that the registrar committed contributory infringement by registering third-party domain names that contained the plaintiff’s mark.<sup>135</sup>

Lacking a clear analogy to prior trademark or common law, the *Lockheed Martin* court had to take a stand on the boundaries of contributory liability. It did so in a way that set a definite limit on the material contribution requirement. The court characterized the previous contributory infringement cases as relying on an assessment of the amount of control the defendant exercised.<sup>136</sup> The court held that if the defendant is not supplying a product as in *Ives*, then contributory liability is possible only if there is “direct control and monitoring of the instrumentality used by a third party to infringe the plaintiff’s mark.”<sup>137</sup> Because the domain name registrar engaged in rote translation and did not conduct any real oversight of its registrants, the court concluded that there was not sufficient “direct control and monitoring.” Consequently, there was no contributory in-

---

132. *Mini Maid Servs. Co. v. Maid Brigade Sys., Inc.*, 967 F.2d 1516, 1521 (11th Cir. 1992). The court cautioned, however, that a franchisor may not be held liable for a single franchisee’s infringement solely because the franchisor failed to exercise reasonable diligence to prevent the violation. *Id.*

133. *Hard Rock Café Licensing Corp. v. Concession Servs., Inc.*, 955 F.2d 1143, 1148-49 (7th Cir. 1992). The court analogized the flea market owner to a landlord, a frequent subject of common law secondary liability actions. A landlord is responsible “for the torts of those it permits on its premises ‘knowing or having reason to know that [they are] acting or will act tortiously.’” *Id.* at 1148-49 (quoting RESTATEMENT (SECOND) OF TORTS § 877(c) (1979)). According to the court, a flea market operator is like a landlord in that it controls the area where the infringement takes place. As a result, the flea market operator has a duty to prevent infringement of which it has constructive knowledge. *Id.* at 1149.

134. 194 F.3d 980, 981 (9th Cir. 1999).

135. *Id.* at 983-85.

136. *Id.* at 984.

137. *Id.*

fringement.<sup>138</sup> Thus, according to the *Lockheed Martin* court, a sufficient relationship will exist between the defendant and the direct infringer to assess liability only when the defendant exerts or can reasonably be expected to exert direct control over the means of infringement.

Other courts have widely adopted the standard articulated in *Lockheed Martin*.<sup>139</sup> Thus, the absence of direct control and monitoring of the means of infringement may preclude contributory trademark liability. When the defendant does not exercise direct control, contributory liability will not attach. For example, a shoe company that sponsored a basketball exhibition was not contributorily liable for an infringing t-shirt its sponsored promotion company distributed.<sup>140</sup> The court emphasized that, despite the shoe company's sponsorship of the exhibition and its endorsement deal with the alleged t-shirt disseminator and direct infringer, there was insufficient evidence to find that the company directly controlled or monitored the promotional materials related to the exhibition.<sup>141</sup>

#### b) The Attenuated Notion of Control in Copyright Law

As with trademark law, in addition to knowledge of the infringement, a contributory copyright infringer must act in a way that materially contributes to the infringement. To make a material contribution, the defendant must either (1) contribute machinery or goods that provide the means to infringe, or (2) engage in personal conduct that furthers the infringement.<sup>142</sup> For example, a radio station that allows its equipment to be used to broadcast advertisements for infringing records may be held liable for contributory copyright infringement.<sup>143</sup>

In determining whether the defendant made a material contribution to the infringement, courts ask whether the defendant had the ability to con-

138. *Id.* at 985.

139. *See, e.g.*, *Habeeba's Dance of the Arts, Ltd. v. Knoblauch*, 430 F. Supp. 2d 709, 714 (S.D. Ohio 2006); *SB Designs v. Reebok Int'l, Ltd.*, 338 F. Supp. 2d 904, 913-14 (N.D. Ill. 2004); *Size, Inc. v. Network Solutions, Inc.*, 255 F. Supp. 2d 568, 572-73 (E.D. Va. 2003); *Fare Deals, Ltd. v. World Choice Travel.com, Inc.*, 180 F. Supp. 2d 678, 688-90 (D. Md. 2001).

140. *SB Designs*, 338 F. Supp. 2d at 914.

141. *Id.* at 912.

142. 3 NIMMER ON COPYRIGHT, *supra* note 43, § 12.04(A)(2); *see also* *Demetriades v. Kaufmann*, 690 F. Supp. 289, 294 (S.D.N.Y. 1988) (stating that "substantial involvement" with the infringing activity is required (citing to RESTATEMENT (SECOND) OF TORTS § 876(b) (1977))).

143. *Screen Gems-Columbia Music, Inc. v. Mark-Fi Records, Inc.*, 256 F. Supp. 399 (S.D.N.Y. 1966) (denying summary judgment).

trol the use of the copyrighted work.<sup>144</sup> For some courts, the amount of control is the key issue in determining contributory liability.<sup>145</sup> But, formal control over the direct infringer is not necessary for contributory liability in copyright.<sup>146</sup> “The fact that the infringing activity is not done under the direction or supervision of the person furnishing facilities, nor for the person’s benefit . . . does not necessarily immunize him from liability as a contributory [copyright] infringer.”<sup>147</sup> Instead, merely providing the means for infringing activity may suffice under a theory of contributory liability even if not under a theory of direct liability.<sup>148</sup> For example, a commercial operator of sound recording or video duplication facilities may be held liable for the infringing acts of its customers even if the customers bring in the copyrighted materials that they illegally copy.<sup>149</sup> In fact, the mere provision of “the site and facilities for known infringing activity is sufficient to establish contributory liability.”<sup>150</sup> Thus, a swap meet landlord was held contributorily liable for the infringing actions of vendors on its property.<sup>151</sup>

The courts have stretched the definition of control even further, suggesting that any ability to regulate customer conduct constitutes the control necessary for a *material contribution*. For example, one court found that an operator of a computer bulletin board service that automatically distributed all bulletin board postings, infringing or not, to service subscribers could be held contributorily liable for a subscriber’s posting of

---

144. *See Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 437 (1984) (stating that a contributory infringer must be “in a position to control the use of copyrighted works by others”).

145. *See Tim Wu, When Code Isn’t Law*, 89 VA. L. REV. 679, 738 (2003) (stating that, in the *Napster* and *Grokster* cases, the Ninth Circuit “took the issue of control as the sine qua non of contributory liability”).

146. *Sony*, 464 U.S. at 487 (Blackmun, J., dissenting) (“I agree with the Gershwin court that contributory liability may be imposed even when the defendant has no formal control over the infringer.”); *Gershwin Publ’g Corp. v. Columbia Artists Mgmt., Inc.*, 443 F.2d 1159, 1163 (2d Cir. 1971) (stating that “although CAMI had no formal power to control either the local association or the artists for whom it served as agent,” relevant factors for the analysis included “that the local association depended upon CAMI for direction in matters such as this, that CAMI was in a position to police the infringing conduct of its artists, and that it derived substantial financial benefit from the actions of the primary infringers”).

147. 3 NIMMER ON COPYRIGHT, *supra* note 43, § 12.04(A)(3)(b).

148. *See, e.g., Fonovisa, Inc. v. Cherry Auction, Inc.*, 76 F.3d 259, 264 (9th Cir. 1996).

149. *See, e.g., RCA Records v. All-Fast Sys., Inc.*, 594 F. Supp. 335 (S.D.N.Y. 1984).

150. *Fonovisa*, 76 F.3d at 264.

151. *Id.*

infringing work.<sup>152</sup> The court explained that running a bulletin board service constitutes “substantial” participation that went beyond simply renting premises to an infringer because the bulletin board service “does not completely relinquish control over how its system is used, unlike a landlord.”<sup>153</sup> Thus for copyright, parties may be liable for contributory infringement even if they have no real ability to control the acts of the direct infringer.<sup>154</sup>

By contrast, under the trademark standard articulated in *Lockheed Martin*, contributory trademark liability requires direct control and monitoring of the means of infringement. When the defendant does not exercise direct control, contributory liability will not attach. Thus, while the bulletin board service and recording studios were held to have made a material contribution sufficient for contributory copyright infringement, such a passive role would be insufficient for contributory trademark infringement. Real oversight and control of the tools of infringement is necessary in the trademark realm.

In addition, though a somewhat nebulous concept, it appears that the easier something is to control, the more likely a court will recognize contributory trademark infringement. Thus, a court held that a common carrier of gasoline provided a material contribution to infringement by its “physical possession” and delivery of unbranded gasoline to a filling station that was passing off its gasoline as GETTY brand gasoline.<sup>155</sup> Courts expect suppliers to have some control over the products they manufacture and distribute; hence, the material contribution requirement may fade when the alleged contributory infringer is a manufacturer or distributor of the infringing product.<sup>156</sup> Courts also expect real property owners to be able to

---

152. *See* Religious Tech. Ctr. v. Netcom On-line Commc’n Servs., Inc., 907 F. Supp. 1361, 1375 (N.D. Cal. 1995).

153. *Id.*

154. As the court held, if the defendant, Netcom, had knowledge of the infringing activity, “failure to simply cancel [the direct infringer’s] infringing message and thereby stop an infringing copy from being distributed worldwide constitutes substantial participation in [the direct infringer’s] public distribution of the message.” *Id.* at 1374.

155. *Getty Petroleum Corp. v. Aris Getty, Inc.*, 55 F.3d 718, 720 (1st Cir. 1995) (noting that although defendant did not have title, “it had, and supplied, an essential factor—physical possession of the property to which the trademark was to be attached”).

156. *Cf. Lockheed Martin Corp. v. Network Solutions, Inc.*, 194 F.3d 980, 984 (9th Cir. 1999) (explaining that in cases not involving the distribution or manufacture of a product, proof of direct control and monitoring “permits the expansion of *Inwood Lab.*’s ‘supplies a product’ requirement for contributory infringement”); *H-D Mich., Inc. v. Biker’s Dream, Inc.*, 48 U.S.P.Q.2d 1108, 1116 (C.D. Cal. 1998) (holding manufacturer of infringing motorcycles liable for contributory infringement even though it did not sell motorcycles directly to the public).

control what happens on their land or in their buildings.<sup>157</sup> By contrast, courts have found owners of more abstract property, such as trademark licenses, to lack the requisite control under *Lockheed Martin*.<sup>158</sup> Thus, in deciding that there was no contributory liability for a travel agency that licensed its mark to an affiliate who infringed on another party's mark, the court emphasized that the travel agency "licensed no real estate," but "merely licensed its own mark to the alleged direct infringer."<sup>159</sup> In another case, a court granted summary judgment in favor of a domain name registrar accused of contributory infringement on the grounds that the constantly changing nature of the internet made it impossible for the registrar to directly control and monitor the means of infringement.<sup>160</sup>

Thus, the circumstances enabling contributory trademark liability are dramatically limited compared to copyright law. In cases that do not involve manufacturing or distribution, the defendant must directly control and supervise the direct infringer. Moreover, the case law suggests that sufficient control for contributory trademark infringement will only be found when the means of infringement is relatively simple and tangible, such as real property. When the means of infringement are more abstract, contributory liability is less likely. Contributory copyright does not require this direct control. Instead, it is sufficient to merely produce an opportunity to infringe.

### III. UNDERSTANDING THE DIVERGENT EVOLUTION OF SECONDARY TRADEMARK AND COPYRIGHT LIABILITY

As the previous Part and Table 1 below illustrate, an alleged secondary trademark infringer is more likely to escape liability than an alleged secondary copyright infringer. For both copyright and trademark law, vicarious liability requires control over the direct infringer and a financial benefit from infringement. But, while vicarious trademark liability uses tradi-

---

157. See *Lockheed Martin*, 194 F.3d at 985 (contrasting the case at hand where no contributory liability was found with other cases where "defendants licensed real estate, with the consequent direct control over the activity that the third-party alleged infringers engaged in on the premises").

158. See *Mini Maid Servs. Co. v. Maid Brigade Sys., Inc.*, 967 F.2d 1516, 1520 (11th Cir. 1992) (holding that a licensor of a mark does not ordinarily have a duty to prevent a licensee's misuse of another party's mark).

159. *Fare Deals, Ltd. v. World Choice Travel.com, Inc.*, 180 F. Supp. 2d 678, 689 (D. Md. 2001).

160. See *Lockheed Martin Corp. v. Network Solutions, Inc.*, 985 F. Supp. 949, 962 (C.D. Cal. 1997).

tional tort principles of agency to determine whether sufficient control exists, copyright infringement does not require a formal agency relationship. Thus, a defendant may be held vicariously liable for its independent contractors' and licensees' direct copyright infringement, but not for the same parties' trademark infringement. Moreover, the notion of financial benefit for copyright has been stretched to include the potential draw and hypothetical revenue to the vicarious infringer while trademark law continues to demand a direct financial stake in infringement revenue, not projections of future income.

**Table 1**  
**Comparison of Vicarious Trademark and Copyright Liability**

	<b>Trademark Law</b>	<b>Copyright Law</b>
<b>Control over Infringer (Actual or Apparent)</b>	<ul style="list-style-type: none"> <li>• Test: the direct infringer must act on behalf of the defendant or as defendant's alter ego for the control element to be met</li> <li>• Traditional tort principles of agency apply requiring actual or apparent agency</li> <li>• Ability to supervise (e.g., contractual relationships such as licensor/licensee or franchisor/franchisee) is insufficient</li> </ul>	<ul style="list-style-type: none"> <li>• Test: "right and ability to supervise"</li> <li>• No requirement of actual or apparent agency</li> <li>• Can be held liable for actions of an independent contractor</li> </ul>
<b>Financial Benefit from Infringement</b>	<ul style="list-style-type: none"> <li>• Direct financial stake in infringement revenue required</li> <li>• Hallmarks of direct financial benefit include using infringement for entertaining of customers/clients or profit-sharing regime with infringer</li> </ul>	<ul style="list-style-type: none"> <li>• Indirect financial benefit sufficient, including attraction of customers/users to site because of infringement</li> <li>• Hypothetical future revenue from infringement enough</li> </ul>

For both trademark and copyright law, contributory liability requires knowledge of the direct infringement and a material contribution to that infringement. For the knowledge element, as illustrated in Table 2 below, trademark law necessitates that the defendant knew or should have known that he was transacting with a customer who was likely to infringe; mere

awareness of the potential for infringement is insufficient. In contrast, especially post-*Grokster*, knowledge may be imputed to a contributory copyright infringement defendant if the defendant received indirect financial benefit from the infringement or failed to take affirmative precautions against third-party infringers. Meanwhile, proving a sufficient material contribution for contributory trademark infringement requires direct control and monitoring of the means of infringement. On the other hand, the notion of control has become so attenuated in copyright that the mere production of an opportunity to infringe or provision of the means for infringement, even in the absence of meaningful control, is sufficient for liability.

**Table 2**  
**Comparison of Contributory Trademark and Copyright Liability**

	<b>Trademark Law</b>	<b>Copyright Law</b>
<b>Contributory Liability</b>	<ul style="list-style-type: none"> <li>• Test: Would reasonable person in defendant's position realize that she had created a situation likely to result in infringement or was transacting with a customer that she should know would be particularly likely to engage in infringement?</li> <li>• No general affirmative duty to investigate or take precautions against trademark infringement by third parties, barring specialized knowledge of infringement.</li> <li>• Mere awareness of potential infringement not enough for liability.</li> </ul>	<ul style="list-style-type: none"> <li>• Test: Actual knowledge or imputed knowledge if:               <ul style="list-style-type: none"> <li>○ Manufacture or distribute product that is incapable of "commercially significant" or "substantial noninfringing uses;" or</li> <li>○ Actively induce infringement by the defendant manufacturing or distributing the product "with the object of promoting its use to infringe copyright, as shown by clear expression or other affirmative steps taken to foster infringement" as evidenced through advertisements, failure to take affirmative precautions against third-party infringements, and direct and indirect financial benefits from infringement.</li> </ul> </li> </ul>
<b>Material Contribution to Infringement</b>	<ul style="list-style-type: none"> <li>• Requires direct control and monitoring of means of infringement.</li> </ul>	<ul style="list-style-type: none"> <li>• Merely producing opportunity to infringe or providing the means for infringement may be enough for liability.</li> </ul>

The remainder of this Article explores potential reasons for this schism between trademark and copyright law and suggests a more principled and consistent way to evaluate both types of indirect infringement claims. Although Part II demonstrated that the secondary liability doctrines for copyright and trademark have branched out in different directions, these doctrines originate from the same root. Judicial opinions identify the basis of contributory copyright infringement as “the basic common law doctrine that one who knowingly participates in or furthers a tortious act is jointly and severally liable with the prime tortfeasor.”<sup>161</sup> This is the same doctrinal basis often identified as the foundation of contributory liability in trademark law.<sup>162</sup> Moreover, well-established tort law principles of respondeat superior supply the bases for the vicarious liability doctrines in both copyright and trademark law.<sup>163</sup> Yet as the previous Part demonstrated, in practice, the law of secondary liability differs depending on whether infringement of a trademark or a copyright is at issue. Below we analyze some potential reasons for this divergence. We argue that neither the separate lines of original legal authority, variances in the overall scope of protection of the two types of intellectual property, nor a greater concern with chilling the behavior of indirect trademark participants explains the difference. Instead, the contrast in the two secondary liability regimes stems from a less rational and, ultimately, unsatisfactory source: panic over the mass infringement of copyright on the internet.

## A. Differences in the Trademark and Copyright Property Bundles

### 1. Sources of Origin

One potential explanation for the difference between copyright and trademark secondary liability is the separate origin of their underlying rights. Copyright and trademark protection are borne from two distinct sources. In American law, copyrights (and patents) are a product of the Progress Clause of the Constitution, which specifically authorizes Congress to enact laws “To promote the Progress of Science and useful Arts, by securing for limited Times, to Authors and Inventors, the exclusive

---

161. *Demetriades v. Kaufmann*, 690 F. Supp. 289, 293 (S.D.N.Y. 1988) (quoting *Screen Gems-Columbia Music, Inc. v. Marki-Fi Records, Inc.*, 256 F. Supp. 399, 403 (S.D.N.Y. 1966)).

162. See *Hard Rock Café Licensing Corp. v. Concession Servs., Inc.*, 955 F.2d 1143, 1148-49 (7th Cir. 1992); *Transdermal Prods., Inc. v. Performance Contract Packaging, Inc.*, 943 F. Supp. 551, 553 (E.D. Pa. 1996).

163. *Demetriades*, 690 F. Supp. at 292 (describing vicarious and contributory liability as “well established precepts of tort liability”).

Right to their respective Writings and Discoveries.”<sup>164</sup> By contrast, federal trademark protection is a strictly statutory creation. There is no Constitutional provision providing for trademark protection. Instead, pursuant to its inherent power under the Commerce Clause,<sup>165</sup> Congress passed the Lanham Act, which provides federal rights and remedies for trademark holders.<sup>166</sup>

However, the constitutional origin of copyright law does not account for the doctrine’s broader secondary liability reach. Both the Constitution and the Lanham Act are silent as to the liability of non-infringers. The Lanham Act merely states that “[a]ny person” who uses a mark in a way that is likely to deceive consumers as to the association of that mark with its owner shall be liable.<sup>167</sup> The Lanham Act does not explicitly mention contributory infringement or vicarious infringement.<sup>168</sup> Similarly, the Constitution is silent as to liability for indirect copyright infringers, and current copyright law states only that “[a]nyone who violates any of the exclusive rights of the copyright owner” will be an infringer.<sup>169</sup>

Despite this silence as to third-party infringers, courts have freely imported secondary liability principles into both trademark and copyright law. Whether or not common law doctrines are applicable in litigation under a federal statute depends on whether those principles advance the goals of the statute.<sup>170</sup> Courts argue that they should import secondary liability principles into federal trademark law because the Lanham Act is derived “generally and purposefully from the common law tort of unfair competition.”<sup>171</sup> In his concurrence to the *Ives* decision, Justice White remarked that “the purpose of the Lanham Act was to codify and unify the common law of unfair competition and trademark protection.”<sup>172</sup> Accord-

---

164. U.S. CONST. art. I, § 8, cl. 8.

165. U.S. CONST. art. I, § 8, cl. 3.

166. Trademark Act of 1946 (Lanham Act), Pub. L. No. 79-489, 60 Stat. 427 (1946) (codified at 15 U.S.C. §§ 1051-1127 (2006)).

167. 15 U.S.C. § 1125(a)(1) (2006).

168. One scholar maintains that, given the Lanham Act’s silence, it would be an error for a court to imply contributory liability from the Lanham Act, although such liability is available under the courts’ common law powers. See John T. Cross, *Contributory Infringement and Related Theories of Secondary Liability for Trademark Infringement*, 80 IOWA L. REV. 101, 119 (1994).

169. See 17 U.S.C. § 501 (2006).

170. *Am. Soc’y of Mechanical Eng’rs v. Hydrolevel Corp.*, 456 U.S. 556, 570 (1982).

171. *AT&T v. Winback & Conserve Program*, 42 F.3d 1421, 1433 (3d Cir. 1994).

172. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 861 (White, J., concurring); see also *SmithKline Beckman Corp. v. Pennex Products Co.*, 103 F.R.D. 539, 540 (E.D. Pa. 1984) (holding third parties liable as joint tortfeasors for trademark infringement and

ing to the Third Circuit, common law vicarious liability is consonant with the goals of the Lanham Act because the doctrine merely allocates liability for conduct the statute already proscribed rather than expanding the scope of proscribed conduct.<sup>173</sup> In fact, courts have freely imported secondary liability to trademark law since the inception of federal protection.<sup>174</sup>

Similarly, courts have unreservedly read secondary tort liability principles into copyright law, despite the absence of any explicit authority in either the Constitution or in the Copyright Act.<sup>175</sup> As with trademark, courts have justified this importation of third-party liability on the grounds that “copyright is analogous to a species of tort” and vicarious and contributory liability in tort are “well-established” precepts.<sup>176</sup> In determining whether to extend liability to third parties, the Supreme Court has dictated that broader principles of desert and deterrence, not the presence of explicit statutory authorization, should guide jurists:

The Copyright Act does not expressly render anyone liable for infringement committed by another. . . . The absence of such express language in the copyright statute does not preclude the imposition of liability for copyright infringements on certain parties who have not themselves engaged in infringing activity. For vicarious liability is imposed in virtually all areas of the law, and the concept of contributory infringement is merely a species of the broader problem of identifying the circumstances in which it is just to hold one individual accountable for the actions of another.<sup>177</sup>

Thus, because courts generally accept secondary liability in numerous legal realms analogous to trademark and copyright law, it is appropriate to

---

stating that “since trademark infringement and unfair competition are tortious, the doctrine of joint tortfeasors is applicable”) (citations omitted); *Transdermal Prods., Inc. v. Performance Contract Packaging, Inc.*, 943 F. Supp. 551, 553 (E.D. Pa. 1996) (explaining that contributory infringement theory grew out of the common law underpinnings of trademark law).

173. *AT&T*, 42 F.3d at 1430-31.

174. *See William R. Warner & Co. v. Eli Lilly & Co.*, 265 U.S. 526, 530-31 (1924) (finding defendant guilty because “[o]ne who induces another to commit a fraud and furnishes the means of consummating it is equally guilty and liable for the injury”).

175. *See, e.g., Gross v. Van Dyk Gravure Co.*, 230 F. 412, 414 (2d Cir. 1916) (“Why all who unite in an infringement [of copyright] are not, under the statute liable for damages sustained by plaintiff, we are unable to see . . . [A]s all united in infringing, all are responsible for the damages resulting from infringement.”).

176. *Demetriades v. Kaufmann*, 690 F. Supp. 289, 292 (S.D.N.Y. 1988); *see also Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 435 (1984); *Screen Gems-Columbia Music, Inc. v. Mark-Fi Records, Inc.*, 256 F. Supp. 399, 403 (S.D.N.Y. 1966).

177. *Sony*, 464 U.S. at 434-35.

import secondary liability into trademark and copyright. Since this justification applies equally to both of these types of intellectual property, the different legal origins of copyright and trademark do not adequately explain the dramatic variances in secondary liability principles.

## 2. *Differences in Scopes of Protection*

Another possible explanation for the divergence between secondary liability in trademark and copyright law stems from the difference in the underlying rights these laws protect. On the few occasions when courts have rationalized the secondary liability divergence, they have appealed to the distinction between rights granted to copyright holders and rights granted to trademark holders. As the courts have frequently posited, the scope of the trademark privilege pales by comparison to the copyright monopoly. Trademark, therefore, warrants a more restrictive secondary liability regime. As a federal district court recently explained:

Because the property right protected by trademark law is narrower than that protected by copyright law, liability for contributory infringement of a trademark is narrower than liability for contributory infringement of a copyright. Unlike trademark law, copyright law gives owners a generalized right to prohibit all copying, provided that the owner's rights are valid and the material copied is original. Trademark law, on the other hand, tolerates a broad range of non-infringing uses of words that are identical or similar to trademarks.<sup>178</sup>

The courts' logic, though, is flawed. Admittedly, trademark rights differ in scope from those granted by copyright. Trademarks have traditionally only provided their owners with the ability to prevent uses that are likely to confuse consumers. Thus, until recently, any use of another's trademark was allowed under federal law so long as it did not result in public misperception. By contrast, copyrights seemingly provide their owners with a wholesale ability to prevent any copying or improper appropriation, regardless of public perception.<sup>179</sup>

However, even a cursory examination of the two regimes reveals that, in many ways, trademark law is more expansive than copyright. Trademarks, unlike copyrights, are potentially infinite in duration, lasting so long as their owners can and do use them to distinguish a particular good

---

178. *Lockheed Martin Corp. v. Network Solutions, Inc.*, 985 F. Supp. 949, 965 (C.D. Cal. 1997) (citations omitted).

179. The exclusive rights of copyright owners are, of course, subject to fair use and the first sale doctrine.

or service. Moreover, there has been a significant expansion in the rights granted to trademark owners in recent years—a trend makes the continuing limits on trademark secondary liability all the more puzzling. Trademarking no longer confers solely the right to prevent uses of a mark that result in a likelihood of confusion at the point of sale. As courts have grown increasingly concerned with protecting the goodwill and investment of mark holders, they have expanded trademark protection to cover confusion that occurs both before<sup>180</sup> and after<sup>181</sup> the point of sale. The Federal Trademark Anti-Dilution Act<sup>182</sup> now provides potential remedies for both the blurring and tarnishment of a famous mark, regardless of the potential for consumer confusion.<sup>183</sup>

Furthermore, trademark law is not subject to certain limitations found in copyright. Copyright plaintiffs must show illicit copying to prove infringement. Independent creation is an absolute defense to a copyright infringement suit, and the burden lies with the plaintiff to prove that an alleged infringing work was not independently created. In trademark law, by contrast, access and illicit copying are largely irrelevant to the issue of liability<sup>184</sup> and are generally reserved for determining the scope of damages.<sup>185</sup> Further, copyright and trademark protection derive from different sources. Copyright lies almost exclusively in the domain of federal law. Any vindication of rights equivalent to those guaranteed or denied under the Copyright Act is properly preempted.<sup>186</sup> By contrast, trademark is not exclusively federal, and states are permitted to apply their own independent trademark systems. State trademark protection frequently expands upon the rights provided under the Lanham Act.<sup>187</sup> Thus, while the scope of trademark rights differs from copyright in important respects, copyright

---

180. *E.g.*, *Brookfield Commc'ns, Inc. v. W. Coast Entm't Corp.*, 174 F.3d 1036, 1061 (9th Cir. 1999).

181. *E.g.*, *United States v. Torkington*, 812 F.2d 1347, 1352-53 (11th Cir. 1987).

182. 15 U.S.C. § 1125(c) (2006).

183. 15 U.S.C. § 1125(c)(1) (providing relief for use of a famous mark “that is likely to cause dilution by blurring or dilution by tarnishment of the famous mark, regardless of the presence or absence of actual or likely confusion”).

184. Intent is used to help determine whether consumers are likely to be confused by a defendant's use of a mark, but the defendant's intent in adopting the mark is only one of a multitude of factors used by the courts to assess likelihood of confusion. *See, e.g.*, *E. & J. Gallo Winery v. Gallo Cattle Co.*, 967 F.2d 1280, 1290 (9th Cir. 1992); *Polaroid Corp. v. Polarad Elecs. Corp.*, 287 F.2d 492, 497 (2d Cir. 1961).

185. The provisions of the Lanham Act providing for attorneys' fees and treble damages are illustrative in this regard. *See* 15 U.S.C. § 1115(a) (2006).

186. 17 U.S.C. § 301 (2006).

187. CRAIG JOYCE ET AL., *COPYRIGHT LAW* 9 (1998).

does not appear so broad in reach as to justify the application of a more expansive secondary infringement regime.

### **B. Concerns with Chilling Behavior of Indirect Participants**

Since neither the scope nor origins of copyright and trademark law explain the divergence in their respective secondary liability regimes, one might search for a justification based on public policy. One possible basis for narrowing the definition of secondary trademark liability may lie in the potential danger inherent in cracking down on “indirect” infringers. A narrow definition of secondary liability protects intermediaries who interact with direct infringers. By holding the line on trademark secondary liability, courts’ rulings may reflect a concern that broad secondary trademark liability could chill legitimate behavior, particularly if expansion of liability would leave intermediaries uncertain as to what actions constitute infringement.<sup>188</sup> A judge may decide that preserving the status quo makes sense given the beneficial output of intermediaries, or potential “indirect” infringers, particularly on the internet. For example, Google’s ability to lower consumer search costs through targeted internet searches has been cited as one reason for shielding it from contributory trademark liability for its keyword search advertising program.<sup>189</sup>

Yet a desire to protect intermediaries does not explain why the law should hold an intermediate catalyst of trademark infringement to a lower standard of liability than a facilitator of copyright infringement. Both the Lanham Act and federal copyright law already account for the dangers of overzealous enforcement of contributory liability by providing safe harbors and limited remedies for defendants in industries at risk of excessive liability for legitimate activity.<sup>190</sup> The case law offers no policy justification for treading more lightly on indirect trademark infringers than on indirect copyright infringers. Indeed, there are a multitude of public policy grounds that courts could have drawn upon (though they did not) to rationalize the narrowing of secondary copyright liability. In the cases involving peer-to-peer technologies, for example, judges could easily have pointed to the social benefits of such technologies as a basis for shielding facilitators from civil liability. After all, peer-to-peer platforms enable internet users to transfer information more efficiently, thereby promoting resource conservation. Peer-to-peer technologies also facilitate the dis-

---

188. Dogan & Lemley, *supra* note 23, at 832.

189. *Id.* at 831-37.

190. The Lanham Act’s exception for trademark uses “otherwise than as a mark,” 15 U.S.C. § 1115(b)(4) (2006), and the safe harbor for liability of internet service providers in the Digital Millennium Copyright Act, 17 U.S.C. § 512 (2006), are examples.

semination of works in the public domain, advancing critical First Amendment and educational interests.<sup>191</sup> Yet the peer-to-peer cases significantly expanded the breadth of secondary liability in copyright. And while manufacturers and distributors of branded products are surely integral to the functioning of the American economy, so are the publishers, software developers, and technologists who face potential indirect liability when third parties use their services and products to violate copyright law.

### C. Copyright Panic

All told, there appears to be no unifying or rational theoretical basis to explain the divergent courses of the secondary copyright and trademark liability regimes. In fact, further examination suggests a somewhat less salutary and deliberate mechanism at work: copyright panic.

In recent years, the ease of digital reproduction and distribution and the availability of broadband internet access have enabled mass infringement of copyrighted works on an unprecedented scale. The fear of infringement has induced widespread copyright panic within the content-creation industries—a fear widely broadcast throughout the mainstream media. The panic has led to the passage of such ill-conceived legislation as the Digital Millennium Copyright Act and has led to calls for heightened legal protection for content creators.

In part, this wave of copyright protectionism has appealed to the notion of romantic authorship. Most prominently articulated by Peter Jaszi<sup>192</sup> and James Boyle,<sup>193</sup> this theory postulates that the notion of romantic authorship has tacitly served as a central driving force behind the expansion of the modern copyright regime. Specifically, copyright laws gain legitimacy by protecting the vision of authors as mythic, solitary geniuses whose individual efforts result in original works created *ex nihilo*. The sympathetic figure of the romantic author has enabled legislators and courts to rationalize copyright protectionism by elevating the mental labors of the author to “a privileged category of human enterprise.”<sup>194</sup>

---

191. See, e.g., Rebecca Tushnet, *Copy This Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 568–81 (2004) (arguing that the pure reproduction of creative works can advance First Amendment interests in self-expression, persuasion and affirmation).

192. Peter Jaszi, *Toward a Theory of Copyright: The Metamorphoses of “Authorship”*, 1991 DUKE L.J. 455, 455-63 (1991).

193. JAMES BOYLE, SHAMANS, SOFTWARE AND SPLEENS: LAW AND THE CONSTITUTION OF THE INFORMATION SOCIETY 51-60 (1996); James Boyle, *The Search for an Author: Shakespeare and the Framers*, 37 AM. U. L. REV. 625, 629 (1988).

194. Jaszi, *supra* note 192, at 455.

The individual consumer generally associates a copyright with an identifiable figure: a movie with the director, music with the singer or band members, and books with the author. Even when these figures—directors, authors, musicians—assign their copyrights to a recording label, a publisher, or a motion picture distributor, they still benefit from the exclusive rights guaranteed by Section 106 of the Copyright Act. As a result, in testimony before Congress and in advertisements pleading with consumers not to engage in piracy, the face of copyright—the sympathetic artist or creator—breeds the perception among consumers that copyright infringement is a personal, violative act.

By contrast, trademark development has traditionally been viewed as a strictly economic enterprise lacking in creativity or aesthetic value and, therefore, undeserving of the special protections reserved for authors and artists.<sup>195</sup> In an early American trademark case, the Supreme Court commented that a trademark, unlike a patent or copyright, does not “depend upon novelty, invention, discovery, or any work of the brain.”<sup>196</sup> More recently, a court observed that “a man of ordinary intelligence could easily devise a score of valid trade-marks in a short period of time.”<sup>197</sup> Moreover, trademark holders present a rather corporate visage unlikely to elicit public sympathy. After all, trademarks indicate the source or origin of products or services, items that typically enter the stream of commerce via corporate structures, not individual artists.

However, the appeal to romantic authorship cannot completely explain the recent revolution in copyright secondary liability.<sup>198</sup> The technological

---

195. See Mark Bartholomew, *Making a Mark in the Internet Economy: A Trademark Analysis of Search Engine Advertising*, 58 OKLA. L. REV. 179, 202-03 (2005).

196. *In re Trade-Mark Cases*, 100 U.S. 82, 94 (1879).

197. *Ambrosia Chocolate Co. v. Ambrosia Cake Bakery, Inc.*, 165 F.2d 693, 697 (4th Cir. 1947).

198. There are limitations to the explanatory power of the romantic authorship theory that lead us to conclude that the notion of romantic authorship cannot, by itself, explain the divergence between secondary trademark and copyright law. See Mark Lemley, *Romantic Authorship and the Rhetoric of Property*, 75 TEX. L. R. 873, 879 (1997); Pamela Samuelson, *The Quest for Enabling Metaphors for Law and Lawyering in the Information Age*, 94 MICH. L. REV. 2029, 2039 (1996). First, a high percentage of valuable copyrighted works are in the hands of corporations (often as works made for hire), rather than authors or individuals. See *Cnty. for Creative Non-Violence v. Reid*, 490 U.S. 730, 737 n.4 (1989); Dogan & Lemley, *supra* note 23, at 883. As a result, it is difficult to implicate romantic authorship concerns in the divergent evolution of trademark and copyright secondary liability, especially when the cases most noted for expanding the scope of secondary liability have consistently involved plaintiffs who were corporate rights holders, not individual creators. See, e.g., *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2764, 2769 (2005); *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1010-11 (9th

changes precipitated by the internet are equally responsible for the courts' embrace of novel indirect liability theories. The late 1990s saw increasing levels of copyright infringement spawned by a surge in internet use, the development of file compression technology such as the mp3 music format, and the creation and dissemination of peer-to-peer file-sharing technology. Beset by declining album sales, the music industry saw the peer-to-peer revolution as a direct threat to its continued survival and quickly engaged in a no-holds-barred litigation against the developers of peer-to-peer networks. The remarkable infringing potential of new technologies helped convince the courts to expand secondary copyright liability. Meanwhile, trademark owners, who seemingly lacked a highly-publicized technological threat of their own, remained stuck with traditional secondary liability rules. Thus, mass awareness over the digital revolution and its threat to copyright holders, made all the more sympathetic by the appeal to romantic authorship, has undoubtedly contributed to the expansion of the secondary liability regime in copyright law and the resulting gap between copyright and trademark.

1. *Early Judicial Responses to the Digital Era: Copyright vs. Trademark*

Consider the sharp contrast in judicial responses to the advent of secondary liability issues on the internet in trademark versus copyright cases. The first significant internet challenge to the secondary trademark liability regime arose in the context of cybersquatting. Individuals rushed to purchase domain names containing the trademarks of large multinational corporations, hoping to sell the domains to the corporations at a premium. When asking prices surpassed the cost of litigation, corporations began to

---

Cir. 2001); *Fonovisa, Inc. v. Cherry Auction, Inc.*, 76 F.3d 259, 261 (9th Cir. 1996); *Columbia Pictures Indus., Inc. v. Aveco, Inc.*, 800 F.2d 59, 60 (3d Cir. 1986); *RCA Records v. All-Fast Sys., Inc.*, 594 F. Supp. 335, 336 (S.D.N.Y. 1984). Second, while the notion of romantic authorship has existed for decades, the explosion in secondary copyright liability cases, and the resulting gap between copyright and trademark law, has taken place over the last few years. If the romantic authorship myth legitimizes an expansive secondary liability regime in copyright law, but not in trademark law, it is difficult to understand why it has not done so all along. Indeed, an exegesis of the secondary copyright liability jurisprudence fails to reveal any betrayal, either explicit or implicit, of such authorial romanticism. We mention romantic authorship here only as a contributing factor to the divergence between copyright and trademark. While by itself, the romantic view of copyright creation was insufficient to reset the boundaries of indirect liability, we submit that romanticism for authors helped create the foundation necessary for the recent panic over digital copyright infringement, which, in turn, has led to a dramatic expansion in secondary copyright liability and the growing divide between indirect liability for copyright and trademark.

sue these domain-holders for trademark infringement. However, when the judgment-proof status of many of the cybersquatters became clear, companies turned their attention to the domain name registrars. The companies' attempts failed as courts squarely rejected theories holding domain name registrars liable for secondary trademark infringement. In the seminal case on domain-name liability, *Lockheed Martin Corporation v. Network Solutions, Inc.*,<sup>199</sup> the Ninth Circuit affirmed a lower court's grant of summary judgment for Network Solutions ("NSI"). The courts determined that NSI could not be held contributorily liable for trademark infringement by allowing a cybersquatter to obtain multiple top-level domain names containing derivations of Lockheed's trademark "SKUNK WORKS." The Ninth Circuit explained that because domain name registration only involves "rote translation" of a registrant's IP address into a domain name, NSI lacked the direct control necessary for contributory trademark infringement.<sup>200</sup>

NSI's registration system was far from completely automated, though. In fact, NSI admitted that it intervened in ten percent of domain-name applications, either to correct clerical errors or to reject applications containing certain pre-designated "prohibited" character strings, such as Olympic, Red Cross, NASA or certain obscenities.<sup>201</sup> Nevertheless, even though NSI supplied the means of infringement and could police the registration of domain names rather than leave it to rote mechanisms, the court found no liability.<sup>202</sup>

The judicial response to the analogous issue of online copyright infringement was strikingly different. Faced with the prospect of mass infringement on the precursor to websites—bulletin board services—and judgment-proof direct defendants, copyright holders went after secondary defendants with significantly greater success than trademark owners. In the influential *Religious Technology Center v. Netcom* decision, a federal district court found that a bulletin board service that automatically distributed all user postings (with less oversight than NSI in its domain-name registering capacity) could potentially face contributory liability for the infringing actions of a posting user.<sup>203</sup> Earlier, in *Playboy v. Frena*,<sup>204</sup> a

---

199. 194 F.3d 980 (9th Cir. 1999).

200. *Id.* at 985.

201. *Id.* at 982.

202. *Id.* at 985. Previously, another federal court had denied a claim of contributory infringement against NSI for registering domain names containing derivations of the "AVERY DENNISON" trademark. See *Avery Dennison Corp. v. Supton*, 189 F.3d 868, 873 (9th Cir. 1999).

203. *Religious Tech. Ctr. v. Netcom On-line Commc'n Servs., Inc.*, 907 F. Supp. 1361, 1375 (N.D. Cal. 1995) (holding that the online service was not liable for direct or

federal district court had taken an expansive view of direct copyright liability, finding the operator of a bulletin board service liable for the distribution of unauthorized Playboy photos. The operator in *Playboy* had merely stored the photos on his servers, and users of his service had copied them.<sup>205</sup>

Thus, while courts immunized automated third parties from secondary trademark liability, they allowed secondary (and even direct) liability against similarly situated third parties in the copyright context. In the context of trademarks, the lack of adequate remedies for cybersquatting led to the passage of ICANN's Uniform Domain Name Resolution Policy and congressional amendment of the Lanham Act with the Anti-Cybersquatting Consumer Protection Act to create an explicit federal civil action for cybersquatting.<sup>206</sup> By sharp contrast, the courts' willingness to expand the reach of contributory and vicarious liability in copyright law has led to congressional involvement to *limit* secondary liability. Witness, for example, the passage of the safe harbor provisions of the Digital Millennium Copyright Act, which Congress promulgated to overrule *Frena* and to shield internet service providers from liability for the activities of their users.<sup>207</sup> In short, the courts have reacted cautiously to digital trademark issues, holding to the traditional bounds of vicarious and contributory liability. By contrast, courts have responded more actively to digital copyright issues by diluting existing requirements of direct financial benefit, control, knowledge and material contribution or even by establishing entirely new theories of infringement.

2. *"The Unlawful Objective Was Unmistakable": Peer-to-Peer File Sharing, Grokster, and the Fundamental Transformation of the Secondary Copyright Regime*

The judicial response to the peer-to-peer file-sharing revolution, as epitomized by the *Napster* decision and the Supreme Court's unanimous ruling in *Grokster*, provides another vivid illustration of copyright panic and its attendant consequences on the reshaping of secondary liability doctrine. An examination of the legal and technological background of these cases helps to illuminate the courts' jurisprudence.

---

vicarious infringement, but that a triable issue existed as to whether it faced liability for contributory infringement).

204. *Playboy Enters. v. Frena, Inc.*, 839 F. Supp. 1552 (M.D. Fla. 1993).

205. *Id.* at 1554.

206. 15 U.S.C. § 1125(d) (Supp. 2000); S. REP. NO. 106-140, at 11-18 (1999).

207. 17 U.S.C. § 512 (2006); H.R. REP. NO. 105-551, pt. 1, at 24 (1998).

The entertainment industry won its high profile battle against the leading first-generation file-sharing system, Napster, when a federal district court issued—and the Ninth Circuit affirmed—a preliminary injunction that effectively terminated Napster’s operations. Despite this victory, the industry’s success was ephemeral; users quickly turned to new and more sophisticated peer-to-peer technologies. Second-generation systems such as gnutella, Grokster, and KaZaa dramatically expanded infringement both in scope and in type. While Napster enabled the exchange of audio files, new systems and wider broadband access allowed users to swap commercial software, movies, and graphics. The networks also adopted superior file organization and retrieval techniques, enabling users to access copyrighted materials with greater agility.

Most significantly, second-generation networks structured their systems to evade liability. The Ninth Circuit affirmed the preliminary injunction against Napster on the grounds that, “the record supports the district court’s finding that Napster has *actual* knowledge that *specific* infringing material is available using its system, that it could block access to the system by suppliers of the infringing material, and that it failed to remove the material.”<sup>208</sup> Central to the court’s holding was its observation that “Napster has both the ability to use its search function to identify infringing musical recordings and the right to bar participation of users who engage in the transmission of infringing files.”<sup>209</sup> But unlike Napster, second-generation networks utilized a decentralized architecture. Napster housed a centralized index of available files on servers that it owned and operated. As a result, it was able to filter the types of files traded on its network. By contrast, the indices of second-generation networks were maintained on servers that were not owned or operated by the network provider. Thus, second-generation systems could shield themselves from liability by precluding their ability to control or monitor infringing activities on their networks.<sup>210</sup>

---

208. *A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004, 1022 (9th Cir. 2002) (emphasis in original).

209. *Id.* at 1027.

210. The networks also sought refuge in international legal arbitrage. Shell corporations now operate many second-generation networks. These entities can easily relocate their systems and operations to jurisdictions with more favorable laws. The story of KaZaa, the world’s most popular peer-to-peer software, epitomizes the viability of such legal arbitrage. Facing an adverse judgment in the Netherlands, the Dutch owners of KaZaa sold their software and service to the nebulous Sharman Networks Ltd. Sharman Networks is a notoriously secret corporation officially incorporated in the South Pacific tax haven of Vanuatu. Vanuatu recognizes no copyright laws. Thus, the enforceability of judgments against KaZaa is very much in doubt. The transnational characteristics of cy-

Confronted with the new peer-to-peer technology and reports of unprecedented infringement, the *Grokster* Court reconstructed secondary liability doctrine to impose liability. Starting from the premise that something terrible—mass copyright infringement—was occurring, the Supreme Court fashioned custom-made relief for the plaintiffs against the developers of peer-to-peer networks:

The argument for imposing indirect liability in this case is, however, a powerful one, given the number of infringing downloads that occur every day using StreamCast's and Grokster's software. When a widely shared service or product is used to commit infringement, it may be impossible to enforce rights in the protected work effectively against all direct infringers, the only practical alternative being to go against the distributor of the copying device for secondary liability on a theory of contributory or vicarious infringement.<sup>211</sup>

The Court therefore viewed mass infringement as a robust basis for inflicting secondary liability on technology developers. A careful analysis reveals the spurious nature of this logic.

First, the simple existence of mass infringement says little about whether secondary liability should attach. The continued vitality of the *Sony* safe harbor, reaffirmed by *Grokster*, makes this point plain. *Sony* shields technology developers from contributory liability if the technology is capable of substantial noninfringing uses and there is no evidence of active inducement by the developer. Nothing in the *Sony* decision suggests that the defense erodes in the face of mass infringement. After all, even the VCR was capable of promoting large-scale infringement by facilitating the long-term cataloging of movies and other copyrighted telecasts and by enabling video-to-video duplication of copyrighted works.

Second, the Court's argument about the impracticality of pursuing direct infringers is similarly unavailing. Challenges posed in recovery from direct infringers might concern copyright plaintiffs, especially the recording and movie industries. But it is not a proper basis for wholesale alteration of secondary liability law. The difficulty of pursuing direct in-

---

berspace combined with the nature of internet piracy have rendered legal action against peer-to-peer networks increasingly difficult. Ultimately, however, KaZaa settled with the RIAA in the summer of 2006. See Thomas Mennecke, *Kazaa Settles with Entertainment Industry*, SLYCK, July 27, 2003, <http://www.slyck.com/news.php?story=1250> (last visited Dec. 6, 2006) (noting also that "this may be a Pyrrhic victory" for the entertainment industry, as peer-to-peer users "have long left Kazaa and FastTrack behind").

211. *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2764, 2776 (2005).

fringers has never served as a doctrinal basis for the imposition of secondary liability. Such reasoning undermines the stability of legal guidelines, rendering them unreliable to technologists shaping the digital revolution, and erodes the principled bases for secondary liability, transforming copyright's vicarious and contributory liability regimes into amorphous traps to catch perceived bad actors.

Indeed, this dangerous vision of secondary liability is evident in the *Grokster* decision's most memorable line and resounding refrain: "The unlawful objective is unmistakable."<sup>212</sup> The Court predicated its expansion of secondary liability law on this manta. Fatally, however, the Court's manta presupposes the very question the Court was supposed to answer. Indeed, the unlawfulness of *Grokster*'s actions was anything but certain: the case went through several rounds of reversal and spurred a wave of amici briefs on all sides. But in a striking move, the Court molded a novel theory of secondary liability to remedy a perceived injustice. Copyright panic so seized the Court that it assumed an unlawful objective before it even made that determination on the merits.

The Court's reverse engineering is particularly salient in light of its general reluctance to fashion new forms of relief for a litany of plaintiffs suffering from injustices every bit as significant as the threat to copyright holders.<sup>213</sup> Even in the copyright and technology arena, the Supreme Court has consistently hesitated to carve out new theories of liability. One commentator recognized this hesitance in cases such as *White-Smith Music Publishing Co. v. Apollo Co.*,<sup>214</sup> *Fortnightly Corp. v. United Artists*,<sup>215</sup> and *Teleprompter Corp. v. Columbia Broadcasting System, Inc.*<sup>216</sup>

---

212. *Id.* at 2782.

213. *See, e.g.*, *United States v. Morrison*, 529 U.S. 598, 627 (2000) (striking down the Violence Against Women Act and its creation of a federal civil cause of action for victims of gender-motivated violence for exceeding congressional authority under the Commerce Clause); *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 576 (1992) ("Vindicating the public interest (including the public interest in Government observance of the Constitution and laws) is the function of Congress and the Chief Executive.").

214. 209 U.S. 1, 18 (1908) (holding that perforated player piano music rolls did not constitute unauthorized copies within the meaning of existing copyright law).

215. 392 U.S. 390, 400-02 (1968) (finding the unauthorized broadcasting of plaintiff's copyrighted works by defendant's community antenna television systems did not constitute a public performance of the copyrighted work proscribed under existing copyright law).

216. 415 U.S. 394, 409-10 (1974) (finding no liability for infringement for the unauthorized retransmission of copyrighted work on defendant's community antenna television systems).

[T]he [C]ourt confronted the same problem it had in *Grokster* and *Sony*—a new technological industry (the record and piano player, and various kinds of cable television) facing off against an incumbent industry. The Court in those cases said, in essence, we don't have a clue, found no copyright liability, and left things for Congress to fix. The Court in those cases made it clear that the Copyright Act, as written, had no answers to the problem presented, and that the Court did not trust itself to fashion one.<sup>217</sup>

Yet the *Grokster* Court did not hesitate to unanimously provide a new theory of liability to the plaintiffs, thereby providing a salient demonstration of the impact that the popular zeitgeist—here, copyright panic—can have on an area of jurisprudence.

The secondary liability regime's own malleability makes it particularly susceptible to such a panic-driven judicial response to the propagation of digital technology. The doctrines of indirect liability in both copyright and trademark are especially prone to mutation because of the dearth of statutory strictures delimiting them. Unlike patent law, which formulates its secondary liability regime explicitly in the Patent Act,<sup>218</sup> there is no explicit provision for secondary liability in either the Copyright Act or the Lanham Act. In fact, there is almost no legislative acknowledgment of such causes of action,<sup>219</sup> save a backdoor reference in the Digital Millennium Copyright Act<sup>220</sup> and an oblique reference to “authorizing” infringement in a House Report for the 1976 Copyright Act.<sup>221</sup>

---

217. Tim Wu, *The Copyright Paradox: Understanding Grokster*, 2005 SUP. CT. REV. 229, 254 (2005).

218. The Patent Act explicitly lays out causes of action for both inducing and contributory infringement. See 35 U.S.C. § 271(b), (d) (2006).

219. See Jay Dratler, Jr., *Common-Sense (Federal) Common Law Adrift in a Statutory Sea, or Why Grokster was a Unanimous Decision*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 413, 419 (2006).

220. See 17 U.S.C. § 1201(c)(2) (2006) (“Nothing in this section shall enlarge or diminish vicarious or contributory liability for copyright infringement in connection with any technology, product, service, device, component, or part thereof.”).

221. The House Report read:

The exclusive rights accorded to a copyright owner under Section 106 are “to do and to authorize” any of the activities specified in the five numbered clauses. Use of the phrase “to authorize” is intended to avoid any questions as to the liability of contributory infringers. For example, a person who lawfully acquires an authorized copy of a motion picture would be an infringer if he or she engages in the business of renting it to others for purposes of unauthorized public performance.

See H.R. REP. NO. 1476 (1976), *reprinted in* 1976 U.S.C.C.A.N. 5659, 5674.

As a result, common law has served as the sole vehicle for change in the secondary liability regime, enabling it in a rapid and fact-responsive fashion. As Jay Dratler has observed:

Common law decision making is inevitably *ad hoc*. It relies on general principles of justice and common sense. Its tools are analogy and distinction based on facts. By using these tools, courts mimic—on a much smaller scale and for a much smaller subset of factual contingencies—the comprehensive factual inquiries that legislatures are supposed to undertake before prescribing more comprehensive and general rules in statutes.<sup>222</sup>

One significant risk, therefore, is that cases with tough facts can easily result in flawed, and even dangerous, legal precedent. Certainly, one is left to wonder what happened to the *Sony* Court’s admonition—handed down from *White-Smith* to *Fortnightly* and *Teleprompter*—in declining to outlaw the Betamax: “Sound policy, as well as history, supports our consistent deference to Congress when technological innovations alter the market for copyrighted materials.”<sup>223</sup>

### 3. *The Dangers of Panic*

#### a) The Problematic Implications of *Grokster*

Although *Grokster* differed from most copyright cases in that it involved a new technology, the case sets legal precedent for any court weighing the evidence in a contributory infringement claim. The *Sony* safe harbor creates an additional evidentiary hurdle for a plaintiff challenging use of a new technology: a plaintiff must provide evidence of an intent to infringe when the technology at issue is capable of both infringing and noninfringing uses. However, with the guidance provided by *Grokster*, it is clear that evidence of financial motivation—broadly construed—or a failure to develop preventative measures, while not quite enough by itself to refute a *Sony* affirmative defense,<sup>224</sup> becomes a powerful weapon for any plaintiff trying to satisfy the knowledge requirement. Indeed, the *Grokster* decision’s sloppiness threatens to wreak havoc on technology developers in emerging fields.

In announcing the inducement theory of copyright liability, the Court pointed to several factual considerations that evidenced the *Grokster* defendants’ clear intent to promote their products for infringing uses. These

---

222. Dratler, *supra* note 219, at 420.

223. *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417, 431 (1984).

224. *MGM Studios, Inc. v. Grokster, Ltd. (Grokster II)*, 125 S. Ct. 2781, 2782 n.125 (2004).

factors form the basis for application of the inducement doctrine in future infringement suits, but each suffers from analytical uncertainties and could create liability for unwitting parties.<sup>225</sup> First, Grokster's unabashed vision of itself as a Napster-substitute provided the most salient evidence of inducement to the Court. Specifically, the Court focused on Grokster's targeted efforts to capture former Napster users in its advertisements and solicitations, thereby corroborating its illicit motives.<sup>226</sup> Remarkably, however, as Tim Wu points out, these advertisements were never actually released.<sup>227</sup> Counsel had wisely advised the companies against taking them public. Yet the Supreme Court still used the existence of an internal debate over such advertisements as a factor against the defendants—hazardous precedent for future developers of cutting-edge technologies with both infringing and noninfringing applications.

The expansive language in *Grokster* regarding evidence of intent implicates a wide range of previously unscrutinized activities that may now serve as predicates for the imposition of contributory liability. For example, *Grokster* calls into question the continued viability of a number of recent advertising campaigns including, as Rebecca Tushnet has pointed out, Apple Computer's "Rip. Burn. Mix." shibboleth.<sup>228</sup> Additionally, a company might market a product that it believes facilitates fair use of copyrighted works by its consumers. But if that use is ultimately deemed unfair, it is unclear whether that company's statements amount to inducement or whether a good faith belief in a product's fair use capacity shields the product's creator from contributory liability. Given the notoriously imprecise boundaries of copyright's fair use doctrine and the rapid pace of technological change,<sup>229</sup> these unresolved issues remain critical to technology developers.<sup>230</sup>

---

225. See *supra* Section II.C.1.

226. *Grokster II*, 125 S. Ct. at 2773.

227. Wu, *supra* note 217, at 243.

228. See Rebecca Tushnet, June 27, 2005, SCOTUSBLOG, <http://www.scotusblog.com/discussion/archives/grokster/>. Sophisticated technology companies, both in the mainstream and at the legal margins, will likely respond to *Grokster* by assiduously avoiding public and private statements that courts might read as inducing infringement.

229. See John Tehranian, *Et Tu, Fair Use? The Triumph of Natural-Law Copyright*, 38 U.C. DAVIS L. REV. 465, 496-504 (2005).

230. Ironically, *Grokster* expands contributory liability to address a specific problem—the threat of mass internet piracy—but it may well fail at addressing it. The realities of technological development make continued legal struggle potentially futile. The mainstream press touted *Grokster* as a significant victory for the entertainment industry, yet the technology at issue in the case is already antiquated. A third generation of peer-to-peer networks has already emerged, posing new challenges to the legal regime. For example, BitTorrent has supplanted KaZaa as the world's leading peer-to-peer network.

However, the most significant long-term impact of *Grokster* may not involve the inducement theory of infringement it announced, but the strong willingness it signaled to expand the secondary liability regime to meet the perceived needs of immediate justice in the copyright arena. As Tim Wu argues, *Grokster* provides us with the “the first test in copyright history that asks a court to look at a defendant’s business model and decide whether its motives are crooked.”<sup>231</sup> The implications of the Court’s transformation of secondary liability in *Grokster* are therefore dramatic. As new technologies emerge, it is likely that secondary copyright liability law will be subject to further distortions. In fact, *Grokster* does little to dissuade future courts from creating new forms of secondary liability. As Jay Dratler notes, future jurists might “suspect that infinitely fertile human imagination and the advance of technology may create other situations in which it would be just and proper to impose secondary liability.”<sup>232</sup>

b) The Potential Expansion of Secondary Trademark Liability:  
Panic or Sound Policy?

As described above, widespread panic over emergent digital technologies has spurred tremendous expansion of secondary copyright liability in recent years. Since no similar trademark panic has afflicted the popular imagination, we have witnessed only small alterations in the secondary trademark liability regime. Piracy, especially in the wake of the digital revolution, is most frequently construed as a threat to copyright, not trademark, holders. Ironically, however, the internet, globalization and new technologies enable as much mass trademark infringement as copyright infringement. Aided by technologies that allow easy replication,

---

Unlike many prior peer-to-peer iterations, BitTorrent was developed for the express purpose of facilitating noninfringing information transfers, and it operates on a non-commercial basis. Moreover, it possesses a far more decentralized architecture than prior networks. In light of these facts, the creators of BitTorrent appear insulated from liability under the standards promulgated by *Grokster*. Also, the *Grokster* ruling may be distinguished from future actions for contributory copyright infringement because the evidence of the defendant’s knowledge of the unlawful activity was unusually stark. The peer-to-peer software companies associated themselves with “the notorious file-sharing service, Napster” by targeting their advertisements to former Napster users. *Grokster II*, 125 S. Ct at 2772-73. Internal e-mails revealed that the companies sought to include more copyrighted songs on their networks than other file-sharing services. *Id.* at 2773. Not every contributory infringement defendant is likely to get caught so red-handed. *See* Grossman, *supra* note 42, at 201-02 (2005) (criticizing the Ninth Circuit for holding that actual knowledge can prevent a defense of substantial noninfringing use under *Sony*).

231. Wu, *supra* note 217, at 241.

232. Dratler, *supra* note 219, at 425.

counterfeiting has become a global issue.<sup>233</sup> The internet has spurred a host of complex trademark infringement issues, including the use of trademarks in metatags, search engines and advertising services. All told, the U.S. Chamber of Commerce estimates that counterfeiting or piracy costs the U.S. economy between \$200 to 250 billion each year.<sup>234</sup> Indeed, the problem has grown so serious that lawyers filing counterfeit and trademark infringement lawsuits are working with federal prosecutors, customs officials and local law enforcement to combat the problem.<sup>235</sup> When signing recent legislation authorizing criminal penalties for trafficking in counterfeit trademarks, the president described significant injuries resulting from trademark infringement: billions of dollars in domestic economic losses, health and safety risks from exposure to untested products, and the use of counterfeit sales to fund terrorist operations.<sup>236</sup> While we do not advocate trademark panic or a more expansive secondary trademark liability regime, the seriousness of real-world effects of trademark infringement in the digital age (and the absence of any morphing of trademark principles to address these issues) highlights just how striking and unusually cavalier the courts' responses to digital copyright infringement have been.

Despite the evidence of a pressing problem in the trademark arena, the copyright quandary has simply captured our collective attention. However, the secondary liability revolution may soon break in favor of trademark holders. The voices of complaint are rising, contending that the special nature of business transactions on the internet makes it easier to infringe on trademarks than ever before.<sup>237</sup> According to some, increased liability against indirect trademark infringers is justified in the internet context because new technology makes it easier for intermediaries to monitor the conduct of end users.<sup>238</sup> Of course, this is the same argument that the *Grokster* court seized to justify secondary copyright liability against the

---

233. See Amanda Bronstad, *Countering the Counterfeits*, NAT'L L.J., July 13, 2006.

234. *Id.*

235. *Id.*

236. See Press Release, The White House, Fact Sheet: President Bush Signs the Stop Counterfeiting in Manufactured Goods Act (Mar. 16, 2006), available at <http://www.whitehouse.gov/news/releases/2006/03/20060316-6.html> (last visited Dec. 3, 2006).

237. See, e.g., SALLY M. ABEL ET AL., *TRADEMARK LAW & THE INTERNET* 142 (Lisa E. Cristal & Neal S. Greenfield, eds., 2001); Adam S. Chinnock, Note, *Meta Tags: Another Whittle from the Stick of Trademark Protection?*, 32 U.C. DAVIS L. REV. 255, 276 (1998).

238. Ronald J. Mann & Seth R. Belzley, *The Promise of Internet Intermediary Liability*, 47 WM. & MARY L. REV. 239, 240 (2005).

peer-to-peer software distributor.<sup>239</sup> Yet trademark infringement may have just begun to capture the public's attention. Just this year, President Bush signed into law the Stop Counterfeiting in Manufactured Goods Act, providing criminal penalties against those who trade in counterfeit marks.<sup>240</sup>

On the other hand, Google, one of the most prominent online brands, is fighting expansion of secondary trademark liability. Google's AdWords and AdSense program permits advertisers to bid on keywords that will generate an advertising link when consumers search using that keyword or websites contain content using a keyword.<sup>241</sup> Mark holders have sued Google, contending that the unlicensed use of trademarked keywords for AdWords and AdSense constitutes infringement. So far, Google has been successful in preventing an adverse secondary liability verdict,<sup>242</sup> a result that would threaten to decimate the advertising program that represents its main source of revenue.<sup>243</sup> Part of Google's strategy rides on maintaining a positive public image—something the peer-to-peer software distributors failed to do—as epitomized by Google's good works<sup>244</sup> and appealing, happy-go-lucky mantra “Don't Be Evil.” Google's Library Project, a plan to digitize the works held by the United States' finest research universities, has been lauded by most of the public, receiving only limited criticism (and a matching lawsuit) from publishers concerned about devaluation of their copyrights. The Chronicle of Higher Education described the project as “providing researchers and students with an unprecedented tool for finding information.”<sup>245</sup> Google has also endeared itself to privacy rights advocates when it refused to turn over information in response to a Justice

---

239. See *supra* notes 224-230 and accompanying text.

240. H.R. 32, 109th Cong. (2005).

241. *Gov't Employees Ins. Co. v. Google, Inc.*, 330 F. Supp. 2d 700, 701-02 (E.D. Va. 2004).

242. Google recently succeeded in defeating lawsuits in the Northern District of New York and the Central District of California. See *Rescuecom Corp. v. Google, Inc.*, No. 5:04-CV-1055, 2006 WL 2811711 (N.D.N.Y. Sept. 28, 2006); *Perfect 10 v. Google, Inc.*, 416 F. Supp. 2d 828 (C.D. Cal. 2006).

243. See Nicholas Carlson, *Internet Ads Up 30%*, INTERNETNEWS.COM, Apr. 21, 2006, [http://www.internet\\_news.com/ec-news/article.php/3600946](http://www.internet_news.com/ec-news/article.php/3600946).

244. See, e.g., Ryan Kim, *Google Gives City free Wi-Fi: Mountain View Service Could Give S.F. Project a Push*, SF CHRONICLE, Aug. 16, 2006, at C1 (noting Google's “hospitable gesture” of offering a free high speed wireless network to its hometown of Mountain View, California).

245. Scott Carlson & Jeffrey R. Young, *Google Will Digitize and Search Millions of Books from 5 Leading Research Universities*, CHRON. HIGHER EDUC., Jan. 7, 2005, at A37.

Department subpoena for data on public search habits.<sup>246</sup> The strategy has worked.<sup>247</sup> “Google enjoys an unsullied image that sparkles cleaner than Coca Cola, Pepsi, Ford, Gap and AT&T combined.”<sup>248</sup> If Google succeeds in capturing public and judicial sympathies, it may be able to prevent the expansion of secondary trademark liability and avoid the fate of software developers who have felt the brunt of such cases as *Napster* and *Grokster*.

#### IV. TOWARDS A BETTER SECONDARY LIABILITY REGIME

Ideally, any reform of the trademark secondary liability regime should stem from a cost-benefit analysis of indirect liability as well as obeisance to traditional common law principles, not from a battle for public opinion. Yet a gap continues to grow in intellectual property jurisprudence because of courts’ irrational response to panic over the pace of technological change in the copyright realm.

Despite ostensibly common origins and similar policy justifications, the vicarious and contributory liability regimes in trademark vary markedly from those in copyright law. For vicarious liability, trademark law has generally mandated a principal-agent relationship and direct financial benefit to the defendant from the infringement. Despite its shared rhetoric, copyright law has increasingly come to require neither characteristic. Courts have found third-party copyright defendants vicariously liable in the absence of a principal-agent relationship, and they have vastly expanded the notion of financial benefit to include hypothetical sources of revenue, such as the monetization of internet traffic. For contributory liability, trademark law requires direct control and monitoring of the means of infringement. Meanwhile, courts have loosely defined the control element in contributory copyright law by imposing liability based merely on the defendant’s ability to regulate infringing conduct or provision of the facilities for infringing conduct. And while trademark law hews to traditional common law principles to infer knowledge of infringement, courts

---

246. Arshad Mohammed, *Google Refuses Demand for Search Information*, WASH. POST, Jan. 20, 2006, at A1.

247. Of course, Google’s public image is not entirely pristine. The company’s censorship of certain news stories on the Chinese version of Google has resulted in condemnation from free-speech and human-rights activists. See Loren Baker, *Google Responds to Google News China Controversy*, SEARCH ENGINE JOURNAL, Sept. 29, 2004, <http://www.searchenginejournal.com/index.php?p=910>.

248. The V7 Network Web Development Community, *Google Brand Strategy: Moral Superiority*, <http://www.v7n.com/google-branding-strategy.php> (last visited Aug. 17, 2006) (commenting further on Google users’ brand loyalty that “[o]ne can only imagine if given the choice between GoogleGuy and Jesus Christ today, we would most likely be mourning the loss of The Nazarene again”).

may impute knowledge to a copyright defendant based on a wide array of evidence irrelevant to the contributory trademark calculus.

No rational explanation exists in the case law for the copyright-trademark infringement dichotomy. Despite the shared common law origins of both secondary liability regimes, the Supreme Court has failed to provide an explanation for the divergent evolution of the two doctrines. With no intellectual grounding for the distinction, the courts have allowed irrational and unexplained fears to shape their jurisprudence. As epitomized by the *Grokster* decision, the courts have stretched secondary copyright liability almost beyond recognition because of panic over technological change and its impact on digital piracy.

The unexplained nature of the dichotomy leaves the law in an ambiguous state, and this uncertainty threatens to stifle a wide range of legitimate business activity. The absence of legible justification for the difference in the two secondary liability regimes leaves those who would conduct economic activity that indirectly touches on copyright and trademark with few markers to guide their activity. Owners of sites and technology implicating use of copyrighted works may feel the need to restrict access or limit functionality given the wide range of unanswered questions after the *Sony* and *Grokster* decisions. Indirect trademark participants cannot entirely rely on trademark law's narrow interpretation of secondary liability, especially in light of the vast unprincipled expansion of secondary liability doctrine in copyright law. Technologists shaping the digital revolution need rational and clear legal guidelines, not hazy doctrine untethered to historical or prudential argument.

The gap between the two secondary liability regimes is also problematic because it creates improper incentives. Copyright stakeholders have vigorously prosecuted their claims, using massive high-profile litigation to push the boundaries of secondary liability law. Meanwhile, secondary trademark liability has remained largely static, even while the same technology that has alarmed copyright holders and the courts is being used to infringe trademarks on a widespread basis. The courts have rewarded copyright holders for their aggressive litigation, justifying an expansion of secondary liability on the impracticality of pursuing direct infringers rather than on bedrock legal principles. The result is not merely a precarious and unprincipled definition of secondary copyright liability but also a dangerous precedent for future intellectual property suits. Cases such as *Grokster* can only encourage trademark stakeholders to take the same aggressive approach, flooding the courts with litigation and threatening to expand trademark doctrine beyond its current ambit without careful consideration and rationalization.

This Article lays the groundwork for further study on the imbalance between the trademark and copyright secondary liability. In particular, the divergent evolution of the two doctrines suggests a need to reevaluate secondary liability from a more deliberate, policy-oriented perspective. Specifically, jurists and legislators must consider how far modern copyright secondary liability principles have deviated from their common law origins and whether this divergence is reasoned or warranted. Similarly, we must ask ourselves whether it makes sense to bring secondary trademark liability doctrine into line with its analog in copyright law, or if other reforms are needed to address trademark concerns in the digital age. Secondary liability principles in other areas of the law provide another potential avenue for exploration. A comparative analysis could provide important lessons for intellectual property law. For example, criminal law examines in minute detail the mental states and level of intentionality that justify punishment for the acts of others.<sup>249</sup> And tort law doctrines outside of intellectual property, such as the learned intermediary doctrine, recalibrate secondary liability principles to more accurately target those who are best capable of preventing tortious conduct.<sup>250</sup> Perhaps insights are to be gleaned from these and other legal subject areas that have already profited from years of hard thinking about when liability is appropriate for indirect participants.

Finally, the divergence between secondary trademark and copyright principles reflects a key tension found throughout intellectual property law. Secondary liability principles reflect two different, and sometimes inconsistent, goals. First, secondary liability serves a fundamentally economic purpose by shifting risks from direct to indirect infringers. Second, secondary liability law serves a moral end by placing fault on a party deserving of punishment even though that party did not commit the underlying infringing act. The relative significance of these two goals can help determine the boundaries of secondary liability rules. For example, if moral desert dominates the rationale for liability imposition, courts should limit the ability to impute knowledge to a defendant in order to punish only truly bad actors for indirect infringement. On the other hand, if economic risk-shifting concerns animate secondary doctrine, courts may not choose to be so circumspect in their definitions of intent. Further analysis of the philosophical justifications for indirect liability would enable more rigorous evaluation of today's secondary liability rules and could help formulate a reasonable blueprint for future reform.

---

249. SANFORD H. KADISH & STEPHEN J. SCHULHOFER, CRIMINAL LAW AND ITS PROCESSES 204-05, 644-65 (6th ed. 1995).

250. *See* Walls v. Alpharma USPD, Inc., 887 So. 2d 881, 882-86 (Ala. 2004).



**DANGEROUS LIAISONS—SOFTWARE  
COMBINATIONS AS DERIVATIVE WORKS?  
DISTRIBUTION, INSTALLATION, AND EXECUTION  
OF LINKED PROGRAMS UNDER COPYRIGHT LAW,  
COMMERCIAL LICENSES, AND THE GPL**

*By Lothar Determann<sup>†</sup>*

**TABLE OF CONTENTS**

<b>I. INTRODUCTION</b> .....	1423
<b>II. BACKGROUND</b> .....	1425
<b>III. COPYRIGHT LAW ON DERIVATIVE WORKS, COMPILATIONS, AND OTHER COMBINATIONS</b> .....	1429
A. STATUTORY DEFINITIONS .....	1429
B. EXAMPLES OF COMBINATIONS .....	1431
C. OWNERSHIP OF DERIVATIVE WORKS VS. COMPILATIONS .....	1434
D. EXCLUSIVE RIGHTS TO DERIVATIVES WORKS VS. COMPILATIONS ..	1435
E. FIXATION REQUIREMENT .....	1436
F. SUMMARY .....	1437
<b>IV. SOFTWARE UNDER COPYRIGHT LAW</b> .....	1437
A. THE BALANCE BETWEEN ACCESS AND EXCLUSION RIGHTS .....	1438
B. CREATIVE EXPRESSION VS. FUNCTIONALITY .....	1438
C. SUMMARY .....	1442
<b>V. SOFTWARE COMBINATIONS UNDER COPYRIGHT LAW</b> .....	1443
A. SOFTWARE COMBINATIONS VIEWED FROM DIFFERENT TECHNICAL PERSPECTIVES .....	1444

---

© 2006 Lothar Determann

<sup>†</sup> Prof. Dr. Lothar Determann teaches courses on computer and internet law at the University of California, Berkeley School of Law (Boalt Hall), University of San Francisco School of Law, and Freie Universität Berlin ([www.lothar.determann.name](http://www.lothar.determann.name)), and practices law as a partner in the technology practice group of Baker & McKenzie LLP, San Francisco/Palo Alto office ([www.bakernet.com](http://www.bakernet.com)). The author is grateful for assistance from his students, in particular Tal Lavian, Principal Scientist at Nortel Labs (valuable comments from computer science perspective), Steven B. Toeniskoetter, Lars F. Brauer, and Neda Shabahang (legal research and footnote editing).

1. <i>Physical and Logical View of Computer Programs in Combination</i> .....	1444
2. <i>Interoperations and Data Communications Between Programs</i> .....	1449
B. PACKAGE WITH PROGRAMS OF ISOLATED FUNCTIONALITY .....	1450
1. <i>Basic Hypothetical</i> .....	1450
2. <i>Assessment</i> .....	1451
3. <i>Variations</i> .....	1451
C. MODIFICATION OF SCREEN OUTPUT THROUGH SOFTWARE COMBINATION.....	1452
1. <i>Hypothetical</i> .....	1452
2. <i>Assessment</i> .....	1453
a) Screen-Output Level.....	1455
i) Permanence.....	1455
ii) Substantial and Creative Portions.....	1456
iii) Combination Creativity .....	1456
iv) Significant Internal Changes .....	1456
v) Conclusion .....	1457
b) Underlying Code .....	1457
i) Permanence.....	1457
ii) Substantial and Creative Portions.....	1457
iii) Combination Creativity .....	1458
iv) Significant Internal Changes .....	1458
v) Conclusion .....	1459
D. ADD-ON PRODUCT WITH STATIC OR DYNAMIC LINKS, BUT WITH NO IMPACT ON SCREEN OUTPUT. ....	1459
1. <i>Hypothetical</i> .....	1459
2. <i>Assessment</i> .....	1461
a) Permanence .....	1461
b) Substantial and Creative Portions.....	1461
c) Combination Creativity .....	1462
d) Significant Internal Changes .....	1462
e) Summary .....	1463
E. WEB LINKING .....	1463
1. <i>Hypothetical</i> .....	1463
2. <i>Assessment of Y's Hyperlinks</i> .....	1463
3. <i>Assessment of Z's Inlinks</i> .....	1464
F. INTERFACE MODIFICATIONS .....	1465
G. SUMMARY .....	1465
VI. SOFTWARE COMBINATIONS UNDER COMMERCIAL SOFTWARE LICENSES.....	1466
A. FIRST SALE DOCTRINE.....	1467
B. TYPICAL LICENSE CLAUSES.....	1470
C. STATUTORY LIMITATIONS ON LICENSE RESTRICTIONS.....	1471

1. <i>Competition Law</i> .....	1471
2. <i>Copyright Misuse</i> .....	1473
3. <i>Unfair Contract Terms</i> .....	1476
D. LICENSE SCOPE DEFINITIONS VS. LICENSE CONDITIONS VS. CONTRACTUAL COVENANTS .....	1479
E. SUMMARY .....	1480
<b>VII. SOFTWARE COMBINATIONS UNDER THE GENERAL PUBLIC LICENSE</b> .....	1480
A. EFFECTS OF GPL § 2(B) .....	1483
B. GPL TERMINOLOGY AND INTERPRETATION .....	1487
1. “ <i>Works Based on the Program</i> ” .....	1487
2. <i>Derived Works</i> .....	1488
3. <i>Rules of Contract Interpretation</i> .....	1490
a) Parties’ Intent .....	1490
b) Interpretation Against the Drafter .....	1491
4. <i>Summary</i> .....	1491
C. FIRST SALE DOCTRINE.....	1492
D. STATUTORY LIMITATIONS ON LICENSE RESTRICTIONS.....	1492
1. <i>Competition Law</i> .....	1493
2. <i>Copyright Misuse</i> .....	1493
3. <i>Unfair Contract Terms</i> .....	1494
E. SUMMARY .....	1495
F. RECOMMENDATIONS FOR GPL VERSION 3.....	1496
<b>VIII. CONCLUSION</b> .....	1497

## I. INTRODUCTION

Companies have been fighting about software interoperability and substitutability for decades. The battles have usually involved wholesale copying and significant modifications of code to achieve compatibility, and the law seems fairly settled in both respects.<sup>1</sup> More recently, however, software developers and users alike have started to wake up to potential problems regarding combinations of separate programs, particularly in connection with open source software:<sup>2</sup> When do developers and users

---

1. For a good introduction and overview see MARK A. LEMLEY ET AL., *SOFTWARE AND INTERNET LAW* 2-45 (Aspen Publishers 1st ed. 2000) and the references in MARK A. LEMLEY ET AL., *SOFTWARE AND INTERNET LAW* (Aspen Publishers 2d ed. 2003).

2. See, e.g., Lothar Determann & Andrew Coan, *Spoiled Code? SCO v. Linux—A Case Study in the Implications of Upstream Intellectual Property Disputes for Software End Users*, 161 *COMPUTER L. REV. INT’L* (2003) (on file with author); see also The GPL-Violations.Org Project, <http://gpl-violations.org/> (last visited Jan. 21, 2006); Daniel Lyons, *Linux’s Hit Men*, [http://www.forbes.com/2003/10/14/cz\\_dl\\_1014linksys.html](http://www.forbes.com/2003/10/14/cz_dl_1014linksys.html) (last

need to obtain specific authorizations from the copyright owners before they may combine separate programs? What consequences can they expect for failure to obtain required authorizations? Fear, uncertainty and doubt (“FUD”) regarding the answers to these questions are prevalent in all quarters and have become a prominent topic in the computer lawyer community.<sup>3</sup>

This Article begins with a brief introduction to the issue and its context (Part II), examines the relevant copyright law principles in general (Part III) and the application of copyright law to software in particular (Part IV), goes on to illustrate the classification of software combinations under copyright law in a few common technical and commercial scenarios (Part V), and addresses the practical implications in the context of commercial (Part VI) and open source licensing (Part VII), particularly in light of the current debate surrounding the update of the General Public License (GPL). The Article concludes that most forms of software combinations are less dangerous (i.e., less likely to infringe copyrights) than commonly assumed because: (1) they do not constitute derivative works (but instead either compilations or *sui generis* aggregations outside the scope of the copyright owner’s exclusive rights); and (2) a number of statutes and legal doctrines significantly limit a copyright owner’s ability to contractually prohibit software combinations that do not also constitute derivative works under copyright law.

---

visited Dec. 30, 2005); Christian H. Nandan, *Open Source Licensing: Virus or Virtue?*, 10 TEX. INTELL. PROP. L.J. 349, 359 (2002) [hereinafter Nandan, *Open Source Licensing*].

3. FUD has historically referred to a marketing strategy employed by established companies to discredit new market entrants and the open source movement in the eyes of potential customers. Recently, however, even established companies complain about perceived FUD strategies by open source communities through intended vagueness and ambiguity of open source license agreements. At the 2005 ACI Conference on Software License Agreements in San Francisco, one presenter summarized his perception with a joke: “What do you get if you combine the Godfather and the GPL? An offer you can’t understand.” See David S. Evans & Bernard J. Reddy, *Government Preferences for Promoting Open-Source Software: A Solution in Search of a Problem*, 9 MICH. TELECOMM. & TECH. L. REV. 313, 340 (2003); Klaus M. Schmidt & Monika Schnitzer, *Public Subsidies for Open Source? Some Economic Policy Issues of the Software Market*, 16 HARV. J.L. & TECH. 473 (2003). More on the GPL will follow below in Part VII of this Article. For a summary of the history of the FUD expression, see Wikipedia.org, FUD, [http://en.wikipedia.org/w/index.php?title=Fear%2C\\_uncertainty\\_and\\_doubt](http://en.wikipedia.org/w/index.php?title=Fear%2C_uncertainty_and_doubt) (last visited Jan. 21, 2006). The legal community does not seem to be entitled to expect much sympathy regarding the suffering of FUD from the programmer community after deciding to apply the copyright regime to software in the first place. It also still has to play its own part in reducing uncertainties by reaching a clear consensus regarding the definitional scope of “derivative works” in the context of software combinations.

## II. BACKGROUND

Few computer programs function in isolation—most have to be combined with other software to perform their tasks. Just consider common personal computer software packages: when one starts an application program (e.g., Microsoft Word, Outlook, Adobe Acrobat, RealPlayer, etc.), it is actually an operating system program (e.g., Microsoft Windows) that “runs” the application program and saves and prints data files. Application programs are often used in tandem. For example, a user may cut and paste text from an Adobe Acrobat or MS Word document into an e-mail or open such a document from an e-mail attachment. Also, word processing and e-mail processing software may use separate programs (libraries) with definitions for fonts. Many applications use shared libraries instead of including the code within the application itself.<sup>4</sup>

From a technical perspective, in order to function in combination, programs have to be interoperable, meaning that they must be capable of exchanging and mutually using information.<sup>5</sup> Therefore, most software manufacturers try to ensure that their own programs are interoperable in order to market seamlessly integrated software suites. The agenda of other software manufacturers, however, depends on the market situation. For example, a company with an established platform may want to prevent interoperability with third party software in order to protect its market share for add-on programs or interests in hardware sales; new market entrants on the other hand will tend to promote interoperability with third party software in order to establish their platforms or to be able to offer add-ons or substitute programs for already established platforms.<sup>6</sup>

From a copyright law perspective, a combination of copyrighted programs typically requires one or more authorizations from the copyright owners. As a starting point, users normally need permission to copy a program in order to use it. In order to combine software or use software in combination, a user must typically first install (copy) the program(s) from the storage medium—floppy disc, CD, or DVD—to a computer’s hard drive. During the installation and execution of a computer program, the

---

4. See generally Wikipedia.org, Library (Computing), [http://en.wikipedia.org/w/index.php?title=Library\\_\(computing\)](http://en.wikipedia.org/w/index.php?title=Library_(computing)) (last visited Jan. 13, 2006).

5. Wikipedia.org, Interoperability, <http://en.wikipedia.org/wiki/Interoperability#Software> (last visited Dec. 29, 2005).

6. See, e.g., *Lexmark Int’l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 528 (6th Cir. 2004); *Chamberlain Group, Inc. v. Skylink Techs., Inc.*, 381 F.3d 1178, 1182 (Fed. Cir. 2004), *cert. denied*, 544 U.S. 923 (2005); *Sony Computer Entm’t, Inc. v. Connectix Co.*, 203 F.3d 596 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000); *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992).

actual data are literally copied several times between storage medium, hard drive, computer memory, and CPU cache (as further explained in Part V).

By copying the electronic data in different storage units of the computer, the user creates additional physical manifestations of the computer program(s). Courts in the United States qualify such physical manifestations as copies for purposes of copyright law.<sup>7</sup> Consequently, the copyright owner's permission is typically required before a user may install copyrighted software in order to combine it with other programs.<sup>8</sup>

A copyright owner also has the exclusive right to prohibit or permit the preparation of derivative works (adaptation right).<sup>9</sup> Consequently, if and to the extent the use of two copyrighted programs in combination constitutes the preparation of a derivative work, the user needs specific permission to combine the programs,<sup>10</sup> and the owners of the copyrights in the two programs have a statutory right to deny granting such permission.

Traditionally, the adaptation right has been regarded as redundant and commercially irrelevant in practice, given that most adaptations also involve copying.<sup>11</sup> After all, a copy of a derivative work also constitutes a non-literal copy of the adapted original.<sup>12</sup> Many cases involving the com-

---

7. *See, e.g.*, *CoStar Group, Inc. v. Loopnet*, 373 F.3d 544, 550-52 (4th Cir. 2004); *NLFC, Inc. v. Devcom Mid-Am., Inc.*, 45 F.3d 231, 235 (7th Cir. 1995); *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 518 (9th Cir. 1993). The final report of the Congressionally-appointed Commission on New Technological Uses of Copyrighted Works ("CONTU") has taken the position that the text of the new copyright law "makes it clear that the placement of any copyrighted work into a computer is the preparation of a copy, and therefore, a potential infringement of copyright." COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS, CONTU FINAL REPORT 30 (1978), available at <http://digital-law-online.info/CONTU/contu-toc.html>. At least one German court came to a different conclusion. *Landgericht [LG] Mannheim [Mannheim Trial Court]* Sept. 11, 1998, 15 (1999) *Computer und Recht [CR]* 360 (360-62) (F.R.G.).

8. 17 U.S.C. § 106 (2000 & Supp. II 2002).

9. *Id.* § 106(2).

10. Such permission could come in the form of a license from the copyright owner or a statutory exception to the copyright owner's exclusive rights. *See id.* § 117(a)(1).

11. REGISTER OF COPYRIGHTS, SUPPLEMENTARY REPORT ON THE GENERAL REVISION OF THE UNITED STATES COPYRIGHT LAW: 1965 REVISION BILL 17 (House Comm. Print 1965) [hereinafter REGISTER OF COPYRIGHTS, SUPPLEMENTARY REPORT]; Lydia P. Loren, *The Changing Nature of Derivative Works in the Face of New Technologies*, 4 J. SMALL & EMERGING BUS. L. 57, 64 (2000); *see also* MELVILLE B. NIMMER & DAVID NIMMER, 2 NIMMER ON COPYRIGHT § 8.09 (2006) [hereinafter NIMMER ON COPYRIGHT].

12. Hence, the Copyright Office believed that Section 106(2) of the Copyright Act was going to be largely duplicative yet helpful for purposes of clarification. *See* REGISTER OF COPYRIGHTS, SUPPLEMENTARY REPORT, *supra* note 11, at 17.

mercialization of non-literal copies that might also qualify as derivatives can already be resolved by finding infringement of duplication rights.<sup>13</sup>

In the software context, however, it can make quite a difference whether an end user seeking to combine two programs needs only a license to copy or additionally needs a license to prepare a derivative work. First, as a practical matter, commercial software programs typically come with end user licenses that expressly provide for a right to install and execute the software, but which are either silent regarding derivative works or expressly prohibit their preparation.<sup>14</sup> Second, purchasers who acquire ownership of particular software copies obtain certain statutory use rights, which include the right to execute the program, but not necessarily the right to create derivative works.<sup>15</sup> Third, a copyright owner who expressly prohibits certain software combinations is legally in a much stronger and clearer position if she can rely on her statutory adaptation right. By contrast, contractual covenants or conditions require privity and so may be subject to challenge under various legal theories, including copyright mis-

---

13. *See, e.g.*, *Lotus Dev. Corp. v. Borland Int'l*, 49 F.3d 807 (1st Cir. 1995) (discussing methods of operation with respect to copyrightability); *Computer Assocs. Int'l v. Altai, Inc.*, 982 F.2d 693 (2d Cir. 1992) (resolving case based on duplication rights rather than rights to derivative works); *Whelan Assocs., Inc. v. Jaslow Dental Lab., Inc.*, 797 F.2d 1222 (3d Cir. 1986) (finding copying of computer program's structure violated copyright). *But see, e.g.*, *Worlds of Wonder, Inc. v. Veritel Learning Sys.*, 658 F. Supp. 351, 356 (N.D. Tex. 1986) (holding that compatible cassette tapes for toy bear constitute derivative works); *Worlds of Wonder, Inc. v. Vector Intercontinental, Inc.*, 653 F. Supp. 135, 140 (N.D. Ohio 1986) (issuing preliminary injunction preventing competitors from marketing cassette tapes for use in toy bear where activation of the tapes was substantially similar to tapes created by toy bear's manufacturer).

14. *See, e.g.*, *Dun & Bradstreet Software Servs., Inc. v. Grace Consulting, Inc.*, 307 F.3d 197, 204 (3d Cir. 2002). Also, for instance, the Microsoft Office Word 2003 SP2 End User License Agreement grants the user a license to use the software and make a backup copy, but nowhere mentions the preparation of derivative works other than as it pertains to graphics and templates included with the program. Microsoft Office Word 2003 SP2 End User License Agreement, [www.microsoft.com/office/eula/en.msp](http://www.microsoft.com/office/eula/en.msp) (click on "Microsoft Office Word 2003" hyperlink to view PDF) (last visited Dec. 5, 2006). The Macromedia Dreamweaver 8 End User License Agreement explicitly prohibits the user from making derivative works. Macromedia Dreamweaver 8 End User License Agreement, [http://www.adobe.com/products/eula/third\\_party/dreamweaver/](http://www.adobe.com/products/eula/third_party/dreamweaver/) (last visited Dec. 5, 2006).

15. U.S. Copyright law provides a very limited right to create a derivative work, but only if and to the extent the "adaptation is created as an essential step in the utilization of the computer program in conjunction with a machine and that it is used in no other manner." 17 U.S.C. § 117 (2000). Given this narrow language, it is questionable whether the owner of a software copy is permitted to adapt its copy for purposes of interoperability with other software—as opposed to interoperability with a machine. *See generally* *Krause v. Titleserv*, 402 F.3d 119 (2d Cir. 2005) (discussing implementation of 17 U.S.C. § 117).

use, unconscionability, unfair competition law, and antitrust law.<sup>16</sup> Fourth, the GPL and other open source licenses tie specific restrictions and conditions to the creation of derivative works.<sup>17</sup>

Several courts have examined whether software combinations constitute derivative works. In video game cases where the end users had lawfully purchased or licensed game software, the game copyright owners had to rely on their adaptation right to fend off suppliers of add-on software that altered the games when executed in combination with them (e.g., by running the games faster or adding new game “levels”).<sup>18</sup> In another case, where a software manufacturer sought to enjoin a maintenance services provider from offering substitute “sub-programs,” the targeted customers held a valid license that allowed them to use at least unmodified versions of the plaintiff’s larger software suite.<sup>19</sup> In the web-linking context, internet users typically have an express or implied license to view the content on framed or linked websites, but website terms of use commonly prohibit any use of the site’s content that goes beyond mere viewing.<sup>20</sup>

Beyond these few cases, however, courts have not yet developed general rules for the qualification of software combinations as derivative works, nor have commentators articulated what such rules should be. The place and role of derivative works within the statutory context of compilations, collective works and other types of aggregations also have not been examined in depth with respect to software combinations.<sup>21</sup> This Article examines these issues in turn after briefly revisiting the concepts of derivative works and compilations under copyright law in general (Part III) and the characteristics of copyright protection for software in particular (Part IV).

---

16. See *infra* Section VI.C.

17. See *infra* Part VII.

18. *Micro Star v. Formgen*, 154 F.3d 1107, 1110 (9th Cir. 1998); *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965 (9th Cir. 1992), *cert. denied*, 507 U.S. 985 (1993); *Midway Mfg. v. Artic Int’l, Inc.*, 704 F.2d 1009 (7th Cir. 1983).

19. *Dun & Bradstreet*, 307 F.3d at 202.

20. Copying, modifying, creating derivative works, commercially exploiting and the like are typically either prohibited outright or made subject to the copyright owner’s express written consent. See, e.g., Ticketmaster Terms of Use, <http://www.ticketmaster.com/h/terms.html> (last visited Nov. 17, 2006); CNN.com, CNN Interactive Service Agreement, [http://www.cnn.com/interactive\\_legal.html](http://www.cnn.com/interactive_legal.html) (last visited Jan. 21, 2006).

21. See, e.g., Michael Gemignani, *Copyright Protection: Computer-Related Dependent Works*, 15 RUTGERS COMPUTER & TECH. L.J. 383 (1989); Christian H. Nandan, *A Proposal to Recognize Component Works: How a Teddy Bears on the Competing Ends of Copyright Law*, 78 CALIF. L. REV. 1633, 1641-42 (1990).

### III. COPYRIGHT LAW ON DERIVATIVE WORKS, COMPILATIONS, AND OTHER COMBINATIONS

#### A. Statutory Definitions

The U.S. Copyright Act defines and uses the term “derivative work” separately and in contrast to the terms “compilation” and “collective work,” as follows:

A “collective work” is a work, such as a periodical issue, anthology, or encyclopedia, in which a number of contributions, constituting separate and independent works in themselves, are assembled into a collective whole.

A “compilation” is a work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship. The term “compilation” includes collective works.

...

A “derivative work” is a work based upon one or more preexisting works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which a work may be recast, transformed, or adapted. A work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a “derivative work.”<sup>22</sup>

The definitions for all three categories share the primary requirement for copyright protection, namely, each must be an original work of authorship. Copyright law is intended to protect creative expression. Unlike patent law,<sup>23</sup> copyright law does not require novelty.<sup>24</sup> Thus, “originality” does not require that facts or ideas be expressed in an innovative way, but merely that the arrangement of facts not be so logical, mechanical, or rou-

---

22. 17 U.S.C § 101 (2000).

23. Regarding patent protection for software, see *State Street Bank & Trust v. Signature Financial Group, Inc.*, 149 F.3d 1368 (Fed. Cir. 1998) and Peter Weissman, *Computer Software as Patentable Subject Matter: Contrasting United States, Japanese, and European Laws*, 23 AIPLA Q.J. 525 (1995).

24. *eScholar, LLC v. Otis Educ. Sys., Inc.*, No. 04 Civ. 4051(SCR), 2005 WL 2977569, at \*14-\*15 (S.D.N.Y. Nov. 3, 2005); *Key Publ'ns, Inc. v. Chinatown Today Publ'g Enters., Inc.*, 945 F.2d 509, 512-13 (S.D.N.Y. 2005); *Bucklew v. Hawkins*, 329 F.3d 923, 929 (7th Cir. 2003).

tine as to require no creativity whatsoever.<sup>25</sup> A minimum “creative spark” is required.<sup>26</sup> With respect to derivative works, the changes to the original work must be creative; with respect to collective works and other compilations, the selection or arrangement must be creative.<sup>27</sup>

A compilation consists merely of the selection and arrangement of pre-existing material without any internal changes to the compiled material.<sup>28</sup> In comparison, a derivative work is created through internal changes to existing works<sup>29</sup> that actually affect these works—as opposed to additions or parts of a combination appearing merely in some loose context, detached from the original work.<sup>30</sup>

If the creator of a new work takes very little of an existing work,<sup>31</sup> taking only non-protectable content such as ideas or facts, or changing the original so much that the new work differs substantially from the existing work, the new creation is simply a new work of authorship<sup>32</sup> and not a derivative of the existing work.<sup>33</sup> Most new works, after all, are influenced to some extent by existing works.<sup>34</sup>

---

25. *Feist v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991) (finding white pages in a telephone directory insufficiently creative to merit copyright protection).

26. *Silverstein v. Penguin Putnam, Inc.*, 368 F.3d 77, 83 (2d Cir. 2004).

27. *Key Publications*, 945 F.2d at 513; *Silverstein*, 368 F.3d at 83.

28. H.R. Rep No. 94-1476 at 57 (1976).

29. *Id.* See generally 1 NIMMER ON COPYRIGHT, *supra* note 11, § 3.02.

30. *Bucklew v. Hawkins*, 329 F.3d 923, 930 (7th Cir. 2003). The court notes that: [I]f the original expression added by the unauthorized preparer of a derivative work is clearly detachable from the original work itself, so that no confusion, or disruption of the copyright owner’s plans for the exploitation of his work, would be created by allowing the unauthorized preparer to copyright his original expression, the unauthorized preparer might be allowed to do so . . . though this principle may be limited to compilations, where “the infringing portion would be easily severable and the scope of the compilation author’s own work . . . would be easily ascertainable.”

*Id.*

31. *Superchips, Inc. v. Street & Performance Elecs., Inc.*, No. 01 Civ. 309, 2001 WL 1795939, at \*3 (M.D. Fla. Apr. 4, 2001) (“[C]ourts consider whether there is a distinguishing variation between the derivative and underlying work and whether that variation is more than ‘merely trivial.’” (internal citation omitted)); *Vault Corp. v. Quaid Software, Ltd.*, 655 F. Supp. 750, 758-59 (E.D. La. 1987), *aff’d*, 847 F.2d 255, 267 (5th Cir. 1988).

32. *Castle Rock Entm’t, Inc. v. Carol Publ’g Group, Inc.*, 150 F.3d 132, 138-39 (2d Cir. 1998).

33. *Bucklew*, 329 F.3d at 930; *Pickett v. Prince*, 207 F.3d 402, 407 (7th Cir. 2000) (distinguishing “works only loosely connected with some ancestral work claimed to be their original”); *Vault*, 655 F. Supp. at 758; 1 NIMMER ON COPYRIGHT, *supra* note 11, § 3.02.

34. *Emerson v. Davies*, 8 F. Cas. 615, 619 (C.C.D. Mass. 1845).

Thus, compilations, derivative works, and entirely new works typically involve combinations of new creative materials with existing material. In the case of a compilation, the existing material remains intact and unchanged, and the “combination creativity” remains separate and clearly distinguishable from the existing material.<sup>35</sup> In the case of a non-derivative new work, existing material may be remotely reflected in the new work, but its contribution is insubstantial. The derivative work category lies somewhere in the middle: existing creative material constitutes a substantial part of the new derivative work, and the new creative material appears in the form of inseparable changes to the existing material.

### **B. Examples of Combinations**

Examples of derivative works include a translation of a poem into another language, an orchestra arrangement of a piano sonata, a rap version of a Beatles’ song, a movie based on a comic book, and a theatre drama based on a novel. In all these cases, both the derivative and the underlying work have to meet the originality requirement of the Copyright Act.<sup>36</sup> Thus, both the author of the underlying work and the creator of the derivative work have to supply at least a “creative spark.”<sup>37</sup>

The same is true for collective works, which are creative collections of copyrighted works: both the collection (i.e., the selection of works) and the collected works have to be creative in nature.<sup>38</sup> The term “compilations” includes creative collections of creative works and creative compilations of non-creative materials such as names and phone numbers of actual persons. In both cases, the compilation itself has to be creative. Thus, an arrangement of creative or non-creative material in a purely logical order will not receive protection under copyright law.<sup>39</sup> Examples of copyrightable compilations include creative catalogues and collections of poems.<sup>40</sup>

---

35. The creative combination is essentially a meta-layer—the glue that holds the existing material together.

36. *Silverstein v. Penguin Putnam, Inc.*, 368 F.3d 77, 83 (2d Cir. 2004).

37. *See, e.g., id.* at 83.

38. *Id.*

39. *See Feist v. Rural Tel. Serv. Co.*, 499 U.S. 340, 362-64 (1991); *Silverstein*, 368 F.3d at 83; *Matthew Bender & Co., Inc. v. West Publ’g Corp.*, 158 F.3d 674, 682-83 (5th Cir. 1992).

40. 1 NIMMER ON COPYRIGHT, *supra* note 11, § 3.02.

If someone applies minor changes to a work without any originality,<sup>41</sup> the result will not constitute a derivative work, but only a non-literal copy of the work.<sup>42</sup> Similarly, if someone prepares a collection of copyrighted or non-copyrighted material without any creativity, such as in historical or alphabetical order, the result will not constitute a copyrightable compilation, but rather a series of literal copies (if the collected material is copyrighted) or an arrangement that is entirely outside the scope of copyright law.<sup>43</sup> If an author creates a new work, borrowing only minor aspects from existing works, the result qualifies as an independently created new work, not a derivative work.<sup>44</sup>

Thus, for copyright law purposes, a combination of new and existing material may constitute one or more of the following:

- a new (non-derivative) work if only very little of or non-protectable elements of the existing materials are present in the new work or if the new work does not bear a substantial resemblance to the existing work;<sup>45</sup>
- a derivative work if new material changes the substance of the existing material and both are creative, e.g., a song based on a poem;<sup>46</sup>

---

41. Examples of unprotectable works include copies with typos, typo corrections using spell-check, and a few missing words. *See, e.g.*, *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 966 (9th Cir. 1992), *cert. denied*, 507 U.S. 985 (1993); *Silverstein*, 368 F.3d at 83; *Matthew Bender*, 158 F.3d at 681 n.4.

42. *Lewis Galoob Toys*, 964 F.2d 965. *But see* *Bucklew v. Hawkins*, 329 F.3d 923, 923 (7th Cir. 2003); *Sherry Mfg. Co. v. Towel King of Fla., Inc.*, 753 F.2d 1565, 1568 (11th Cir. 1985); *Superchips, Inc. v. Street & Performance Elecs., Inc.*, No. 01 Civ. 309, 2001 WL 1795939, at \*3-\*4 (M.D. Fla. Apr. 4, 2001).

43. *Feist*, 499 U.S. at 362-64; *see also Silverstein*, 368 F.3d at 83.

44. *See Pickett v. Prince*, 207 F.3d 402, 407 (7th Cir. 2000) (distinguishing “works only loosely connected with some ancestral work claimed to be their original”).

45. *See, e.g.*, *Well-Made Toy Mfg. Corp. v. Goffa Int’l Corp.*, 354 F.3d 112, 117 (2d Cir. 2003) (finding that the toy doll was not substantially similar to competitor’s product and therefore did not constitute an infringing derivative work); *Ty, Inc. v. Publ’ns Int’l Ltd.*, 292 F.3d 512, 521 (7th Cir. 2002) (holding that “Beanie Babies” collectors’ guide was not a derivative work, but a “public evaluation,” which the toy manufacturer was not entitled to control under its copyright).

46. *See, e.g.*, *Lamb v. Starks*, 949 F. Supp. 753, 757 (N.D. Cal. 1996) (regarding a trailer as a derivative work of a movie); *Shoptalk, Ltd. v. Concorde-New Horizons Corp.*, 897 F. Supp. 144, 147 (S.D.N.Y. 1995) (finding a movie as derivative work of screenplay), *aff’d in part, vacated in part*, 168 F.3d 586 (2d Cir. 1999), *cert. denied*, 527 U.S. 1038 (1999).

- a compilation if existing creative or non-creative material is arranged in a creative manner, e.g., a collection of songs, poems, and/or facts related to the holiday season;<sup>47</sup>
- a non-literal copy if new material makes insubstantial, non-creative changes to the substance of the existing material, only the existing material is creative, and the end result is nearly identical, e.g., publication of a poem with typo corrections;<sup>48</sup>
- a literal copy if existing creative material is arranged and reproduced in a non-creative manner, e.g., a collection of all poems by a particular author by titles in alphabetic order;<sup>49</sup> and/or
- an arrangement that is neither restricted nor protected by copyright law if the combination does not involve any changes, duplication, or creative arrangement, e.g., storage of books on a shelf sorted by the author's name in alphabetical order.<sup>50</sup>

---

47. See, e.g., *Mason v. Montgomery Data, Inc.*, 967 F.2d 135, 141 (5th Cir. 1992) (holding collection of real estate ownership maps incorporating information from various sources was “sufficiently creative to qualify . . . as original ‘compilations’” because author had engaged in “selection, coordination, and arrangement of the information . . . depicted”); *Corsearch, Inc. v. Thomson & Thomson*, 792 F. Supp. 305, 322 (S.D.N.Y. 1992) (finding state trademark computer database was protectable compilation where copyright proponent had selected, coordinated, arranged, enhanced, and programmed the trademark data); cf. HOWARD B. ABRAMS, *THE LAW OF COPYRIGHT* § 1:16 (2006) (stating that the “majority of compilations will pass [minimal level of creativity] test”).

48. See, e.g., *Signo Trading Int'l, Ltd. v. Gordon*, 535 F. Supp. 362, 365 (N.D. Cal. 1981) (holding that a list of words, translated into foreign language, did not constitute a copyrightable compilation because the author of the translation had not selected words on the list). For infringement liability by reason of non-literal copying, see generally *Castle Rock Entertainment, Inc. v. Carol Publishing Group, Inc.*, 150 F.3d 132, 140 (2d Cir. 1998), finding that a “Seinfeld” trivia book infringed copyright in the television show, even though “direct quotations or close paraphrases . . . copied from the Seinfeld series [were] few and almost irrelevant.” See also *Twin Peaks Prods., Inc. v. Publ'ns Int'l, Ltd.*, 996 F.2d 1366, 1372-73 (2d Cir. 1993) (finding infringement of a copyrighted television series, as shown by “comprehensive nonliteral similarity,” where the book contained a “detailed recounting of . . . episodes of the series”).

49. See *Silverstein v. Penguin Putnam, Inc.*, No. 01 Civ. 309, 2003 WL 1797848, at \*7 (S.D.N.Y. April 4, 2003) (granting summary judgment for plaintiff on claim of infringement of poem compilation where defendant copied “nearly the entire work,” including the original author’s “selection and . . . guiding principles”), *rev'd, vacated and remanded*, 368 F.3d 77 (2d Cir. 2004) (finding that a summary judgment and an injunction were not appropriate because issue of fact existed as to whether plaintiff’s arrangement was copyrightable in the first instance).

50. See, e.g., *Paramount Pictures Corp. v. Video Broad. Sys., Inc.*, 724 F. Supp. 808, 821 (D. Kan. 1989) (regarding defendant who inserted ads for local business at beginning of movie videotapes). While both ads and movie were copyrightable works, adding the

These categories are not mutually exclusive: an author could translate poems (prepare derivative works), add some poems without translation but edit for typographical corrections (non-literal copies) and then creatively arrange the poems (create a collective work).

### C. Ownership of Derivative Works vs. Compilations

The Copyright Act treats compilations and derivative works similarly with respect to copyright subject matter and ownership rights:<sup>51</sup> the author of a derivative work or compilation owns the copyrights in her creative contributions, but not in the underlying work.<sup>52</sup> Consequently, the creator of a derivative work or compilation can exclude anyone—including the owner of the copyrights in the underlying work(s)—from copying or distributing the derivative work or compilation.<sup>53</sup>

If a creator of a derivative work or compilation bases her work unlawfully on copyrighted works of others, she does not acquire any copyrights in the derivative work or compilation.<sup>54</sup> Additionally, a licensee who creates literal or non-literal copies that do not amount to derivative works under license from the copyright owner does not acquire any copyright in such copies.

---

ads did not amount to “recasting, transforming or adapting the motion picture,” such as creation of a derivative work. *Id.*

51. 17 U.S.C. § 103 (2000). The section provides the following:

Subject matter of copyright: Compilations and derivative works. (a) The subject matter of copyright as specified by section 102 includes compilations and derivative works, but protection for a work employing preexisting material in which copyright subsists does not extend to any part of the work in which such material has been used unlawfully. (b) The copyright in a compilation or derivative work extends only to the material contributed by the author of such work, as distinguished from the preexisting material employed in the work, and does not imply any exclusive right in the preexisting material. The copyright in such work is independent of, and does not affect or enlarge the scope, duration, ownership, or subsistence of, any copyright protection in the preexisting material.

*Id.*

52. *Id.* § 103(b).

53. *Id.* §§ 103(b), 106.

54. *Id.* § 103. In this respect, creators of derivative works and compilations are in slightly different positions: the creator of a derivative work needs an authorization (i.e., license) from the owner of the copyright in the underlying work(s) to create the derivative work in the first place—and then later to reproduce or distribute it. Such a license must specifically allow the preparation of derivative works. With respect to compilations, such a specific authorization is not required to ensure the lawfulness and thus acquisition of ownership rights in the compilation. *See infra* Section III.D.

#### D. Exclusive Rights to Derivative Works vs. Compilations

Given the similar treatment of derivative works and compilations for ownership purposes, it is worth noting—and particularly relevant for the permissibility of software combinations—that the Copyright Act treats derivative works and compilations very differently with respect to the scope of exclusionary rights. A copyright owner has the exclusive right to prohibit or authorize the preparation of derivative works.<sup>55</sup> The copyright owner does not, however, have an exclusive right to prohibit or authorize the preparation of compilations or non-creative arrangements of works.<sup>56</sup> Thus, the Copyright Act specifically empowers the creator of artwork to prohibit a buyer of prints from selling tiles with framed prints attached,<sup>57</sup> but not from arranging unmodified prints in a creative or non-creative compilation such as side-by-side with other artwork.<sup>58</sup> Consequently, the distinction between derivative works and other categories of combinations is the most crucial for software interoperability.

The special treatment of derivative works under copyright law is rooted in a recognition of the special relationship that authors traditionally have with their works, in contrast to the treatment of compilations and, under patent law, in contrast to the treatment of improvements.<sup>59</sup> An author of a copyrighted novel, painting, or symphony is affected personally and commercially if a publisher adds a happy ending to the novel, removes potentially offensive scenes from the painting, or eliminates a movement from the symphony. Such changes can affect the author's reputation and ability to commercialize future works if the public cannot easily separate the changes from the original work.<sup>60</sup> Transparent and detached

---

55. 17 U.S.C. § 106(2).

56. 17 U.S.C. § 106(1)-(3) mention only derivative works, but not collective works or other types of compilations..

57. *See* *Mirage Editions, Inc. v. Albuquerque A.R.T. Co.*, 856 F.2d 1341, 1343-44 (9th Cir. 1988). *But see* *Lee v. A.R.T.*, 125 F.3d 580, 582 (7th Cir. 1997).

58. Of course, a copyright owner can independently prohibit copying of her work, so any combinations that involve duplication require separate permission. 17 U.S.C. § 106(1). Thus the creator of the artwork here could prohibit the buyer of the prints from arranging the unmodified prints in a book, since that would involve duplicating the prints.

59. In discussing this issue, Mark Lemley uses the term “improvers” to refer to those who make works or inventions based upon, to varying degrees, an underlying work or patent. He divides “improvers” into three categories: minor improvers, significant improvers, and radical improvers. *See generally* Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 1007 (1997) [hereinafter Lemley, *Economics of Improvement*].

60. This phenomenon is relevant beyond the context of continental European copyright laws that protect authors against distortion of their works, because even jurisdictions that consider an author's rights against distortion inalienable will typically allow an as-

combinations that can easily be undone conceptually, such as compilations or improvements on patented inventions, do not affect the interests in the underlying intellectual property in the same way because others can easily differentiate between the original version and the improvements or additions. Therefore, the requirement of internal changes to the adapted work is an important definitional element for derivative works under copyright law that justifies the legislative decision granting copyright owners the right to prohibit adaptations.

### E. Fixation Requirement

The Copyright Act provides a fixation requirement for ownership purposes, but does not specify how permanent a derivative work must be to infringe.<sup>61</sup> In some cases, courts have simply assumed that the same permanency threshold applies for both ownership and infringement purposes.<sup>62</sup> In other cases, courts have asserted that a different standard applies for infringement purposes, albeit without clearly defining exactly what constitutes that standard.<sup>63</sup> Overall, however, it seems generally accepted that adaptations do not constitute derivative works for infringement or ownership purposes if they are fleeting and lack any significant permanency—such as a work viewed through a pink filter.<sup>64</sup>

---

signment or perpetual license to derivative works. *See, e.g.*, C. PROP. INTEL. [Intellectual Property Code], art. L111-1, L113-1 (Fr.); *see also Lee*, 125 F.3d at 582 (criticizing the 9th Circuit's decision in *Mirage Editions* as granting a backdoor moral right to authors via the derivative works doctrine); 2 PAUL GOLDSTEIN, COPYRIGHT § 5.3 n.12 (2d ed. 2002); Edward J. Damich, *The Right of Personality: A Common-Law Basis for the Protection of the Moral Rights of Authors*, 23 GA. L. REV. 1, 41 (1988); Lemley, *Economics of Improvement*, *supra* note 59, at 1033. *See generally* Amy B. Cohen, *When Does a Work Infringe the Derivative Works Right of a Copyright Owner?*, 17 CARDOZO ARTS & ENT. L.J. 623, 645-46 (1999); Ian Eagles & Louise Longdin, *Technological Creativity and Moral Rights: A Comparative Perspective*, 12 INT'L J.L. & INFO. TECH. 209, 232-33 (2004).

61. 17 U.S.C. § 101 defines when a work is “created,” but 17 U.S.C. § 106(2) does not refer to the “creation,” but rather to the “preparation” of derivative works, a more generic term that is used also in other sections of the Copyright Act.

62. *See, e.g.*, *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1518 (9th Cir. 1992).

63. *See, e.g.*, *Micro Star v. Formgen*, 154 F.3d 1107, 1110-11 (9th Cir. 1998); *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 967-68 (9th Cir. 1992), *cert. denied*, 507 U.S. 985 (1993); 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.09[A] (differentiating between infringements of the adaptation right through performances as opposed to copies).

64. One court used the “low tech” example of a pink piece of cellophane placed in a frame over a television screen. *Micro Star*, 154 F.3d at 1111 n.4; *see also* Tyler Ochoa, *Symposium Review: Copyright, Derivative Works and Fixation: Is Galoob a Mirage, or*

## F. Summary

In addition to the right to control duplication, copyright owners have an express statutory right (adaptation right) to prohibit or permit combinations of their works with other materials if and to the extent such combinations qualify as derivative works. Combinations qualify as derivative works only if they are sufficiently permanent, contain significant amounts of existing copyrighted works, and involve significant and creative changes to such pre-existing works. Combinations without internal changes can qualify as compilations if the combination is creative, but unlike with derivative works, a copyright holder has no express statutory right to prohibit end users from making compilations.<sup>65</sup> Neither insignificant nor non-creative changes to existing copyrighted works, or entirely insignificant adaptations from existing works, result in the creation of a derivative work. Instead, they result in the creation of non-literal copies or non-derivative new works, respectively.

## IV. SOFTWARE UNDER COPYRIGHT LAW

The peculiar relationship between software and copyright law has already been well analyzed.<sup>66</sup> To lay the groundwork for the following analysis, however, it is helpful to briefly recall how two important principles of copyright law apply in the software context: (1) copyright law strikes a delicate balance between access and exclusion rights, and in do-

---

*Does the Form(gen) of the Alleged Derivative Work Matter?*, 20 SANTA CLARA COMPUTER & HIGH TECH. L.J. 991 (2004). *But see* Lydia P. Loren, *The Changing Nature of Derivative Works in the Face of New Technologies*, 4 J. SMALL & EMERGING BUS. L. 57, 83-84 (2000); Emilio B. Nicolas, *Why the Ninth Circuit Added Too Much to Subtract Add-on Software from the Scope of Derivative Works Under 17 U.S.C. § 106(2): A Textual Argument*, 2004 SYRACUSE SCI. & TECH. L. REP. 1, 44-45 (2004).

65. The copyright holder does retain the right to authorize copying, so the compilation cannot involve any unlicensed copying of the underlying works. If an end user holds a license to execute and use two programs, however, and their execution in combination constitutes a compilation, such combination would not infringe the software copyright owners' exclusive rights under Section 106 of the Copyright Act. Where a combination qualifies as a derivative work, on the other hand, and all other things are equal, the end user would need a separate license to combine the programs; without such license, the combination would infringe the copyright owners' adaptation rights under Section 106(2) of the Copyright Act.

66. *See, e.g.*, Mark A. Lemley, *Convergence in the Law of Software Copyright*, 10 HIGH TECH. L.J. 1, 3-6 (1995); Peter S. Menell, *Tailoring Legal Protection for Computer Software*, 39 STAN. L. REV. 1329 (1987); Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994).

ing so, (2) copyright law protects only creative expression, not functionality, however valuable such functionality may be.

### A. The Balance Between Access and Exclusion Rights

As contemplated by the U.S. Constitution,<sup>67</sup> the Copyright Act protects investments in creative works through exclusion rights that encourage further creation and public availability of such works.<sup>68</sup> Exclusion rights enable the owner to permit or prohibit the prescribed activities—and charge fees for use (licenses). The prospect of such license fees are intended to incentivize creators to create and adapt original works of authorship.

Both under- and over-protection can harm the public interest in creative works.<sup>69</sup> This risk is particularly obvious with respect to the adaptation right. Since most creative works borrow, to some extent, from existing material, overbroad adaptation rights could seriously stifle further developments.<sup>70</sup> Thus, legislatures and courts have over the years struck a delicate balance between granting and limiting exclusion rights for authors and access rights for the public.<sup>71</sup> Copyright owners who upset this balance by abusing their rights can be penalized by the denial of copyrights under the doctrine of copyright misuse.<sup>72</sup>

### B. Creative Expression vs. Functionality

The Copyright Act, in balancing creative expression and functionality, takes great care to limit copyrightable subject matter to creative, artistic expression and to keep underlying ideas and functionality in the public domain. The Act states that

---

67. U.S. CONST. art. I, § 8 provides that “Congress shall have the power . . . [t]o promote the Progress of Science and useful Arts, by securing for limited times to Authors and Inventors the exclusive right to their respective Writings and Discoveries.”

68. See Pamela Samuelson, *Computer Programs, User Interfaces, and Section 102(b) of the Copyright Act of 1976: A Critique of Lotus v. Paperback*, 55 SPG LAW & CONTEMP. PROBS. 311, 339 (1992) [hereinafter Samuelson, *Computer Programs*].

69. See *id.* at 338-39.

70. 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.09. See also Daniel S. Hurwitz, *A Proposal in Hindsight: Restoring Copyright’s Delicate Balance by Reworking 17 U.S.C. § 1201*, 2005 UCLA J.L. & TECH. 1 (2005); cf. Christine Jeanneret, *The Digital Millennium Copyright Act: Preserving the Traditional Copyright Balance*, 12 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 157 (2001).

71. See, e.g., *Eldred v. Ashcroft*, 537 U.S. 186 (2003); *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539 (1985); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984). Where this line should be drawn has been the subject of extensive academic debate, especially in the last decade.

72. See *infra* Section VI.C.2.; *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 977 (4th Cir. 1990). See generally 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A].

[i]n no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.<sup>73</sup>

This express exception historically made software an unlikely candidate for copyright protection given the fact that the value of most<sup>74</sup> computer programs lies in their functionality and efficiency.<sup>75</sup> Software licensees typically appreciate and pay for the speed, reliability, and the operational simplicity of particular programs, but do not value, for example, the creativity or originality of the underlying code.

Yet, the threshold question of whether software is protected by copyright law at all has long been settled in the United States.<sup>76</sup> For over three decades, U.S. courts have consistently found that computer programs (both object and source code versions) are generally protected under the U.S. Copyright Act.<sup>77</sup> The underlying code constitutes a literary work,<sup>78</sup> whereas output such as screen displays and user interfaces can be protected separately as audiovisual works.<sup>79</sup> It is possible to generate the same screen displays and user interfaces with substantially different underlying code. Accordingly, a program's screen display may be infringing while the actual code is not.<sup>80</sup>

---

73. 17 U.S.C § 102(b) (2000).

74. Video and computer games and other entertainment software are noteworthy exceptions.

75. David G. Luetgen, *Functional Usefulness vs. Communicative Usefulness: Thin Copyright Protection for the Nonliteral Elements of Computer Programs*, 4 TEX. INTEL. PROP. L.J. 233, 249-60 (1996).

76. See, e.g., *Apple Computer, Inc. v. Franklin Computer Corp.*, 714 F.2d 1240 (3d Cir. 1983), cert. denied, 464 U.S. 1033 (1984); *Harcourt, Brace & World, Inc. v. Graphics Controls Corp.*, 329 F. Supp. 517 (S.D.N.Y. 1971). For an overview, see MARK LEMLEY ET AL., *SOFTWARE AND INTERNET LAW*, 1-45, 97-98 (1st ed. 2000); MARK LEMLEY ET AL., *SOFTWARE AND INTERNET LAW*, 33-35 (2d ed. 2003).

77. See, e.g., *Franklin Computer*, 714 F.2d 1240; *Williams Elecs., Inc. v. Artic Int'l, Inc.*, 685 F.2d 870 (3d Cir 1982); *Harcourt, Brace & World*, 329 F. Supp. 517.

78. *Franklin Computer*, 714 F.2d at 1246-49.

79. *Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435 (9th Cir. 1994); *Data E. USA, Inc. v. Epyx, Inc.*, 862 F.2d 204 (9th Cir. 1988); *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852 (2d Cir. 1982).

80. *Lotus Dev. Corp. v. Borland Int'l*, 49 F.3d 807 (1st Cir. 1995) (involving user interface command hierarchies); *Computer Assocs. Int'l v. Altai, Inc.*, 982 F.2d 693, 703 (2d Cir. 1992). For an analysis of the decision, see Dennis S. Karjala & Peter S. Menell, *Applying Fundamental Copyright Principles to Lotus Development Corp. v. Borland International, Inc.*, 10 HIGH TECH. L.J. 177 (1995).

Despite the conceptual mismatch, copyright law has established itself globally as the primary intellectual property regime for software, and jurisdictions that initially rejected this approach have now adopted it.<sup>81</sup> Yet, since courts first made the decision that software was copyrightable, they have struggled with the fundamental problem that copyright law is designed to protect creative expression as an incentive for further creative activity, whereas the value in software is usually measured by functionality and efficiency,<sup>82</sup> aspects that are expressly excluded from copyright protection.<sup>83</sup> In the process, courts complained about having to fit a square peg into the round hole<sup>84</sup> and thus developed a number of tests and approaches to separate protectable creative elements from non-protectable functional elements in software.<sup>85</sup> Creative elements are protected against

---

81. An example is Germany: software had been expressly recognized in Section 2 of the German Copyright Act as a category of copyrightable works since 1985. Urheberrechtsgesetz, UrhG [Copyright Law] Sept. 9, 1965, as amended May 8, 1998, § 2 no. 2 (F.R.G.). However, prior to the implementation of the EC Software Directive into German law in 1993, German courts had required a very high level of originality before they would afford copyright protection for software. The leading cases are from 1985 and 1991: Bundesgerichtshof [BGH] [Federal Court of Justice] May 19, 1985, 12 Gewerblicher Rechtsschutz und Urheberrecht [GRUR] 1041 (F.R.G.) (known as the collection program case—"Inkasso-Programm") and Bundesgerichtshof [BGH] [Federal Court of Justice] Oct. 4, 1990, 44 (1991) *Neue Juristische Wochenschrift* [NJW] 1231 (F.R.G.) (known as the operating system case—"Betriebssystem"). Many programs that would have easily qualified as copyrightable in the United States were not found to be so in Germany. For an overview of software copyright protection in the European Union, see Pamela Samuelson, *Comparing U.S. and EC Copyright Protection for Computer Programs: Are They More Different Than They Seem?*, 13 J.L. & COM. 279 (1994).

82. See *Bucklew v. Hawkins*, 329 F.3d 923, 928 (7th Cir. 2003).

83. *Altai*, 982 F.2d at 712; see also Mark A. Lemley & David W. O'Brien, *Encouraging Software Reuse*, 49 STAN. L. REV. 255 (1997) (arguing that copyright protection for software inhibits efficiency); Peter S. Menell, *An Analysis of the Scope of Copyright Protection for Application Programs*, 41 STAN. L. REV. 1045 (1989); Pamela Samuelson et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994); cf. Stacey L. Dogan & Joseph P. Liu, *Copyright Law and Subject Matter Specificity: The Case of Computer Software*, 61 N.Y.U. ANN. SURV. AM. L. 203, 204 (2005) (discussing judicial treatment of computer software copyright protection in light of its dual nature); Jane C. Ginsburg, *Four Reasons and a Paradox: The Manifest Superiority of Copyright Over Sui Generis Protection of Computer Software*, 94 COLUM. L. REV. 2559 (1994) [hereinafter Ginsburg, *Four Reasons*]; Arthur R. Miller, *Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?*, 106 HARV. L. REV. 977 (1993).

84. See, e.g., *Altai*, 982 F.2d at 712.

85. *Chamberlain Group, Inc. v. Skylink Techs., Inc.*, 381 F.3d 1178 (Fed. Cir. 2004), cert. denied, 544 U.S. 923 (2005); *Lexmark Int'l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 534-36 (6th Cir. 2004); *Lotus Dev. Corp.*, 49 F.3d at 815-17;

literal and non-literal copying, whereas functional elements are in the public domain and can be freely duplicated, even where idea and expression or functionality and creativity merge, for example, because a particular technical solution can be programmed efficiently only in one particular manner.<sup>86</sup>

Courts have defined the dividing line on a case-by-case basis in light of the underlying public policy considerations. Artistic screen displays of computer games (with fantasy figures and landscapes)<sup>87</sup> bear a greater resemblance to traditional subjects of copyright protection (like novels and paintings) than to software in executable form (not readable by humans and consisting of zeros and ones) or to the functionality-driven user interfaces for application programs.<sup>88</sup> Therefore, computer game screen displays have generally fared better in cases where litigants have raised issues of non-literal copying and the idea-expression dichotomy.

Some courts had to confront situations where companies used otherwise creative works in a purely functional manner, for example, as interfaces, passwords, or lock-out mechanisms. Courts denied copyright protection for such works, regardless of how creative and original they were, in the interest of preserving the balance between protection and access rights described above.<sup>89</sup> Thus, as a general matter, copyright owners cannot rely on the protections that copyright law affords where they deploy copyrighted works in a software context for the sole purpose of forcing others to infringe (by copying or adapting copyrighted code) in order to establish interoperability.<sup>90</sup>

Other courts tried to stretch the boundaries of copyright law in order to protect investments, even where the Copyright Act did not literally cover the material or activities that concerned the plaintiffs.<sup>91</sup> Since the U.S. Supreme Court vehemently rejected the “Sweat of the Brow” doctrine in 1991, however, it is important to start by thoroughly examining the crea-

---

Gates Rubber Co. v. Bando Chem. Indus., Ltd., 9 F.3d 823, 836-38 (10th Cir. 1993); *Altai*, 982 F.2d at 714.

86. See *Altai*, 982 F.2d at 707-08.

87. See, e.g., *Micro Star v. Formgen*, 154 F.3d 1107, 1110 (9th Cir. 1998).

88. See, e.g., *Lotus Dev. Corp.*, 49 F.3d 807; *Altai*, 982 F.2d 693.

89. See *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992) (holding that reverse engineering to obtain interface information for compatibility is permissible under narrow circumstances); see also *Skylink*, 381 F.3d 1178; *Lexmark*, 387 F.3d 522.

90. See *Lexmark*, 387 F.3d 522; *Sega*, 977 F.2d 1510.

91. See, e.g., *Midway Mfg. v. Artic Int'l, Inc.*, 704 F.2d 1009, 1014 (7th Cir. 1983) (“[T]he amount by which the language of Section 101 must be stretched to accommodate speeded-up video games is, we believe, within the limits within which Congress wanted the new Act to operate.”).

tivity vs. functionality dichotomy in any software copyright analysis. One must then filter out ideas, processes, methods, facts, elements dictated by external factors or efficiency, material in the public domain, expression which has “merged” with any of the foregoing, and expression which is so standard or common as to be a “necessary incident” to any of the foregoing.<sup>92</sup>

This filtering technique is equally important for the characterization of software combinations as derivative works, compilations, or other aggregations. Given the utilitarian nature of software, modifications that are dictated by external factors or functionality considerations do not generally affect the special relationship between the author and her creative work, which adaptation rights are intended to protect under copyright law. Contrast this situation with patent law, which, in order to avoid stifling innovation in functionality, does not grant a patent holder exclusive rights to improvements to her invention. Hence, courts have to be particularly careful to focus on creative expression—as opposed to functionality—when drawing the line between combinations that infringe adaptation rights (derivative works) and combinations that do not (because they constitute compilations or other aggregations that are generally permissible without the copyright owner’s consent). Internal changes to code that software users or other developers cannot easily separate or distinguish tend to be capable of affecting an author’s interests in adaptation of her creative work, thereby strongly supporting classification as a derivative work. In contrast, combinations with separable add-on programs or interface modifications dictated by functional requirements do not support a classification as derivative works, as they may be relevant for the program’s functionality, but do not significantly affect an author’s reasonable interest in controlling adaptations of her creative works.

### C. Summary

Given its typically utilitarian nature, software usually contains many elements that are excluded from copyright protection and need to be filtered out at the outset of any copyright analysis. Consequently, the coverage and strength of copyright protection for software varies depending on its different forms (including object code, source code, graphical user interface, etc.) and functionalities. The screen output of computer games, for example, can be highly creative and thus enjoy similarly strong protection as more traditional works of authorship, such as paintings, novels, or musical compositions. The underlying code of the same games, on the other

---

92. *Gates Rubber Co. v. Bando Chem. Indus., Ltd.*, 9 F.3d 823, 837-38 (10th Cir. 1993).

hand, is usually more functional in nature, derives value from execution efficiency as opposed to artistic creativity, and contains a thinner layer of copyrightable expression (over uncopyrightable functionality). Copyright protection excludes interfaces and lock-out mechanisms if duplication or adaptation is necessary to overcome such mechanisms and establish interoperability between programs, because idea and expression merge when there is only one way the code at issue can be written in order to achieve the desired functionality.

Thus, one can further refine the general test formulated at the end of Part III of this Article: a combination of a copyrightable computer program with another constitutes a derivative work of the program if the combination (a) is sufficiently permanent, (b) contains significant and creative portions of the program as opposed to interface information or purely functional incorporation of lock-out code, (c) is creative in its own right as opposed to representing the only or most efficient technical combination solution, and (d) involves significant and creative internal changes to the other program that cannot be easily separated or distinguished from the other program. If the adapted portions or the changes caused by the combination are not creative, but merely functional in nature, the combination will not constitute a derivative work. If a software combination does not involve such qualified internal changes to the combined program(s), it may constitute a compilation or fall outside the scope of copyright law altogether.

## V. SOFTWARE COMBINATIONS UNDER COPYRIGHT LAW

In this Part, the abstract rule developed in the previous Sections is illustrated and further refined through application to a few common technical and commercial scenarios. A number of definitions and general explanations (Section A) are followed by five hypotheticals<sup>93</sup> (Sections B through F) and a summary (Section G).

---

93. The hypotheticals are based on factual scenarios discussed in the existing case law, but simplified to address only issues relevant for purposes of this Article and expanded to allow a review of surrounding questions. The assessment of the hypotheticals generally comes to the same ultimate conclusions as the courts did in the underlying cases, except, however, that it is not entirely clear whether *Dun & Bradstreet* addressed static (as assumed herein) or dynamic linking and whether the courts in *Micro Star* and *Midway* relied (or should have relied) on direct or indirect (as assumed herein) infringement. *Dun & Bradstreet Software Servs., Inc. v. Grace Consulting, Inc.*, 307 F.3d 197 (3d Cir. 2002); *Micro Star v. Formgen*, 154 F.3d 1107 (9th Cir. 1998); *Midway Mfg. v. Artic Int'l, Inc.*, 704 F.2d 1009 (7th Cir. 1983).

### A. Software Combinations Viewed from Different Technical Perspectives

Before attempting to shed some light on technical details, it is important to acknowledge the limitations of this effort. Computer science and information technology are complex and rapidly evolving disciplines, and much of the applicable terminology, details, and context are specific to particular programming architectures. Thus, this Article must address technical details in a simplified and exemplary manner, guided by the objective of illustrating the relevant legal principles. Along those lines, this Article refers to different types of software loosely as “software” or “computer programs” and mentions characteristics of applications, libraries, drivers, operating systems, and other categories of software only to the extent it is relevant for purposes of the legal analysis. Subject to these caveats, it is important to realize at the outset of any copyright analysis that (a) the installation and execution of any computer program involves countless instances of copying in excerpts, mixed up with bits of other software and dictated by data storage efficiency requirements, and (b) interoperations between programs can be achieved in many technical ways, which are dictated by program interoperability and efficiency requirements.

#### 1. *Physical and Logical View of Computer Programs in Combination*

Software users are accustomed to the logical view of programs, which is provided and managed by the Operating System’s (OS) File System (FS).<sup>94</sup> In the Microsoft Windows environment, users can see each application (for example, Adobe Reader, Microsoft Word, Outlook, and PowerPoint) separately, which are distinguishable as installed (through Windows Explorer), as copied to RAM, and as executed by the computer processor (CPU).

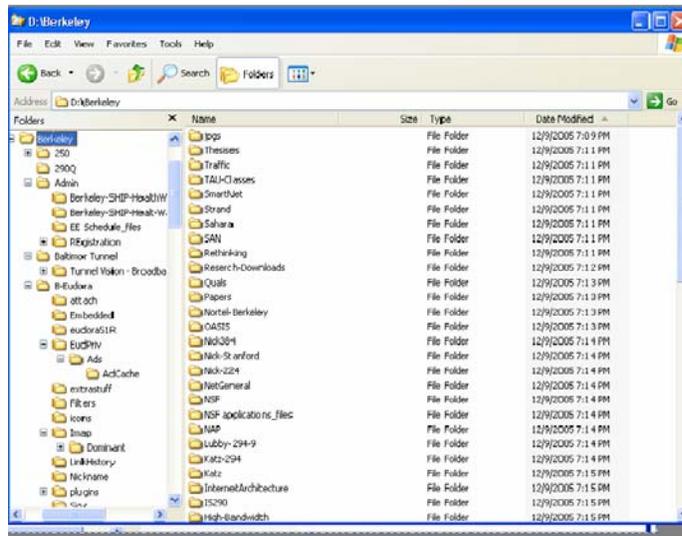
Physically, however, the bits of data comprising the various computer programs are located on various different places on the media providing the memory space, such as a memory stick, floppy disc, CD, computer hard disc, RAM, and CPU cache memory. The physical arrangement of the data bits serves purely technical storage efficiency considerations and bears little or no relationship to the logical function of the various programs. In order to save space and preserve execution speed, each of the

---

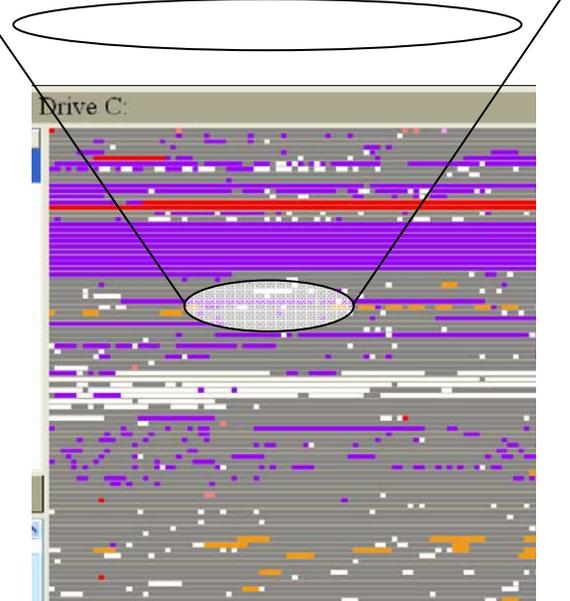
94. A file system (FS) is a method for storing, organizing, and retrieving files on a storage device. Wikipedia.org, File system, [http://en.wikipedia.org/wiki/File\\_system](http://en.wikipedia.org/wiki/File_system) (last visited Dec. 12, 2006).

different types of memory usually contains different amounts of code of a particular program.

A good example is a typical software installation from a CD to a personal computer's hard disc, followed by execution of the program from the hard disc. The full amount of the applications' electronic data (object code, data files, libraries) comes on a CD set. Based on whether the installation is full or partial, some portion of this electronic data is installed (copied) on the hard drive, where it is parceled up and placed into various spaces on the disc (entirely unrelated to the logical context of the application; for example, bits and pieces of code related to Word will sit right next to, and intermingled with, pieces of code related to Outlook and Adobe Reader).



Logical View (screenshot in MS Explorer)



Physical View (screenshot of Speed-Disc)

Figure 1: Logical View and Physical View of Software on Hard Disc

During the application execution, usually only a relatively small portion of these data are loaded from the hard drive into RAM because not all

portions of the program code are typically needed in RAM.<sup>95</sup> The CPU accesses the electronic data on the RAM and during the operation will copy some smaller amount of the data into the cache memory of the CPU (L1 or L2 cache),<sup>96</sup> where the data reside only for extremely short time periods and can be accessed and executed much faster than in RAM. During the operation of the CPU, much fewer data are copied into the CPU registers, which compose a minuscule amount of memory inside the CPU.<sup>97</sup> The closer that memory is to the CPU, where all data are processed via “current” and “no current” signals, the more accessible (faster) and smaller in size it is, and the shorter the time it is stored.

If one were to draw up a physical map to identify where the various portions of a particular program reside on a hard drive or in RAM, alone or in combination with other programs, one would find them mixed up with bits and pieces from other programs all over the storage media—like dozens of CD pieces mixed up on a pavement after falling from a twelve story building or a truck load of books and newspaper stamped into pulp in a recycling factory.<sup>98</sup>

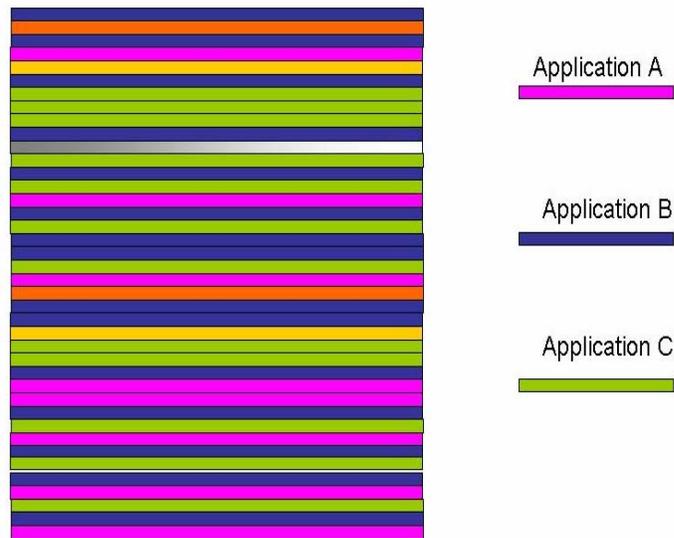
---

95. In general terms, there are differences in the amounts of space and availability relatively in a hard disc, RAM, cache, and registers. For example, Microsoft Office suite has many features that are not required in the day-do-day operation. Most Word installations do not need to copy Equations Editor and not all design templates are needed in Power Point. Thus, for execution (i.e., use), computers create “excerpts” of software, based on purely functional and efficiency considerations.

96. See Paul Geuna, *A Cache Primer*, FREESCALE SEMICONDUCTOR (Freescale Semiconductor, Inc., Austin, Texas), Oct. 2004, at 1-16, available at [http://www.freescale.com/files/32bit/doc/app\\_note/AN2663.pdf](http://www.freescale.com/files/32bit/doc/app_note/AN2663.pdf); see also Wikipedia.org, CPU cache, [http://en.wikipedia.org/wiki/CPU\\_cache](http://en.wikipedia.org/wiki/CPU_cache) (last visited Dec. 12, 2006).

97. A register is a small amount of very fast memory in a CPU providing quick access to commonly used values. See Wikipedia.org, CPU register, [http://en.wikipedia.org/wiki/CPU\\_register](http://en.wikipedia.org/wiki/CPU_register) (last visited Feb. 2, 2006).

98. See *supra* p. 24, Fig.1, Physical View.

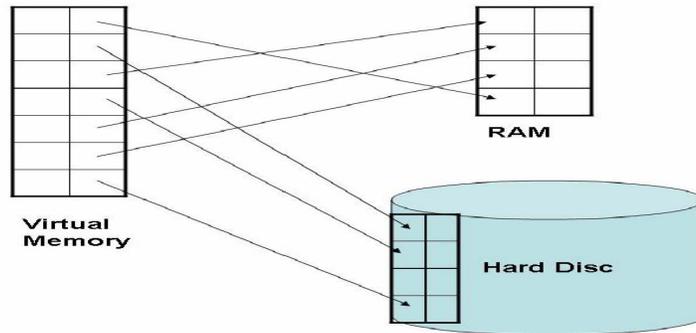


**Figure 2: The physical address space of applications is mixed on RAM.**

In order to ensure that program data storage and execution is efficient and that separate programs do not interfere on a logical level, the OS and its components allocate different memory spaces to each program on the hard drive, in RAM and cache, and translate between the human view and the computer view. Often, the amount of memory needed by applications is not available in RAM and memory pages are copied back and forth between the RAM and the hard drive. While running many applications or applications that require more memory than is available in RAM, personal computer users can see the disc light blinking and hear the disc drive working. This copying back and forth between the RAM and the hard disc is called “paging” and is part of the virtual memory<sup>99</sup> mechanism. Virtual memory addressing allows the operating system to manage the memory with non-contiguous address space in the interest of storage and execution efficiency. The cost of address space on RAM is much more expensive than the cost of address space on hard disc, while RAM access is much faster and available to the CPU.

---

99. See DAVID PATTERSON & JOHN HENNESSY, COMPUTER ORGANIZATION & DESIGN: THE HARDWARE-SOFTWARE INTERFACE 579-603 (2d ed. 1998).



**Figure 3: Virtual memory address space on RAM and hard disc**

These examples illustrate that from a purely physical perspective, computer program code is constantly copied, in excerpts, and intermingled with excerpts of other code: on a hard disc, in RAM, and in CPU cache. These activities are dictated entirely by external functionality requirements (storage and execution efficiency) and are completely unrelated to particular software applications or the creative expression embodied in them. Therefore, the dissection and combination events occurring on a physical level seem generally irrelevant for purposes of copyright law analysis. Instead, copyright analysis should focus on the logical perspective, which views programs as separate works based on how they are written by their authors and perceived by their users.

## 2. *Interoperations and Data Communications Between Programs*

Interoperability and communications between applications in a computer system are mainly realized through Inter-Process Communication (IPC). A number of different mechanisms are available, including the following: signals provide one-way asynchronous communications between application processes; a process sends a signal to a target process, and the operating system generates an event to the target application; and the target application has a signal handler to handle those signals respectively for further actions. Named pipes are a pre-defined pipeline in the computer system and provide one-way communication between application processes. Usually two applications attach to a named pipe; one application writes to the pipeline while the other application reads from the pipeline. An application can use a file to input and output data, in binary or text format.

Application communications can be realized by opening the file to read or write data. Shared memory uses a section of RAM to store data

that is shared between applications.<sup>100</sup> Shared memory provides a communication mechanism between applications by allowing applications to access the memory to read and write data.<sup>101</sup> In addition, sockets provide a data transport application program interface (API) for a network.<sup>102</sup> A socket is generally bound to the network transport protocol such as TCP/IP, including a host IP address, a protocol (TCP or UDP),<sup>103</sup> and a port number. Typically, an application opens a socket and listens on a certain port number, while another application, which can run locally or remotely, opens a socket and connects to the listening sockets. After the connection is completed, the two applications can perform a two-way communication, so they can read and write to each other. All these IPC mechanisms achieve different levels of communication and interoperability efficiency, but programs that are combined through such IPC mechanisms remain separate and distinguishable in the logical program view. The selection of the most appropriate IPC is dictated by external functionality requirements (a question of which IPC will allow efficient interoperability). Hence, programs combined like this would generally not be considered part of a larger derivative work due to a lack of combination creativity and significant internal changes, as the following hypothetical situations will further illustrate.

## **B. Package with Programs of Isolated Functionality**

### *1. Basic Hypothetical*

A reseller buys application programs from two software manufacturers. The applications have isolated, independent functionality (e.g., two alternative photo editors) that do not interact with each other. Each manufacturer ships its respective program on pre-packaged CDs to the reseller. The reseller takes one pre-packaged CD from each manufacturer and places them in a cardboard box to sell them in combination. Consumers purchase the two-CD packages and usually install only one of the two alternative programs on a personal computer.

---

100. *Id.* at 727-28.

101. *Id.*

102. LARRY PETERSON & BRUCE S. DAVIE, *COMPUTER NETWORKS: A SYSTEM APPROACH* 332-35 (1st ed. 1996).

103. *Id.*

## 2. *Assessment*

The combination of the programs in a sales package is too loose and detachable,<sup>104</sup> and not sufficiently creative, to qualify as a protectable work for purposes of copyright law. Thus, due to its lack of fixation and originality, the combination would also not qualify as a derivative work, collective work, or compilation. Further, neither the reseller nor any consumers make any changes to the programs and the consumers do not typically install the programs on the same hardware or execute programs in the same RAM. Thus, the programs do not modify each other even temporarily. This is another reason why neither the reseller nor the consumers create derivative works by distributing, installing, or executing the programs in combination.

## 3. *Variations*

The assessment should not change if the reseller receives golden masters for duplication from the two software manufacturers, and then copies two programs on one CD or the hard drive of a personal computer. The fixation requirement for a compilation would be met under such circumstances, but the selection could still not qualify as creative enough in nature to comprise a copyrighted compilation. Physically, the programs' individual bits of data would stay fairly separate on the CD where data are organized sequentially, but would be "mixed up" on the hard drive of the personal computer. The sequential arrangement of data bits on the CD and the "mix up" of data bits on the PC hard drive, at the physical level, do not involve even a spark of creativity with respect to the software itself, because the storage arrangements are dictated entirely by technical storage efficiency and functionality requirements. Therefore, the physical arrangement lacks creative internal changes required to qualify as a derivative work. From a logical perspective, the two programs remain entirely separate and detachable on either disc (CD and hard drive) and will be perceived and interpreted by any computer user as separate, unrelated programs.

---

104. See *Bucklew v. Hawkins*, 329 F.3d 923, 930 (7th Cir. 2003). The court in *Bucklew* noted that unauthorized combination of copyrighted works might be allowed where:

[the] original expression added by the unauthorized preparer of a derivative work is clearly detachable from the original work itself, so that no confusion, or disruption of the copyright owner's plans for the exploitation of his work, would be created by allowing the unauthorized preparer to copyright his original expression.

*Id.*

### C. Modification of Screen Output Through Software Combination

#### 1. Hypothetical<sup>105</sup>

Company A makes video games that consist of three programs, which are delivered as separate object code files on a CD: a game engine program, a source art library, and a MAP file. Consumers who purchase a game CD receive an express license to install and play the game uploaded into the RAM of a computer, but not a right to create derivative works. Printed on a piece of paper, the object code of each of the three programs would appear as thousands of lines of zeros and ones. The source code would appear as a set of instructions to a machine to create a certain screen output, which audio-visualizes the story of a fighter who is to some extent controlled by the game player and who has to confront various enemies and challenges in numerous fantasy worlds depicted by different background environments and game levels.

In the RAM of a computer, the three programs' individual instructions are stored in different address spaces where they are accessed by the CPU for execution. During execution, the programs automatically call on each other through links embodied in the programs. The game engine program instructs the computer when to read data, save and load games, play sounds, and project images onto the screen. In order to create the audio-visual display for a particular game level, the game engine calls the MAP file, which contains a series of instructions that tell the game engine (and, through it, the computer) what artwork from the source art library (e.g., a mountain, city, or tree) to display where on the screen. For instance, the MAP file might indicate that the scuba gear should be at the bottom of the screen. The game engine then goes to the source art library, finds the image of the scuba gear, and displays it at the screen location specified by the MAP file. The MAP file describes the level environment in painstaking detail, but it does not actually contain any of the art itself; everything that appears on the screen actually comes from the source art library program. The creation of the game's audiovisual display functions similar as a paint-by-numbers kit: The MAP file tells the engine program to put blue paint in section number 565, but it does not contain any blue paint itself; the blue paint comes from a palette, which is the low-tech analog of the source art library, while the painter plays the role of the game engine.<sup>106</sup>

---

105. The details of this hypothetical are adapted from *Micro Star v. Formgen*, 154 F.3d 1107 (9th Cir. 1998), *Lewis Galoob Toys, Inc. v. Nintendo of America, Inc.*, 964 F.2d 965 (9th Cir. 1992), and *Midway Manufacturing v. Artic International, Inc.*, 704 F.2d 1009 (7th Cir. 1983).

106. See *Micro Star*, 154 F.3d at 1110.

Company B enters the market and sells add-on products that enable consumers to manipulate the screen output of Company A's games by speeding up the game sequence,<sup>107</sup> changing existing parameters (speed or strength of game characters),<sup>108</sup> or correcting errors.<sup>109</sup> Company C creates new MAP files that contain instructions for alternative combinations of game level environments (e.g., with new creative arrangements of backgrounds and threat-posing characters).<sup>110</sup> The underlying code of the add-on programs does not bear any resemblance to Company A's code at the source or object code level.

Consumers purchase and install copies of the add-on programs developed by Companies B and C on the same computer as Company A's games and execute them simultaneously. Each of the add-on programs is executed in a different memory address space of the computer's RAM. They provide additional instructions to the computer, which result in modified screen outputs during the execution of the programs. Depending on how the game is played, the computer will execute the programs in different sequences and variations. When consumers finish playing, they close the programs and no permanent changes remain in the copy of Company A's game that is installed on the hard disc.

## 2. *Assessment*

The object and source code of Company A's game engine program and MAP file each constitute a literary work; assuming a minimal creative spark, these programs are sufficiently original to enjoy protection against literal copying.<sup>111</sup> The source art library program, on the other hand, constitutes a collection of instructions, arranged to optimize access by the MAP file and game engine program, to create the various background items. While individual items may be copyrightable, their arrangement seems purely functional and thus not sufficiently creative to constitute a collective work or other compilation.<sup>112</sup>

When executed in combination, the three programs create game sequences with a reasonably consistent plot, set of characters, and back-

---

107. In *Midway Manufacturing v. Artic International, Inc.*, this result was achieved through a circuit board as opposed to software, which is not relevant for purposes of copyright law, because 17 U.S.C. § 106(2) prohibits any adaptations regardless of the technical means of creation. *Midway*, 704 F.2d at 1009-10.

108. *Lewis Galoob Toys*, 964 F.2d at 967.

109. *Lewis Galoob Toys* addresses add-on spell-check programs in dictum. *Id.* at 969.

110. *Micro Star*, 154 F.3d at 1110, 1112.

111. *See supra* Section III.A.

112. *See supra* Section III.C.

ground variations.<sup>113</sup> This screen output is protectable as an audiovisual work.<sup>114</sup>

After the development process,<sup>115</sup> Companies B and C distribute their add-on programs independently from Company A's games. The add-on programs do not contain any expression taken from Company A's copyrighted games at the code or screen output level.<sup>116</sup> Since derivative works and compilations by definition require content from other works, the add-on products themselves cannot qualify as either.<sup>117</sup> Nevertheless, since B and C make, advertise, and distribute their add-on programs for the sole intended purpose that users combine the add-on programs with A's games, B and C would be contributorily liable for their customers' actions.<sup>118</sup>

Consumers execute the add-on programs simultaneously with the game programs to modify the game's screen output. These combinations affect the games at two levels: (i) the screen output, which is protected as

---

113. See generally *Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852 (2d Cir. 1982).

114. 17 U.S.C. § 102(a); *Stern Elecs.*, 669 F.2d at 855-56.

115. It seems likely—yet outside the scope of the hypothetical—that in the process of developing its products, Company B may have modified the copyrighted software of Company A products. However, such preparatory activities are subject to a number of different considerations, including the Fair Use doctrine. See, e.g., 17 U.S.C. §§ 107, 1201(f) (2000); *Sony Computer Entm't, Inc. v. Connectix Co.*, 203 F.3d 596, 602 (9th Cir. 2000), cert. denied, 531 U.S. 871 (2000); *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1520 (9th Cir. 1992).

116. In both *Micro Star v. Formgen*, 154 F.3d 1107 (9th Cir. 1998), and *Midway Manufacturing v. Artic International, Inc.*, 704 F.2d 1009 (7th Cir. 1983), the court did not explain whether add-on products even had a user-interface or any independent screen output.

117. See *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 969 (9th Cir. 1992) (distinguishing *Midway* by pointing out that Galoob's add-on module for the Nintendo system, the Game Genie, "does not physically incorporate a portion of the copyrighted work [such as the Nintendo console or games]."). But see Mitchell L. Stoltz, Note, *The Penguin Paradox: How the Scope of Derivative Works in Copyright Affects the Effectiveness of the GNU GPL*, 85 B.U. L. REV. 1439, 1458 (2005) [hereinafter Stoltz, *Penguin Paradox*]. In both *Micro Star*, 154 F.3d 1107, and *Midway*, 704 F.2d 1009, the courts seemed less concerned with this requirement and focused on the respective "story lines" being retold by the makers of the add-on. However, the facts in either case would also seem to support a finding of infringement based on a contributory liability theory, and in that context, the derivative work could have been the audio-visual work created by the end user through the combination of the add-on program with the original game. Thus, it does not seem that these cases create authority for a general proposition that an add-on program on its own—outside the context of the intended combination work—can constitute a derivative work or compilation.

118. See *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005); *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984); *A&M Records, Inc. v. Napster, Inc.*, 284 F.3d 1091 (9th Cir. 2002).

an audiovisual work and (ii) the underlying code, which is protected as a literary work. The combination of each add-on program with the game is illegal if an unauthorized derivative work is created at either level. Whether the combination creates a derivative work depends on whether the combination (a) is sufficiently permanent, (b) contains significant and creative portions of the program(s), (c) is creative in its own right, and (d) involves significant and creative internal changes to the other program that cannot be easily separated or distinguished from the other program.

a) Screen-Output Level

i) Permanence

By deploying Company B's add-on products, the consumers do not permanently modify Company A's game screen outputs. At the end of the combined use of Company A's and Company B's products, Company A's software is back to "normal"—if the users execute the games next time without the add-on programs, the screen output will be exactly as designed by Company A. Thus, the consumers do not create permanent derivative works through the combination with the add-on programs. While the games and add-on programs are in RAM, however, users can manipulate the speed and other aspects of the games through the add-on products made by Company B, and the MAP files supplied by Company C provide new game level background environments. RAM copies are generally sufficiently permanent fixations for purposes of copyright law.<sup>119</sup> Yet, the manipulations made by players through the add-ons supplied by Company B are fleeting in nature, and will differ during each game session. Therefore, they lack the minimum permanence required to qualify as derivative works.<sup>120</sup> The MAP files supplied by Company C, on the other hand, are permanent in nature and always create the same art combinations in conjunction with the source art library files. Thus, the MAP files create combinations that are sufficiently permanent to constitute derivative works or compilations.<sup>121</sup>

---

119. *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 518 (9th Cir. 1993) ("After reviewing the record, we find no specific facts (and Peak points to none) which indicate that the copy created in the RAM is not fixed.")

120. According to *Lewis Galoob Toys*, 964 F.2d at 967-68, fixation is only relevant for the question whether derivative works are protectable, not whether they infringe; however, in order to infringe, derivative works have to exist in a "concrete or permanent" form. *See also Micro Star*, 154 F.3d at 1110-11; 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.09[A] (requiring "permanence").

121. *Cf. Stern Elecs., Inc. v. Kaufman*, 669 F.2d 852, 856 (2d Cir. 1982).

ii) Substantial and Creative Portions

Upon execution of the add-on products, the screen outputs still comprise only artwork created by Company A because all background items originate from Company A's source art library program. Some of the individual art items are most likely sufficiently creative to qualify for copyright protection. Therefore, the screen outputs created by the combination of the game programs and any of the add-on programs contain substantial and creative portions from Company A's copyrighted programs.

iii) Combination Creativity

The next question, then, is whether the changes to the screen outputs created by the combination of the games and the add-on programs are merely functional or sufficiently creative. Simple changes to the speed of program execution,<sup>122</sup> game character parameters,<sup>123</sup> and error corrections<sup>124</sup> caused by Company B's add-on programs will typically not meet even minimal creativity requirements, whereas the detailed rearrangement of background motives from the source art library program caused by Company C's program will typically suffice.<sup>125</sup>

iv) Significant Internal Changes

Whether or not the combination causes internal changes to existing copyrighted works determines whether the combination can qualify as a derivative work (as opposed to a compilation or other arrangement). By rearranging the items or sequence of events on the screen outputs designed by Company A, all add-on products apply internal changes to the audiovisual work(s) created by Company A. In the case of Company C's MAP files, the changes seem significant, whereas the mere manipulation of the playing speed caused by Company B's add-ons will likely not suffice.<sup>126</sup>

---

122. *Lewis Galoob Toys*, 964 F.2d at 969. In 1983, the Seventh Circuit came to the same conclusion regarding a lack of creativity, yet ultimately found for infringement due to a perceived need to honor the equities involved in the case outside the scope of the Copyright Act: "[T]he amount by which the language of Section 101 must be stretched to accommodate speeded-up video games is, we believe, within the limits within which Congress wanted the new Act to operate." *Midway*, 704 F.2d at 1014. Such a pure economic analysis does not seem justifiable anymore post-*Feist*. See *Feist v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991).

123. *Lewis Galoob Toys*, 964 F.2d at 967.

124. *Id.* at 969 (addressing add-on spell-check programs in dictum).

125. *Micro Star*, 154 F.3d at 1111-12.

126. For instance, a DJ would not create a derivative work by playing at 45 RPMs a musical record designated for 33 RPM speed. See *Midway*, 704 F.2d at 1013. Similarly, small add-on toolbars to internet browsers or additional menu features for other programs

v) Conclusion

When executed in combination with the games, Company C's add-on program creates a derivative work of Company A's screen outputs. Combinations with Company B's add-on programs, on the other hand, lack sufficient permanence, significant internal changes, and creativity to qualify as derivative works. Therefore, consumers, and contributorily, Company C, would infringe on Company A's adaptation right to the screen outputs by deploying Company C's add-on program, whereas Company B does not infringe on Company A's adaptation right.

b) Underlying Code

Given how different the software at issue looks at the code and screen output levels, it is not surprising that the characteristics of the respective combinations also show some distinctions:

i) Permanence

The combination of Company A's game engine code and the add-ons made by Companies B and C depends much on user interaction and does not meet the permanence requirement at the code level either. The same is true for any combinations that occur in CPU cache, given the fleeting nature of these extremely short-lived copies in the execution process.<sup>127</sup> The interaction between Company C's MAP file and the source art library, on the other hand, occurs in a similar way whenever the two programs are executed in combination, which indicates some form of permanency. Yet, given the fact that the RAM copies and execution sequences are logically separated in RAM (in different address space) and on the hard drive of the computer, the combination of the executables appears too detached<sup>128</sup> and separate to qualify as a permanent combination for purposes of constituting a derivative work or compilation.

ii) Substantial and Creative Portions

If one were to focus on the physical view and consider the functionally segmented aggregation of programs in RAM a permanent combination—contrary to the preceding assessment—the substantiality requirement would be met. During execution, the RAM contains large portions of Company A's code, which contains at least some creative portions.

---

may not qualify as “significant” in the generally functional user interface environment. *See generally* Lotus Dev. Corp. v. Borland Int'l, 49 F.3d 807, 815-16 (1st Cir. 1995).

127. *See* CoStar v. Loopnet, 373 F.3d 544, 550-52 (4th Cir. 2004) (addressing quantitative and qualitative thresholds regarding the “fixation” requirement).

128. *See* Bucklew v. Hawkins, 329 F.3d 923, 930 (7th Cir. 2003).

### iii) Combination Creativity

One must examine the question of whether the changes to the RAM copy created by the combination of the games and the add-on programs are merely functional or sufficiently creative separately from the question of whether the changes to the screen output are functional or creative.<sup>129</sup> In this context, the idea-expression merger phenomenon comes into play: If there is only one efficient mechanism to manipulate game sequence, change game character parameters, correct errors, or compile level background artwork, then the software implementation of this mechanism cannot qualify as creative for copyright purposes because the functional requirements do not leave any room for creativity.<sup>130</sup>

### iv) Significant Internal Changes

Whether or not the combination causes significant and creative internal changes to existing copyrighted works determines whether a combination that otherwise qualifies as a copyrightable work will qualify as a derivative work or a compilation. The screen outputs of Company A's games constitute works in their own right, and by rearranging the items or sequence of events on the screen outputs, all add-on products apply internal changes to the work(s) created by Company A.<sup>131</sup> On the code level, however, the situation looks quite different.<sup>132</sup> From a logical perspective, Company A's underlying programs and subprograms remain separate literary works, consisting of instructions to a computer which are formed and arranged for a purely functional reason, namely, to cause the desired changes on the screen output level. Company C's MAP file does not make any internal changes to the source art library program; it merely calls individual items from the library into RAM in a different manner than Company A's MAP file would have. Therefore, at the code level, Company C's MAP file lacks another requirement to qualify as a derivative literary work of Company A's game program code: significant and creative internal changes. Consequently, at the code level, Company C's program constitutes a new (non-derivative) work.

---

129. The latter question is examined *supra* Section V.C.2.a.

130. *See supra* Section IV.B.

131. *See supra* Section V.A.

132. In *Micro Star v. Formgen*, the court clarified that the infringed work at issue was the audiovisual screen output and story line; thus, the fact that the underlying code was not changed did not matter for the outcome of the case. 154 F.3d 1107, 1112 (9th Cir. 1998). The courts in *Sega Enterprises, Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), and *Sony Computer Entertainment, Inc. v. Connectix Co.*, 203 F.3d 596 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000), did not even discuss a potential adaptation right infringement through code combinations during the execution phase.

v) Conclusion

At the code level, combinations with the add-on programs do not qualify as derivative works because the combination lacks permanency and typically also combination creativity (the means to achieve interoperability are dictated by externalities). Even if any combination creativity was present, the add-on programs could only qualify as compilations, which would not affect Company A's adaptation right under Section 106(2) of the Copyright Act<sup>133</sup>, due to the lack of internal changes.

**D. Add-on Product with Static or Dynamic Links, but with No Impact on Screen Output.**

1. *Hypothetical*<sup>134</sup>

Company A owns and markets a payroll management software package that consists of numerous programs, including database software in which employers can store human resources data and an application program ("AW-2") that allows employers to create W-2 year-end reports for income tax purposes. Customers receive the program package under a license that permits installation and execution and prohibits the creation of derivative works.

Company B develops a substitute W-2 application program ("BW-2") and markets it to Company A's customers. BW-2 comes with a separate user interface and neither its code, its user interface, nor any screen output of BW-2 contains any elements of AW-2. However, BW-2 must extract data from the human resources database in order to perform its functions. Thus, BW-2 has to interact with Company A's database software. For example, if a payroll administrator enters "Create W-2 report for John Doe," BW-2 sends commands to Company A's database software to find the information required for the report. In order to ensure Company A's program understands the requests from BW-2, Company B must use certain commands and interface specifications defined by Company A. Also, BW-2 calls on other functions offered by sub-programs or libraries in Company A's overall payroll management software package, such as scripts, macros, mathematical calculations, and currency conversion. Once BW-2 has gathered all the required information, it creates John Doe's W-2 report, which the payroll administrator can then save and print.

---

133. 17 U.S.C. § 106(2) (2000).

134. These facts are adapted from *Dun & Bradstreet Software Services, Inc. v. Grace Consulting, Inc.*, 307 F.3d 197 (3d Cir. 2002), but are subject to clarifications and intentional modifications.

Company B can achieve interoperability through static or dynamic links. If Company B follows the static linking approach,<sup>135</sup> it permanently copies and pastes necessary commands and other code lines from Company A's programs into the BW-2 code; this action would usually occur at the source code level prior to compilation into object code. Thus, on the CD delivered by Company B, the copies of the BW-2 code would actually contain copied portions of code written by Company A developers. In the dynamic linking option, on the other hand,<sup>136</sup> Company B programs call functions into BW-2 that instruct the computer to obtain certain data or other input by executing a Company A program (in a separate address space of the RAM). Subsequently, after the Company A program produces

---

135. For an explanation of static linking in the context of a library program, see Wikipedia.org, Static Library, [http://en.wikipedia.org/wiki/Static\\_Library](http://en.wikipedia.org/wiki/Static_Library) (last visited Jan. 21, 2006). The Static Library provides that:

[i]n computer science, a static library, also referred to as a statically linked library, is a computer library in which links are resolved at compile-time by a linker. Static libraries may be merged with other libraries and executables to form a single object file, or they may be loaded at run-time into the address space of the linking executable or library, at a static memory offset determined at link-time.

*Id.*

136. For an explanation of dynamic linking in the context of a library program, see Wikipedia.org, Dynamic Linking, [http://en.wikipedia.org/wiki/Dynamic\\_linking](http://en.wikipedia.org/wiki/Dynamic_linking) (last visited Jan. 21, 2006). An explication of dynamic linking has been provided as the following:

Dynamic linking means that the data in a library is not copied into a new executable or library at compile time, but remains in a separate file on disk. Only a minimal amount of work is done at compile time by the linker—it only records what libraries the executable needs and the index names or numbers. The majority of the work of linking is done at the time the application is loaded (load time) or during the execution of the process (runtime). The necessary linking code, called a loader, is actually part of the underlying operating system. At the appropriate time the loader finds the relevant libraries on disk and adds the relevant data from the libraries to the process's memory space. Some operating systems can only link in a library at load time, before the process starts executing; others may be able to wait until after the process has started to execute and link in the library just when it is actually referenced (i.e., during runtime). The latter is often called "delay loading." In either case, the library is called a dynamically linked library. This term is sometimes shortened to "dynamic link library" or DLL, but this last initialism is most common in Microsoft Windows environments where dynamic libraries use the filename extension .dll.

*Id.*

certain data results, Company B's call functions instruct the computer to resume the execution of BW-2 with the input so obtained.<sup>137</sup>

## 2. Assessment

BW-2 does not impact the screen output or user interface of Company A's programs, so it does not create a derivative pictorial or audiovisual work. At the actual code level, however, software combinations occur both in the static and dynamic linking scenario. Such combination would qualify as a derivative work of Company A's programs if the combination (i) is sufficiently permanent, (ii) contains significant and creative portions of Company A's programs, (iii) is creative in its own right, and (iv) involves significant and creative internal changes to Company A's program(s) that one cannot easily separate or distinguish.

### a) Permanence

In the static link scenario, the combination is permanent because the copied sections of Company A's code appear within the code lines of Company B's code.<sup>138</sup> The combination created by dynamic links, on the other hand, does not appear to be permanent for two reasons. First, the dynamic links are activated by a user requesting a particular operation from the BW-2 program, and therefore the precise character of the combination will be different each time the programs are executed. Second, the programs appear in separate address space of RAM and are connected only through sequential calls, so they do not permanently become part of a larger work.

### b) Substantial and Creative Portions

In the static link scenario, BW-2 incorporates significant amounts of code from Company A's program(s) into its own code, which will typically result in the adoption of substantial and creative portions of copyrighted code owned by Company A. However, in the dynamic link scenario, the code lines whose functionality is required are not incorporated,

---

137. See *Dun & Bradstreet*, 307 F.3d at 204-05, 208, 213 (describing dynamic linking, but also mentioning that Grace's substitute "W-2 program actually consists of 62% Geac code," and that "Grace admitted that the installation, testing, compiling and link editing of its W-2 programs required copying Geac's software and link editing the Geac code," which seems to describe static linking). Grace asserted that its W-2 program accessed customer's data, not Geac's copyrighted code. *Id.* at 212; see also 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.08[D][2]; David McGowan, *Legal Aspects of Free and Open Source Software* 23 (2005), <http://www.cogsci.ucsd.edu/~rik/courses/readings/McGowanD-OpenSource.pdf> [hereinafter McGowan].

138. Company A's duplication right is infringed as well.

but rather executed by the computer separately and in sequence.<sup>139</sup> In order to reach the desired functions in the Company A programs, BW-2 may have to copy some necessary interface code lines, which are typically neither substantial nor creative.<sup>140</sup>

c) Combination Creativity

The creativity aspect does not seem to depend on the method of linking (static or dynamic). Either way, the combination serves the sole purpose of extracting data or functions from one program and sharing them with another. If Company B limits the combination to the absolute minimum necessary for functionality purposes, the combination itself should generally be classified as purely functional and, therefore, not sufficiently creative to constitute a derivative work or a compilation.<sup>141</sup>

d) Significant Internal Changes

Assuming all other requirements are met, a classification as a derivative work or compilation can depend on the linking technique. In the static link solution alternative, Company B permanently copies and pastes lines of code from Company A's copyrighted programs into the context of Company B's own code. If such code lines are taken out of a creative context (as opposed to a merely functional aggregation), this would involve internal changes to a copyrighted work owned by Company A. If, however, Company B incorporates entire subprograms, which are works in their own right, the combination product may lack significant internal changes to Company A's works. Also, in the dynamic link solution alter-

---

139. The computer executes BW-2 until it calls for a function that is only available from a Company A program. The computer then executes the Company A program, until the requested result is available and subsequently uses this result to continue the execution of BW-2. The execution of the linked code in RAM is generally permitted under the applicable license agreements. See 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.08[D][2].

140. See *Computer Assocs. Int'l v. Altai, Inc.*, 982 F.2d 693, 704-05 (2d Cir. 1992).

141. See *Sony Computer Entm't, Inc. v. Connectix Co.*, 203 F.3d 596, 602-03 (9th Cir. 2000), *cert. denied*, 531 U.S. 871 (2000); *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1522-23 (9th Cir. 1992). The practical relevance of this assessment is greater for dynamically linked programs as compared to statically linked programs, because static linking involves, by definition, literal or non-literal copying that is typically not covered by any license—whereas the RAM copies involved in dynamic linking are usually permitted in standard end user license agreements (EULAs). See *infra* Part VI. However, U.S. courts have generally been sympathetic towards defendants who copied minor amounts of code in the interest of achieving interoperability (as opposed to substitution). See, e.g., *Lexmark Int'l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522, 544 (6th Cir. 2004); *Sega*, 977 F.2d at 1518; *Vault Corp. v. Quaid Software, Ltd.*, 847 F.2d 255, 267-68 (5th Cir. 1988).

native, Company A's code remains generally unchanged, viewed from a logical perspective, even in RAM where the programs are kept in different address spaces. Thus, the requirement of significant internal changes is typically not met by a software combination created through dynamic linking or by statically linking detachable subprograms.

e) Summary

Software combinations that pursue functional goals and do not involve changes to, or combinations of, screen outputs or user interfaces will typically not meet the creativity requirement to qualify as derivative works. Additionally, dynamic link solutions will typically lack the permanence, substantial and creative elements, and internal changes elements of derivative works.<sup>142</sup> Static link solutions that do not qualify as derivative works nevertheless typically infringe the duplication right to the linked program.

## E. Web Linking

### 1. Hypothetical

X makes various articles and other information on currencies as well as a number of currency conversion programs (with simple user interfaces) available through her home page. Y has an entirely separate website on which he recommends various websites with hyperlinks, including X's homepage. If a website visitor clicks on the hyperlink to X's page, Y's website disappears and X's website appears in the visitor's browser. Z programs her own website with "in-line linking" references to X's currency conversion programs; visitors of Z's website see X's currency conversion program interfaces as part of Z's website.<sup>143</sup>

### 2. Assessment of Y's Hyperlinks

Y creates a combination between his and other websites through hyperlinks. The permanence of the links collection resembles more the dy-

---

142. An example of dynamic linking in the "off-line" world would be a hornbook supplement for a law school textbook, which is designed to slip into the pocket part of the law textbook. The hornbook does not use any creative content from the textbook, but instead consists of a series of "links" which point the user to a specific page (e.g., the hornbook might say "See Page 15 for Liquidated Damages" and list a few new cases in the field.) The hornbook also makes no internal changes to the original textbook. It should not be considered a derivative work.

143. For a more detailed description of the technological background of in-line linking, see *Kelly v. Arriba Soft*, 336 F.3d 811, 816 (9th Cir. 2003). On liability for in-line linking in general, see Allison Roarty, Note, *Link Liability: The Argument for Inline Links and Frames as Infringements of the Copyright Display Right*, 68 FORDHAM L. REV. 1011, 1045 (1999).

dynamic link solution discussed above in Section V.D than the MAP file scenario discussed in Section V.C because user interaction is required to realize any combination. Y's hyperlinks do not adopt "significant portions" of X's website, since the linked pages are displayed separately and after Y's home page is processed. Finally, Y's hyperlinks do not cause any internal changes to X's website or the website of others. Thus, Y's link collection does not constitute a derivative work in combination with the linked websites.

### 3. *Assessment of Z's Inlinks*

Any visitor of Z's website sees the same combination of Z's website and elements from X's website. This combination takes X's user interfaces out of the context of X's home page and combines them with Z's own website layout. Thus, the "permanency" and "substantial and creative portions" requirements seem generally fulfilled. If the combination is sufficiently creative, it would qualify as a derivative work or compilation; if not, it would qualify as a non-literal copy. Whether Z's in-line links create a derivative work of X's website, on the one hand, or a compilation or non-copyrightable, functional arrangement of individual works displayed on X's websites, on the other hand, depends on how creatively the various items are arranged on the respective websites. While many websites look very similar these days and are designed by functional considerations without much originality or creativity,<sup>144</sup> it seems more likely that Z's website lacks the elements of a derivative work. In any event, the combination of X's and Z's website elements is not directly realized by Z, but rather by the website visitors. When they visit Z's website, their browsers create the derivative works, compilations, or non-literal copies of X's website. Z could only be contributorily responsible for this result if the combination created by the website visitors' browsers constitutes a derivative work.<sup>145</sup>

---

144. *But see* MyWebGrocer, LLC v. Hometown Info, Inc., 375 F.3d 190, 193-94 (2d Cir. 2004).

145. If X's website is generally available to the public without password or other restrictions, the website visitors should normally be deemed to receive an implied license to view X's website and the individual elements thereon. Thus, if Z's website does not constitute a derivative work of X's website, X could not succeed in a suit against Z under copyright law, but X may have other cause of actions, for example, under breach of website use contract terms or trademark law. *See, e.g.*, Ticketmaster Corp. v. Tickets.com, Inc., No. 03-55641, 2005 U.S. App. LEXIS 6227, at \*2 (9th Cir. Apr. 11, 2005).

## F. Interface Modifications

For completeness sake, one point should be mentioned again at this junction: If programs are not already designed to be interoperable, their combination will often require changes to the program interfaces and possibly other code segments. Such changes typically involve copying and modification of interfaces, which are usually permissible so long as the activities remain limited to that which is absolutely necessary for purposes of achieving interoperability.<sup>146</sup>

## G. Summary

Applying the refined test developed in the previous Part to the examples in this Part confirms results that courts have reached in similar cases<sup>147</sup>, that is, users do not normally create a derivative work by executing two separate programs, unless one program significantly and creatively changes the creative screen output of another program. Developers can avoid or at least substantially reduce the risk of becoming liable for infringing duplication or adaptation rights by creating add-on programs using only dynamic linking techniques as opposed to static links.<sup>148</sup>

Of course, every individual case is different and technology is constantly evolving. Yet, given that the above examples address the main scenarios that have been subject to court decisions or discussion in the GPL-

---

146. This is based on one of a number of threshold requirements and defenses regarding infringement, including the idea/expression merger and the Fair Use doctrine. *See generally* Chamberlain Group, Inc. v. Skylink Techs., Inc., 381 F.3d 1178 (Fed. Cir. 2004), *cert. denied*, 544 U.S. 923 (2005); *Lexmark*, 387 F.3d 522; *Sony*, 464 U.S. 417; *Computer Assocs. Int'l v. Altai, Inc.*, 982 F.2d 693 (2d Cir. 1992); *Sega*, 977 F.2d 1510.

147. The only exception may be *Dun & Bradstreet Software Services, Inc. v. Grace Consulting, Inc.*, if the software at issue in that case did contain dynamic as opposed to static links, which is not entirely clear from the decision. 307 F.3d 197 (3d Cir. 2002). If so, the case would be at odds with prior cases. *See, e.g.*, *Lotus Dev. Corp. v. Borland Int'l*, 49 F.3d 807 (1st Cir. 1995) (recommending add-on products as an alternative to substitute products with architectural similarities); *Lewis Galoob Toys, Inc. v. Nintendo of Am., Inc.*, 964 F.2d 965, 967 (9th Cir. 1992) (insisting that derivative works must contain creative expression copied from the adapted original). Other cases have dealt with a situation in which the add-on products contained creative content that adapted a creative storyline from another, copyrighted product. *Worlds of Wonder, Inc. v. Vector Intercontinental, Inc.*, 653 F. Supp. 135 (N.D. Ohio 1986); *Worlds of Wonder, Inc. v. Veritel Learning Sys.*, 658 F. Supp. 351 (N.D. Tex. 1986).

148. As noted above, even static linking may not affect the copyright owner's adaptation right to the linked program. For example, the linking program may constitute merely a compilation if it includes copies of entire programs without making internal changes, or it could even constitute an entirely unprotectable combination for lack of combination creativity. Nevertheless, static links will often infringe duplication rights and, therefore, are generally more problematic than dynamic links.

context, it seems that the question raised in the title of this article can tentatively be answered as follows: software combinations do not normally constitute derivative works, or dangerous liaisons—at least not according to the U.S. Copyright Act. The following two Parts of this Article examine whether a review of commercial and open source licensing practices mandates a different answer.

## **VI. SOFTWARE COMBINATIONS UNDER COMMERCIAL SOFTWARE LICENSES**

A copyright owner does not have a statutory right under copyright law to prohibit software combinations, except to the extent that a combination qualifies as a derivative work of the copyright owner's software. Nonetheless, a copyright owner can try to prohibit combinations contractually, whether or not they constitute derivative works. Part VI of this Article examines this potential danger to software developers by briefly reviewing commercial software licensing practices and then discussing laws that limit a copyright owner's ability to contractually prohibit software combinations.

For several reasons discussed in more detail below, copyright owners find contractual or quasi-contractual measures less effective than a statutory prohibition sanctioned by the Copyright Act. First, contract or license clauses bind only licensees and not the public at large. Second, licensees may simply refuse to agree to such clauses. Third, even if agreed upon, the clauses may be invalid as a matter of statutory law under the doctrine of copyright misuse, competition law, or laws against unfair contract terms as an unlawful attempt to contractually expand exclusivity rights granted and intentionally limited by statute.

Despite such weaknesses, contractual or quasi-contractual clauses can have similar effects as statutory copyright law itself if they appear uniformly in mass-market end user license agreements (EULAs). Just imagine if a dominant software manufacturer, such as Microsoft, included a clause in the EULAs of its Windows and MS Office Suite, which stated that the "licensee agrees that he or she will not install or execute Netscape Navigator, RealNetworks RealPlayer or any other non-Microsoft program on any PC on which Microsoft products are used or installed." If such a clause appeared in a shrink-wrap or click-through license agreement, many—if not most—consumers would probably accept it without hesita-

tion.<sup>149</sup> If such a clause were also legally valid and enforceable, it could have serious implications for software interoperability.

Given the general freedom of contract and the variety of commercial licensing practices and market situations, the following discussion can only be exemplary in nature and aims to flag potential legal issues. At the same time, it is worth broadening the jurisdictional scope of the discussion given the global reach of most software licenses. Copyright laws are territorial, but fairly harmonized among the many member states of the Berne Convention.<sup>150</sup> Contract and competition laws, on the other hand, vary substantially from jurisdiction to jurisdiction. Nevertheless, many software license agreements apply to software use globally and many licensors port their licensing models to other jurisdictions without regard to local laws. German law seems well-suited for comparison because its civil law tradition often differs substantially from the U.S. German courts also have decided many software-related cases that were similar factually to cases that have been brought in the United States.<sup>151</sup>

#### A. First Sale Doctrine

First, it is helpful to remember why copyright owners are even able to require end user license agreements in the software field. Copyright owners of other types of works, such as music records, movie DVDs, or books, are usually prevented from soliciting such agreements by the First Sale Doctrine. According to this doctrine, the copyright owner's exclusive right to prohibit (or permit) further distribution of a particular copy becomes

---

149. See, e.g., Lydia P. Loren, *Slaying the Leather—Winged Demons in the Night: Reforming Copyright Owner Contracting with Clickwrap Misuse*, 30 OHIO N.U. L. REV. 495, 496 n.6 (2004). Anecdotal evidence suggests that few, if any, consumers ever read such clauses, but instead merely select “I agree” and continue with little hesitation. Based on the evidence discovered in connection with recent antitrust lawsuits against Microsoft, it seems quite possible that OEMs, distributors, and resellers would also have tolerated such clauses in EULAs, because they themselves accepted similar restrictions on higher levels of trade. See *United States v. Microsoft Corp.*, 253 F.3d 34, 60-63 (D.C. Cir. 2001).

150. For more information on the Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, see WIPO, *Treaties and Contracting Parties: Berne Convention*, <http://www.wipo.int/treaties/en/ip/berne/> (last visited Jan. 21, 2006). See also SAM RICKETSON, *THE BERNE CONVENTION FOR THE PROTECTION OF LITERARY AND ARTIST WORKS: 1886-1986* (1987).

151. See, e.g., Lothar Determann & Aaron X. Fellmeth, *Don't Judge A Sale by its License: Software Transfers Under the First Sale Doctrine in the United States and the European Community*, 36 U.S.F. L. REV. 1 (2001) [hereinafter Determann & Fellmeth, *First Sale Doctrine*].

exhausted after that particular copy is sold.<sup>152</sup> Consequently, the first purchaser (e.g., a distributor, retailer, or end user) can freely resell a lawfully acquired copy to a secondary purchaser (e.g., a retailer, a friend, or a used book or CD trader).<sup>153</sup>

The U.S. Supreme Court created the First Sale Doctrine to strike a balance between (a) the interest in rewarding the copyright owner for her creativity with a right to control initial distribution, and (b) the public's interest in free flow of goods and access to copyrighted works after the copyright owner has been rewarded in connection with the first sale of a particular copy.<sup>154</sup> With respect to software, the traditional First Sale Doctrine alone cannot have the same effect that it does with respect to books or CDs. Software end users usually need to create additional copies in order to enjoy the use of a software copy—typically making one permanent copy in the process of installing the software on the hard drive of a computer and further temporary copies in RAM and CPU cache in the process of executing the program.<sup>155</sup> Neither the First Sale Doctrine nor the first purchaser of a software copy could confer such a right on a secondary purchaser. Therefore, the legislature added the software-specific Section 117(a)(1) to the Copyright Act to complement the First Sale Doctrine:<sup>156</sup>

[I]t is not an infringement for the owner of a copy of a computer program to make or authorize the making of another copy or adaptation of that computer program provided:

(1) that such a new copy or adaptation is created as an essential step in the utilization of the computer program in conjunction with a machine and that it is used in no other manner . . . .<sup>157</sup>

Pursuant to Section 117(a)(1) of the Copyright Act, end users who own software copies are not only permitted to duplicate their software, but also to adapt it as an essential step in the utilization of the software with a

---

152. 17 U.S.C. § 109 (2000).

153. Neither the first nor the secondary purchaser is permitted to make any additional copies or prepare derivative works—these rights remain with the copyright owner and are not exhausted after the first sale of a copy.

154. The first case to acknowledge this doctrine was *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908), which prevented the copyright owner from controlling retail pricing through a license condition. See also Determann & Fellmeth, *First Sale Doctrine*, *supra* note 151.

155. See *MAI Sys. Corp. v. Peak Computer, Inc.*, 991 F.2d 511, 518-19 (9th Cir. 1993).

156. See 2 NIMMER ON COPYRIGHT, *supra* note 11, § 8.08[D]; see also *MAI Sys.*, 991 F.2d at 519 n.6.

157. 17 U.S.C. § 117(a) (1).

machine,<sup>158</sup> that is, to create derivative works in the interest of interoperability.

From the outset, software companies have tried to circumvent this law by insisting that they never transfer ownership to copies of their software.<sup>159</sup> With few exceptions, including a relatively recent case,<sup>160</sup> the software industry has been able to prevail by adopting this position in U.S. courts.<sup>161</sup> Therefore, Section 117(a)(1) of the Copyright Act has not had a significant practical effect in terms of authorizing the preparation of derivative works for purposes of achieving interoperability.<sup>162</sup> German courts, on the other hand, have ignored the labels as well as most of the content of software license agreements and have qualified perpetual software transfers for fixed fees as sales.<sup>163</sup> This resulted in the availability of a (broader) equivalent of Section 117(a)(1)<sup>164</sup> for end users, developers, and distributors alike. Moreover, EULAs are often invalid due to a lack of deemed acceptance.<sup>165</sup>

---

158. *See, e.g.*, Krause v. Titleserv, 402 F.3d 119, 125 (2d Cir. 2005).

159. DSC Commc'ns Corp. v. Pulse Commc'ns Inc., 170 F.3d 1354, 1360 (Fed. Cir. 1999); Davidson v. Internet Gateway, 334 F. Supp. 2d 1164, 1177 (E.D. Miss. 2004); Novell, Inc. v. Unicom Sales, Inc., No. C-03-2785 MMC, 2004 WL 1839117, at \*7 (N.D. Cal. Aug. 17, 2004); Softman Prods. Co., LLC v. Adobe Sys. Inc., 171 F. Supp. 2d 1075, 1084 (C.D. Cal. 2001); Novell, Inc. v. CPU Distrib., Inc., No. H-97-2326, 2000 U.S. Dist. LEXIS 9975, at \*17-\*18 (S.D. Tex. May 4, 2000); Novell, Inc. v. Network Trade Ctr., Inc., 25 F. Supp. 2d 1218, 1230 (D. Utah 1997); Microsoft Corp. v. Harmony Computers & Elecs., 846 F. Supp. 208, 212-13 (E.D.N.Y. 1994); ISC-Bunker Ramo v. Altech, Inc., 765 F. Supp. 1310, 1314 (N.D. Ill. 1990); Data Prods. v. Reppart, 18 U.S.P.Q.2d 1058, 1061-62 (D. Kan. 1990).

160. *Softman Prods.*, 171 F. Supp. 2d at 1084, 1086.

161. *See generally* Christian H. Nadan, *Software Licensing in the 21st Century: Are Software "Licenses" Really Sales, and How Will the Software Industry Respond?*, 32 AIPLA Q.J. 555 (2004).

162. *See* NIMMER ON COPYRIGHT, *supra* note 11, § 8.08 [B], [D][2]. *But see* Krause, 402 F.3d 119.

163. *See* Determann & Fellmeth, *First Sale Doctrine*, *supra* note 151.

164. The German Copyright Act implements Council Directive 91/250/EEC of 14 May 1991 on the legal protection of computer programs, OJ L 122 pp. 42-46, which permits, in Articles 5 and 6, adaptations "where they are necessary for the use of the computer program by the lawful acquirer in accordance with its intended purpose, including for error correction."

165. If an end user does not need a license to execute a program, licensors find it much more difficult to persuade courts that the end user should be deemed to assent by way of implication because he or she tore away a shrink-wrap or clicked on a technically compulsory "accept" button during installation. *See, e.g.*, Jorge L. Contreras & Kenneth H. Slade, *Click-Wrap Agreements: Background and Guidelines for Enforceability*, 16 COMPUTER UND RECHT INT'L 104, 104-09 (2000).

## B. Typical License Clauses

Copyright owners commonly include a number of definitions, limitations, and restrictions regarding permissible software use in their software license agreements regarding, for example, the licensed territory in which the software may be used, the number of copies (archival and production copies) the licensee may create, the type and number of individual end users (e.g., licensee's employees) who may access the software copies, and the type of equipment on which the software may be used (defined by serial number or performance categories, e.g., MIPS, processor speed, etc.).<sup>166</sup>

Additionally, most commercial software companies typically prohibit any reverse engineering and any modifications to their copyrighted object code.<sup>167</sup> They may also state that the limited warranty does not cover malfunctions caused by operating the licensed software in conjunction with other software products or programs for which the licensed software is not recommended according to the documentation.

Commercial software license agreements, however, do not commonly contain express prohibitions on executing or combining the licensed software with other products as suggested in the preamble to Part VI of this article. The author of this Article has never seen such a contractual prohibition in any commercial license agreement outside the "free software" context<sup>168</sup> or encountered any case where a commercial software manufacturer has complained about software combinations, except with respect to video game screen manipulations and cases involving other infringements.<sup>169</sup> Most EULAs are silent on this point and contain relatively broad license grants. This, coupled with the fact that accompanying documentation often expressly contemplates combination with particular products, may be interpreted to permit the preparation of derivative works in the

---

166. See, e.g., Microsoft, Microsoft Windows XP Professional End User Licensing Agreement, <http://www.microsoft.com/windowsxp/sp2/proeula.msp> (last visited Nov. 16, 2006); Adobe, Adobe Acrobat Reader End User Licensing Agreement, <http://www.adobe.com/products/acrobat/acreula.html> (last visited Jan. 21, 2006).

167. See 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][3][b].

168. The GPL and other open source license agreements contain such restrictions, but are usually not referred to as "commercial" agreements, even though commercial software companies are using these agreements. See *infra* Part VI.

169. The computer game cases are addressed *supra* Part IV. *Dun & Bradstreet Software Services, Inc. v. Grace Consulting, Inc.*, 307 F.3d 197 (3d Cir. 2002) involved numerous other alleged infringements, so that the potential infringement of Geac's adaptation right was not ultimately dispositive. See also McGowan, *supra* note 137.

program combination context.<sup>170</sup> If software companies want to prevent consumers from using their programs in combination with other software, the companies usually design the programs or file formats to be incompatible and then take actions based on copyright law against competitors when they try to circumvent the compatibility hurdles.<sup>171</sup>

Thus, it seems that prevailing commercial licensing practices do not necessitate a more cautious answer than that based on statutory copyright law. Yet, given the possibility that commercial licensing practices may change—for example, in response to the “free software” licensing practices discussed in more detail in Part VII—applicable legal limitations are summarized in the remainder of this Part.

### C. Statutory Limitations on License Restrictions

As a matter of public policy, restrictions in license agreements (and other contracts, for that matter) are subject to scrutiny under a number of different statutes and legal theories, including competition laws, the doctrine of copyright misuse, and laws against unfair contractual terms.

#### 1. Competition Law

Section 1 of the U.S. Sherman Act—and numerous laws in other jurisdictions modeled after it—prohibits or restricts contractual restraints of trade.<sup>172</sup> A restraint on trade is any agreement wherein one of the contract-

---

170. For example, many EULAs contain a license to “use” the program. While the term “use” has a specific meaning in the patent context, it is not defined in the Copyright Act and courts have decided that a contractually granted right to “use” may imply broad rights under copyright law, including a right to prepare derivative works. *See, e.g., Kennedy v. Nat’l Juvenile Detention Ass’n*, 187 F.3d 690, 695 (7th Cir. 1999).

171. *See, e.g., Lexmark Int’l, Inc. v. Static Control Components, Inc.*, 387 F.3d 522 (6th Cir. 2004); *Chamberlain Group, Inc. v. Skylink Techs., Inc.*, 381 F.3d 1178 (Fed. Cir. 2004); *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992); *see also* Open Design Alliance, <http://www.opendesign.com> (last visited Nov. 16, 2006) (discussing the Open Design Alliance and the struggle of competitors of Autodesk Inc.). Another recent example is the fight about interoperability and music file formats, e.g., between Apple and RealNetworks. *See* Sean Captain, *My Songs, My Format*, N.Y. TIMES, Oct. 6, 2005, at C9.

172. The Sherman Act, 15 U.S.C. § 1 (2000), provides “Trusts, etc., in restraint of trade illegal; penalty: Every contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States, or with foreign nations, is declared to be illegal.” Treaty Establishing the European Community, art. 81, Dec. 24, 2002, 2002 O.J. (C 325) 33 provides that “[t]he following shall be prohibited as incompatible with the common market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the common market . . . .”

ing parties is restrained in the manner it conducts business with third parties.<sup>173</sup> Courts in the U.S. qualify relatively few types of restraints as “per se illegal”; they otherwise apply a “rule of reason,” which considers market power as a threshold requirement for antitrust scrutiny.<sup>174</sup> The European Community (EC) follows a similar approach in its block exemption regulations based on Article 81(3) of the EC Treaty.<sup>175</sup> While intellectual property laws do not provide absolute immunity from competition law,<sup>176</sup> competition law does not generally prohibit intellectual property owners from exercising the very “monopoly rights” that intellectual property laws confer upon them.<sup>177</sup> Thus, a copyright owner is generally free to exercise exclusionary rights and prohibit others from preparing derivative works.<sup>178</sup> If a copyright owner grants conditional permission to prepare derivative works, however, such conditions are fully subject to competition law scrutiny. For example, grant-back clauses (i.e., clauses that allow licensees to further develop licensed technology on the condition that they transfer ownership of all improvements back to the licensor) are considered anti-competitive, because of the potential concentration of additional intellectual property rights under the control of the licensor.<sup>179</sup>

---

173. JOSEPH P. BAUER & EARL W. KINTNER, 1-2 FEDERAL ANTITRUST LAW § 2.4 (1998).

174. *See, e.g.*, WILLIAM C. HOLMES, ANTITRUST LAW HANDBOOK, § 1.04 n.7 (1987).

175. *See, e.g.*, Andreas Heinemann, *Kartellrecht und Informationstechnologie*, 21 COMPUTER UND RECHT 715, 718-19 (2005) (F.R.G.); Romina Polley, *Softwareverträge und ihre Kartellrechtliche Wirksamkeit*, 20 COMPUTER UND RECHT 641 (2004) (F.R.G.).

176. *See, e.g.*, United States v. Microsoft Corp., 253 F.3d 34, 63 (D.C. Cir. 2001); U.S. DEP’T OF JUSTICE & FED. TRADE COMM’N, ANTITRUST GUIDELINES FOR THE LICENSING OF INTELLECTUAL PROPERTY § 2.2 (1995), available at <http://www.usdoj.gov/atr/public/guidelines/0558.pdf> [hereinafter DOJ, ANTITRUST GUIDELINES]; Commission Notice – Guidelines on the application of Article 81(3) of the EC Treaty to Technology Transfer Agreements (TTBER), 2004 O.J.(C 101/02) 107-11.

177. *See* Gerald Spindler, *Rechtsfragen der Open Source Software*, Rechtsgutachten im Auftrag des Verbandes der Softwareindustrie Deutschlands e. V. (VSI), [http://lehrstuhl-spindler.uni-goettingen.edu/extern/ross/downloads/studie\\_final.pdf](http://lehrstuhl-spindler.uni-goettingen.edu/extern/ross/downloads/studie_final.pdf) (last visited Dec. 12, 2006) [hereinafter Spindler]. *See generally* Judgment of the European Court of Justice of Apr. 6, Joined Cases C-241/91 P & C-242/91 P, *Radio Telefis Eireann v. Comm’n*, 1995 E.C.R. I-743, available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:61991J0241:EN:HTML>; WILLIAM C. HOLMES, INTELLECTUAL PROPERTY AND ANTITRUST LAW § 36:1 (2005) [hereinafter HOLMES, INTELLECTUAL PROPERTY AND ANTITRUST LAW].

178. *See* Spindler, *supra* note 177. For materials on the Noerr Pennington doctrine, see 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][1]; DOJ, ANTITRUST GUIDELINES, *supra* note 176, § 2.2.

179. *See, e.g.*, *Transparent-Wrap Mach. Corp. v. Stokes & Smith Co.*, 329 U.S. 637, 642 (1947); *Santa Fe-Pomeroy, Inc. v. P & Z Co.*, 569 F.2d 1084, 1101 (9th Cir. 1978); *Robintech, Inc. v. Chemidus Wavin, Ltd.*, 450 F. Supp. 817, 822 (D.D.C. 1978), *aff’d*,

Thus, a licensor can rely on copyright law to contractually prohibit licensees from creating derivative works through software combinations or otherwise. Such a restraint should generally be exempt from antitrust scrutiny as a legitimate exercise of statutory rights granted under copyright law. A licensor that contractually prohibits the combination of its software with any other programs, however, would exceed the safe haven established by copyright law and be subject to full competition law scrutiny. A licensor with appreciable market power risks violating applicable antitrust laws and providing copyright infringers with a defense.<sup>180</sup>

## 2. *Copyright Misuse*

According to the doctrine of copyright misuse, copyright owners may not, through a contract or otherwise,<sup>181</sup> magnify their rights beyond those sanctioned by the Copyright Act.<sup>182</sup> The policy behind the doctrine of copyright misuse bears some resemblance to competition law policies. However, two main differences apply: the copyright misuse doctrine provides only a defense, and never an offensive claim, and its viability does not depend on market power of the misuser.<sup>183</sup>

---

628 F.2d 132 (D.C. Cir. 1980); *Old Dominion Box Co. v. Cont'l Can Co.*, 273 F. Supp. 550 (S.D.N.Y. 1967), *aff'd*, 393 F.2d 321 (2d Cir. 1968); *United States v. Associated Patents, Inc.*, 134 F. Supp. 74, 82 (E.D. Mich. 1955), *cert. denied*, 350 U.S. 960 (1956); DOJ, ANTITRUST GUIDELINES, *supra* note 176, § 5.6; Commission Regulation 772/2004, art. 5(1) (a)-(b), On the Application of Article 81 (3) of the Treaty to Categories of Technology Transfer Agreements, 2004 O.J. (L 123) 12, *available at* [http://eur-lex.europa.eu/LexUriServ/site/en/oj/2004/l\\_123/l\\_12320040427en00110017.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/oj/2004/l_123/l_12320040427en00110017.pdf); Fair Trade Commission of Japan, *Guidelines for Patent and Know-how Licensing Agreements under the Antimonopoly Act of Japan*, 27-28, <http://www.jftc.go.jp/e-page/legislation/ama/patentandknow-how.pdf> (last visited Jan. 23, 2006); HOLMES, INTELLECTUAL PROPERTY AND ANTITRUST LAW, *supra* note 177, § 23:2; RAYMOND T. NIMMER, THE LAW OF COMPUTER TECHNOLOGY § 7:139-41 (3d ed. 2006). The concern is that more and more related intellectual property rights, and thus market power, accumulates under the control of the original licensor, which seems particularly dangerous in the patent context where competition cannot develop based on independent development. Whether the same concerns apply in the copyright law context can be questioned, but to the extent copyright law is used to protect functional software, the situation seems at least more similar to patent protection than to the more traditional works of authorship (e.g., novels or music).

180. NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][2].

181. *Assessment Techs. v. WIREDData, Inc.*, 350 F.3d 640, 647 (7th Cir. 2003).

182. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 977 (4th Cir. 1990); 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][3][c]; David Nimmer, et al., *The Metamorphosis of Contract Into Expand*, 87 CALIF. L. REV. 17 (1999) [hereinafter Nimmer, *Metamorphosis*].

183. *See* 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][2]; *see also* *Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc.*, 342 F.3d 191, 205 (3d Cir. 2003) (finding that copyright misuse may be possible even absent anticompetitive behavior), *cert.*

In the seminal case, *Lasercomb America, Inc. v. Reynolds*,<sup>184</sup> Lasercomb and Holliday Steel were competitors in the manufacture of steel rule dies. Lasercomb developed a software program that directed the mechanized creation of a steel rule die from a computerized template. Reynolds, a computer programmer for Holliday Steel, made unauthorized copies of the program and then marketed a new program that was extremely similar to Lasercomb's program. A clause in Lasercomb's standard licensing agreement prohibited the development and distribution of any competing programs.<sup>185</sup> The court considered this clause to amount to misuse, because the licensor used its limited exclusionary rights under copyright law to prohibit activities entirely outside the scope of its copyright, namely, the independent development of programs with similar functionality. As a consequence, the court denied copyright protection to the copyright owner even though the defendant sold what would otherwise be considered pirated copies and had not even concluded the questionable license agreement.<sup>186</sup>

Subsequent federal appellate courts have found copyright misuse in a number of other situations: where a copyright owner agreed to license a product in exchange for an agreement by the licensee not to use a competitor's product,<sup>187</sup> where a copyright owner used its software copyrights to prevent licensees from substituting hardware supplied by the copyright owner,<sup>188</sup> where a copyright owner tried to prohibit criticism in the context in which its copyrighted video clips would appear on licensee's website,<sup>189</sup> and where an owner of copyrighted software attempted to use its copyright

---

*denied*, 540 U.S. 1178 (2004); *Alcatel U.S.A., Inc. v. DGI Techs., Inc.*, 166 F.3d 772, 792 (5th Cir. 1999) (stating that the doctrine of copyright misuse "bars a culpable plaintiff from prevailing on an action for the infringement of the misused copyright").

184. *Lasercomb*, 911 F. 2d 970.

185. The clause stated: "Licensee agrees . . . that it will not permit or suffer its directors, officers and employees, directly or indirectly, to write, develop, produce or sell computer assisted die making software." *Id.* at 973.

186. Neither Holliday Steel nor Reynolds had signed the licensing agreement or was bound by it; however, the court held that, due to public policy concerns, the copyright misuse defense was available to defendants who had not been injured by the misuse. *Id.* at 979.

187. *Practice Mgmt. Info. Corp. v. Am. Med. Ass'n*, 121 F.3d 516, 520 (9th Cir. 1997), *cert. denied*, 522 U.S. 933 (1997).

188. *Alcatel*, 166 F.3d at 793; *DSC Commc'ns Corp. v. DGI Techs., Inc.*, 81 F.3d 597 (5th Cir. 1996). For additional information regarding the history and the results of the proceedings in these related cases, see 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A][2][b] n.45.

189. *Video Pipeline, Inc. v. Buena Vista Home Entm't, Inc.*, 342 F.3d 191, 205-06 (3d Cir. 2003), *cert denied*, 540 U.S. 1178 (2004).

to prevent access to non-copyrightable data.<sup>190</sup> In each instance, the courts were concerned copyright owners could have used their copyrights as leverage to control matters that the legislature intentionally kept outside the scope of the limited copyright “monopoly.”<sup>191</sup>

Courts that have embraced the doctrine of copyright misuse<sup>192</sup> initially adopted a rationale developed in the patent context.<sup>193</sup> The doctrine of patent misuse penalizes patent holders who try to expand their limited legal monopoly over the patented invention beyond the “four corners of the patent” and thus upset the balance that patent law has struck between protection and public access.<sup>194</sup> In patent cases, courts have found a number of licensing practices to be abusive, including royalty requirements for components, territories, or time periods outside the scope of the patent grant, covenants not to deal in competing products, and package licensing.<sup>195</sup>

A licensor who contractually prohibited the combination of its software with other programs in situations where adaptation rights are not affected would exceed the scope of its copyright by seeking to control external activities and subject matter—namely, the use of independent pro-

190. *Assessment Techs. v. WIREData, Inc.*, 350 F.3d 640, 646-47 (7th Cir. 2003).

191. For an overview, see 4 NIMMER ON COPYRIGHT, *supra* note 11, § 13.09[A].

192. *See id.*, § 13.09[A][2]. Some courts have not been as reluctant. According to *Davidson v. Internet Gateway*, 334 F. Supp. 2d 1164, 1181 (E.D. Miss. 2004), (citing *Schoolhouse, Inc. v. Anderson*, No. 91-2324, 2001 WL 1640081, at \*7 (D. Minn. Nov. 8, 2001)), the court stated that “[a]buse of copyright is generally recognized as an equitable affirmative defense to a copyright infringement claim.”

193. *See United States v. Loew’s, Inc.*, 371 U.S. 38, 46-47 (1962); *United States v. Paramount Pictures*, 334 U.S. 131, 157-59 (1948).

194. *See Morton Salt Co. v. G.S. Suppiger Co.*, 314 U.S. 488, 492-93 (1942). The legislature somewhat limited the applicability of the patent abuse doctrine in the 1988 Patent Misuse Reform Act, which provides that:

No patent owner otherwise entitled to relief for infringement or contributory infringement of a patent shall be denied relief or deemed guilty of misuse or illegal extension of the patent right by reason of his having done one or more of the following: . . . (5) conditioned the license of any rights to the patent or the sale of the patented product on the acquisition of a license to rights in another patent or purchase of a separate product, unless, in view of the circumstances, the patent owner has market power in the relevant market for the patent or patented product on which the license or sale is conditioned.

35 U.S.C. § 271(d)(4)-(5) (2000).

195. *See generally* 6 DONALD S. CHISUM, CHISUM ON PATENTS § 19.04 (2005). Whether it is appropriate to transfer the abuse doctrine from patent law to copyright law may seem questionable in the context of traditional works of authorship, given the relatively smaller impact of the copyright exclusion rights on innovation and commerce; in the context of protection for functional software, however, it seems more likely that similar problems to those in the context of patent might arise from misuse.

grams. Depending on the context, such a clause could, in effect, constitute a prohibition on using competing products. In any event, such a clause would limit a licensee's right to create compilations and non-creative combinations—rights that the Copyright Act declares to be free, in contrast to the right to prepare derivative works.<sup>196</sup> Thus, such a copyright owner would seem to run a significant risk that a court would classify such a clause as copyright misuse with the dramatic result that the copyright owner would be denied copyright protection even against outright piracy. A licensor who merely prohibits licensees from creating derivative works, as the term is defined by statute and through combinations or otherwise, would generally remain within the scope of its statutory rights and not risk a finding of copyright misuse.

The need to prevent an abuse of intellectual property law is internationally recognized.<sup>197</sup> Nevertheless, a doctrine similar to copyright misuse does not seem common in national copyright laws outside the United States.<sup>198</sup> One reason for this may be that courts in other jurisdictions find more effective tools to sanction abuse in stricter competition or unfair contract terms laws.

### 3. *Unfair Contract Terms*

Clauses in license agreements, as in other contracts, are subject to scrutiny under doctrines such as contract of adhesion<sup>199</sup> and unconscionability.<sup>200</sup> Unilateral licenses, which are permissions subject to certain

---

196. 17 U.S.C. §§ 103, 106(2); *see supra* Section III.D.

197. *See, e.g.*, Agreement on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Counterfeit Goods, Apr. 15, 1994, 1869 U.N.T.S. 299, 33 I.L.M. 1125, arts. 8.2, 40.2 [hereinafter TRIPS]. *See generally* Lucie Guibault, *Limitations Found Outside of Copyright Law—The Exceptions and Limitations to Copyright*, ALAI Study Days, Cambridge, Mass., September 14-17, 1998, S. 1. For a perspective on the EU position regarding TRIPS obligations and competition law, *see generally* Hanns Ullrich, *Expansionist Intellectual Property Protection and Reductionist Competition Rules: A TRIPS Perspective*, 7 J. INT'L ECON. L. 401 (2004).

198. *See* P.E. Geller/M.B. Nimmer, *International Copyright Law and Practice*, Release No. 17, October 2005, *e.g.*, *France* (A. Lucas/P. Kamina), *Germany* (A. Dietz) and *United Kingdom* (L. Bently).

199. An adhesion contract is “[a] standard-form contract prepared by one party, to be signed by the party in a weaker position, usu. a consumer, who adheres to the contract with little choice about the terms.” BLACK’S LAW DICTIONARY 341-42 (8th ed. 2004).

200. Unconscionability is “[t]he principle that a court may refuse to enforce a contract that is unfair or oppressive because of procedural abuses during contract formation or because of overreaching contractual terms, esp. terms that are unreasonably favorable to one party while precluding meaningful choice for the other party.” *Id.* at 1560-61.

conditions but do not contemplate the conclusion of a negotiated contract, are generally subject to similar interpretation and limitation rules.<sup>201</sup>

Some foreign jurisdictions apply relatively strict scrutiny when evaluating contracts, especially consumer contracts and contracts between businesses that are not individually negotiated. For example, the European Community has enacted the Directive on Unfair Terms in Consumer Contracts.<sup>202</sup> Germany has extremely rigid rules on “standard contract terms,”<sup>203</sup> and the UK passed the Unfair Contract Terms Act in 1977.<sup>204</sup> Courts in these jurisdictions usually strike contract clauses in their entirety if they contain sections that are considered to be unfair, and refuse to blue-pencil overreaching clauses (i.e., strike only unfair portions of a clause or otherwise reduce an overreaching clause to what would have been permissible) in order to discourage overreaching contract drafting.<sup>205</sup>

Courts in the United States, on the other hand, generally restrain themselves from questioning contract terms except in extreme circumstances. This has resulted in seemingly unbalanced and poorly considered licensing practices, particularly in the context of shrink-wrap and click-through agreements.<sup>206</sup> Litigators have so far attacked such agreements commonly on procedural grounds,<sup>207</sup> as the law seems less developed and less favor-

---

201. See, e.g., *Womack+Hampton Architects, LLC v. Metric Holdings Ltd.*, 102 F. Appx. 374, 378 (5th Cir. 2004); *Kennedy v. Nat'l Juvenile Detention Ass'n*, 187 F.3d 690, 694 (7th Cir. 1999); *Fantastic Fakes, Inc. v. Pickwick Int'l, Inc.*, 661 F.2d 479, 483 (5th Cir. 1981); *Nimmer, Metamorphosis*, *supra* note 182, at 30.

202. Council Directive 93/13/EEC of 5 April 1993 on Unfair Terms in Consumer Contracts.

203. See generally Bürgerliches Gesetzbuch [BGB] [Civil Code] Aug. 18, 1896, Reichsgesetzblatt [RFBI] 195, §§ 305-310.

204. Unfair Contract Terms Act, 1977, c. 50, §§ 3-7 (U.K.).

205. Examples for standard clauses in software license agreements invalidated by German courts: Bundesgerichtshof [BGH] [Federal Court of Justice] No. 28, 1979, 33 (1980) *Neue Juristische Wochenschrift* [NJW] 832 (F.R.G.) (clause voiding warranty in case of reassembly of software); Oberlandesgericht [OLG] Nürnberg [Nürnberg Court of Appeals] June 20, 1989, 6 (1990) *Computer und Recht* [CR] 118 (121), (F.R.G.) (prohibition of resale); Oberlandesgericht [OLG] Hamm [Hamm Court of Appeals] Dec. 12, 1988, 44 (1989) *Neue Juristische Wochenschrift* [NJW] 1041 (F.R.G.) (deemed acceptance upon delivery); Bundesgerichtshof [BGH] [Federal Court of Justice] Feb. 25, 1981, 34 (1981) *Neue Juristische Wochenschrift* [NJW] 1501 (F.R.G.) (warranty disclaimer).

206. 1 NIMMER ON COPYRIGHT, *supra* note 11, § 3.04[B][3][a]. See generally Mark A. Lemley, *Intellectual Property and Shrinkwrap Licenses*, 68 S. CALIF. L. REV. 1239 (1995); Loren, *supra* note 150.

207. Many shrink-wrap and click-wrap cases focus on whether contract formation mechanisms were flawed, e.g., because customers did not receive sufficient notice of the contract terms. See, e.g., *Specht v. Netscape Commc'ns*, 306 F.3d 17, 31 (2d Cir. 2002);

able on substantive challenges. However, as software companies are updating their contract formation procedures, taking existing case law into consideration, more and more substantive challenges can be expected.

Under general principles of contract law in the United States, a contract of adhesion is a “standardized contract prepared entirely by one party to the transaction for the acceptance of the other.”<sup>208</sup> A party presented with a contract of adhesion has no opportunity for bargaining;<sup>209</sup> she is forced to accept or reject the contract terms on a “take it or leave it” basis, usually having no other way of obtaining the desired product or service.<sup>210</sup> A contract is not automatically unenforceable simply because it is formed according to one party’s initiative, without negotiation, and/or using boilerplate terms.<sup>211</sup> However, courts subject contracts of adhesion to heightened scrutiny, inquiring whether their terms fall within the “adhering” party’s reasonable expectations.<sup>212</sup> Similarly, the doctrine of unconscionability allows courts to refuse to enforce an unconscionable provision in a one-sided contract.<sup>213</sup> The test for unconscionability is “whether, in light of the general background and the needs of a particular case, the clauses involved are so one-sided as to be unconscionable under the circumstances existing at the time of the making of the contract.”<sup>214</sup>

---

*see also* Lothar Determann & Saralyn Ang-Olson, *Comment on Specht v. Netscape*, 6 INTELL. PROP. L. BULL. 38, 39 (2001).

208. *Painters Dist. Council No. 33 v. Moen*, 128 Cal. App. 3d 1032, 1039 (Ct. App. 1982).

209. *Id.* at 1039.

210. *Id.*

211. *See* RESTATEMENT (SECOND) OF CONTRACTS § 211(1) (1981) (“[W]here a party to an agreement . . . manifests assent to a writing and has reason to believe that like writings are regularly used to embody terms of agreements of the same type, he adopts the writing as an integrated agreement with respect to the terms included in the writing.”).

212. *Ericksen, Arbuthnot, McCarthy, Kearney & Walsh, Inc. v. 100 Oak Street*, 673 P.2d 251, 257 (Cal. 1983).

213. *See, e.g.*, CAL. CIV. CODE § 1670.5 (West 1985); *Am. Online, Inc. v. Super. Ct.*, 90 Cal. App. 4th 1 (Ct. App. 2001).

214. CAL. CIV. CODE § 1670.5 cmt. 1; *see also* *Cooper v. MRM Inv. Co.*, 367 F.3d 493, 503-04 (6th Cir. 2004); *Davidson v. Internet Gateway*, 334 F. Supp. 2d 1164, 1179 (E.D. Miss. 2004). Unconscionability has both a procedural and substantive element. *Cooper*, 367 F.3d at 503; *Freeman v. Wal-mart Stores, Inc.*, 111 Cal. App. 4th 660, 669 (Ct. App. 2003). Similar to the doctrine of adhesion contracts, the procedural element focuses on oppression or surprise due to unequal bargaining power, and substantive unconscionability focuses on overly harsh or one-sided results. Generally, courts require that both procedural and substantive unconscionability be present, although not to the same degree, before a court will refuse to enforce a contract or clause due to unconscionability. *See Freeman*, 111 Cal. App. 4th at 669; *Pardee Constr. Co. v. Super. Ct. of San Diego County*, 100 Cal. App. 4th 1081, 1088 (Ct. App. 2002) (“In other words, the more

With respect to the doctrine of copyright misuse and competition law, copyright owners that maneuver within the express grant of rights afforded by the Copyright Act are far less likely to be successfully challenged because it is usually not unfair to contractually exercise rights specifically granted by the legislature.<sup>215</sup> Thus, a copyright owner who simply prohibits the preparation of derivative works in an end user license agreement is generally on safe ground, whereas a licensor who prohibits program combinations without regard to copyright law will more likely be subject to scrutiny under unfair contract terms law.<sup>216</sup>

#### **D. License Scope Definitions vs. License Conditions vs. Contractual Covenants**

Copyright owners can largely stay clear of the risks described in the preceding Section if they keep in mind a simple rule: license scope limitations within the concepts expressly provided by copyright law, such as prohibiting the preparation of derivative works, are generally permissible; conditions and covenants that are alien to copyright law, such as prohibiting the execution of one program with another independent program or requiring a grant-back of copyright in modifications, are subject to scrutiny and can be invalid, illegal, and/or copyright misuse.

Conceptually, restrictions can be introduced through different types of mechanisms. The least intrusive of these mechanisms is a license scope definition, a clause that strives to prevent granting an implied license by clarifying that the copyright owner does not intend to authorize a certain activity.<sup>217</sup> Most intrusive would be a license condition, a clause that conditions the license grant on compliance with all contractual prohibitions in the license agreement.<sup>218</sup> Somewhere in between is a contractual covenant

---

substantively oppressive the contract term, the less evidence of procedural unconscionability is required to come to the conclusion the term is unenforceable, and vice versa.”).

215. Spindler, *supra* note 177, at 51.

216. *Id.*

217. Such a clause could follow the actual license grant and be phrased, for example: Licensor reserves all other rights, including, without limitation, the right to prepare derivative works of the Licensed Software. Licensee understands that Licensor does not positively authorize the installation or execution of any programs made by other manufacturers in combination with the Licensed Software. Whether or not such combinations are permissible depends on applicable law, including, without limitation, 17 U.S.C. § 106(2). Licensee agrees to apply with all applicable laws.

218. Such a clause could follow the actual license grant and be phrased, for example: “Aforesaid license grant is made under the condition that Licensee refrains from installing or executing any programs made by other manufacturers in combination with the Licensed Software. If Licensee violates this condition, the License grant shall immediately

that imposes obligations which the licensor can enforce against the licensee independent of the license grant.<sup>219</sup> The stronger the mechanism, the greater the chance that a court will invalidate an overreaching restriction.

### E. Summary

The author of this Article is not aware of any commercial software licenses that try to control software combinations beyond general, unspecific contractual prohibitions of adaptations (which are quite common). If this practice were to change, and copyright owners were to use their rights under copyright law to prohibit software combinations regardless of whether they qualify as derivative works, they would risk running afoul of competition laws, the doctrine of copyright misuse, and laws prohibiting or invalidating unfair contract terms. The extent of the actual risks depends on a number of circumstances, including the intrusiveness of the restriction mechanism (a license condition being more problematic than a scope limitation, for example), the licensor's market power, and the applicable law. For example, licensors would face relatively high risks under German competition and contracts law, or under the U.S. doctrine of copyright misuse.

## VII. SOFTWARE COMBINATIONS UNDER THE GENERAL PUBLIC LICENSE

Since courts first decided to afford copyright protection to computer programs, commercial software development companies have had a strong incentive to avoid reusing existing code owned by others.<sup>220</sup> Independent creation is a defense to copyright infringement,<sup>221</sup> and so software devel-

---

expire." GPL § 4, discussed in more detail in Part VII of this Article, also contains such a condition: "You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License." More commonly in practice—and less clear for licensees and courts—copyright owners preface the license grant sections with a phrase such as "subject to all terms and conditions of this Agreement, Licensor grants Licensee a license to . . ." and include the limitations as separate restrictions in different clauses without express designation as a license condition.

219. Such a clause could be phrased, for example, in the following manner: "Licensee shall refrain from installing or executing any programs made by other manufacturers in combination with the Licensed Software." If a licensee violates such a clause, its license rights would not necessarily be affected and the licensor would have to seek recourse based on breach of contract remedies. *See, e.g., Sun Microsystems, Inc. v. Microsoft Corp.*, 188 F.3d 1115 (9th Cir. 1999).

220. *See* Lemley, *supra* note 83.

221. 3 NIMMER ON COPYRIGHT, *supra* note 11, §§ 12.10[B][2][b], 12.11[D].

opment companies often opt for creating programs from scratch, ideally in a “clean room” environment, so they can prove that their products are not copies of existing programs with similar functionality.<sup>222</sup> Thus, the decision in favor of software copyrightability had a rather dramatic impact on the professional lives and day-to-day activities of programmers: instead of being asked to further develop and improve the “state of the art” and to focus on cutting-edge problems, programmers were asked to spend most of their time reinventing the wheel. Why? Because lawyers did not have the energy or wit to come up with a more fitting intellectual property law regime tailored to software.<sup>223</sup>

So, the programmers had their revenge on the lawyers.<sup>224</sup> Programmers invented “copyleft”<sup>225</sup> to fight copyright law, published a manifesto,<sup>226</sup> and created a new set of license terms intended to free software from the shackles and chains of copyright protection and become the new standard for licensing software to the public: the General Public License (GPL).<sup>227</sup>

---

222. See Wikipedia.org, Cleanroom (Software Engineering), [http://en.wikipedia.org/wiki/Cleanroom\\_\(Software\\_Engineering\)](http://en.wikipedia.org/wiki/Cleanroom_(Software_Engineering)) (last visited Jan. 21, 2006).

223. Many practitioners and academics have proposed various types of *sui generis* protection regimes for software in both the U.S. and abroad. See, e.g., John C. Phillips, *Sui Generis Intellectual Property Protection for Software*, 60 GEO. WASH. L. REV. 997 (1992); Pamela Samuelson, et al., *A Manifesto Concerning the Legal Protection of Computer Programs*, 94 COLUM. L. REV. 2308 (1994); Richard H. Stern, *A Sui Generis Utility Model Law as an Alternative Legal Model for Protecting Software*, 1 U. BALT. INTELL. PROP. L.J. 108 (1993); Steven B. Toeniskoetter, *Protection of Software Intellectual Property in Europe: An Alternative Sui Generis Approach*, 10 INTELL. PROP. L. BULL. 65 (2005). But see Brett A. Carlson, *On the Wrong Track: A Response to the Manifesto and a Critique of Sui Generis Software Protection*, 37 JURIMETRICS J. 187 (1997); Ginsburg, *Four Reasons*, *supra* note 83.

224. More specifically, the leader of the free software movement is Richard M. Stallman, Founder of the Free Software Foundation and the GNU Project, and primary author of the current GNU GPL. See generally Free Software Foundation, Leadership, <http://www.fsf.org/about/leadership.html> (last visited Nov. 16, 2006). For a more objective restatement of the GPL’s history, see Wikipedia.org, GPL, <http://en.wikipedia.org/wiki/Gpl> (last visited Jan. 21, 2006).

225. RICHARD M. STALLMAN, *Copyleft: Pragmatic Idealism*, in FREE SOFTWARE FREE SOCIETY: SELECTED ESSAYS OF RICHARD M. STALLMAN 91 (Joshua Gay ed., 2002), available at <http://www.fsf.org/licensing/essays/pragmatic.html>.

226. RICHARD M. STALLMAN, *The GNU Manifesto*, in FREE SOFTWARE FREE SOCIETY: SELECTED ESSAYS OF RICHARD M. STALLMAN 31 (Joshua Gay ed., 2002), available at <http://www.gnu.org/gnu/manifesto.html>.

227. GNU Project, GNU General Public License, Version 2 (June 1991), <http://www.gnu.org/licenses/gpl.html> [hereinafter GPL License]. Richard M. Stallman has recently unveiled a draft of the next revision of the GPL. See Mark Baard, *New GPL is Free at Last*, WIRED, Jan. 17, 2006, [http://wired.com/news/technology/0,70028-0.html?tw=wn\\_tophead\\_1](http://wired.com/news/technology/0,70028-0.html?tw=wn_tophead_1).

The ultimate goal was to “free”<sup>228</sup> software and spread the “freed software”, licensed under the GPL, to replace and ultimately eliminate<sup>229</sup> all proprietary software commercialized through restrictive licensing.<sup>230</sup> Fighting fire with fire, the means to reach this goal would be copyright law: any copyright owner who released software under the GPL would require anybody else to apply the GPL to any new versions of that code and would bring copyright infringement actions against anyone who breached the GPL. Anybody who distributes software outside the scope of the applicable license agreements lacks a valid authorization required by the Copyright Act<sup>231</sup> and thus commits copyright infringement.<sup>232</sup> This cunning plan for revenge worked quite well, judging by the increasingly publicized success of the free software movement,<sup>233</sup> the outcries from the legal community,<sup>234</sup> and the software companies with proprietary licensing models.<sup>235</sup>

---

228. See Free Software Foundation, The Free Software Foundation Definition, <http://www.fsf.org/licensing/essays/free-sw.html> (last visited Jan. 21, 2006) (“Free software is a matter of the users’ freedom to run, copy, distribute, study, change and improve the software.”).

229. “The enemy is proprietary software.” RICHARD M. STALLMAN, *Why “Free Software” is Better than “Open Source,”* in FREE SOFTWARE FREE SOCIETY: SELECTED ESSAYS OF RICHARD M. STALLMAN 31 (Joshua Gay ed., 2002), available at <http://www.fsf.org/licensing/essays/free-software-for-freedom.html>.

230. See Free Software Foundation, GNU Project, Categories of Free and Non-Free Software, <http://www.fsf.org/licensing/essays/categories.html#ProprietarySoftware> (last visited Jan. 21, 2006). Other supporters of the movement have taken more moderate positions, stressing the importance of both proprietary and open source software. See, e.g., Lawrence Lessig, *Open-Source, Closed Minds*, EWEEK.COM, Oct. 1, 2003, <http://www.eweek.com/article2/0%2C1895%2C1309535%2C00.asp> (“[A] balance between proprietary and nonproprietary property is better than either extreme.”).

231. 17 U.S.C. § 106(3). See, e.g., *Apple Computer, Inc. v. Microsoft Corp.*, 35 F.3d 1435, 1441 (9th Cir. 1994).

232. See The GPL-Violations.Org Project, <http://gpl-violations.org/> (last visited Jan. 21, 2006) (cataloging past and present GPL infringers).

233. See, e.g., James Boyle, *The Public Domain: The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33 (2003); Peter Galli, *GPL Could Put Heat on Microsoft*, EWEEK.COM, Nov. 29, 2004, <http://www.eweek.com/article2/0,1759,1732567,00.asp>.

234. See, e.g., Heidi S. Bond, *What’s So Great About Nothing? The GNU General Public License and the Zero-Price-Fixing Problem*, 104 MICH. L. REV. 547 (2005); Dan Hunter, *Culture War*, 83 TEX. L. REV. 1105 (2005); Robert P. Merges, *The End of Friction? Property Rights and Contract in the “Newtonian” World of On-Line Commerce*, 12 BERKELEY TECH. L.J. 115 (1997); Mathias Strasser, *A New Paradigm in Intellectual Property Law? The Case Against Open Sources*, 2001 STAN. TECH. L. REV. 4 (2001); Greg Vetter, *Infectious Open Source Software: Spreading Incentives or Promoting Resistance?*, 36 RUTGERS L.J. 53, 162 (2004); Marc Kaufman et al., *Licensing Open Source*

Some commentators note that combinations incorporating code licensed under the GPL are prone to create derivative or “derived” works and consequently infringement risks.<sup>236</sup> In order to determine whether such combinations are as dangerous as these scholars and practitioners suggest, this Part reviews the mechanism of Section 2(b) of the GPL; possible interpretations of the definitional scope of Section 2(b); the potential impact of the First Sale Doctrine; and the potential impact of statutory and other legal limitations on license restrictions in the GPL. It concludes with an assessment that combinations with GPLed code are less dangerous than the GPL’s drafters claim and advances a few recommendations for the current GPL update process.

### A. Effects of GPL § 2(b)

Section 2(b) of the GPL requires makers of “derived works” to license those works under the GPL as well:

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1... provided that you also meet all of these conditions: [ . . . ]

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or

---

*Software*, Apr. 7, 2004, [http://www.nixonpeabody.com/copyright\\_article.asp?ID=1&PubType=A](http://www.nixonpeabody.com/copyright_article.asp?ID=1&PubType=A); MICHAEL TSUR & SHAY DAVID, A LICENSE TO KILL (INNOVATION): ON OPEN SOURCE LICENSES AND THEIR IMPLICATIONS FOR INNOVATION 37, [http://www.shaydavid.info/papers/shaydavid\\_a\\_license\\_to\\_kill\\_043005.pdf](http://www.shaydavid.info/papers/shaydavid_a_license_to_kill_043005.pdf) (last visited Jan. 27, 2006).

235. See, e.g., ComputerWire, *Microsoft Anti-GPL Fine Print Threatens Competition*, THE REGISTER, April 17, 2002, [http://www.theregister.co.uk/2002/04/17/microsoft\\_antigpl\\_fine\\_print\\_threatens](http://www.theregister.co.uk/2002/04/17/microsoft_antigpl_fine_print_threatens); Remarks by Bill Gates, Seattle, Washington, Apr. 17, 2002, <http://www.microsoft.com/billgates/speeches/2002/04-17glc.asp>; Thomas C. Greene, *Ballmer: “Linux is a Cancer” Contaminates All Other Software with Hippy GPL Rubbish*, THE REGISTER, June 2, 2001, [http://www.theregister.co.uk/2001/06/02/ballmer\\_linux\\_is\\_a\\_cancer/](http://www.theregister.co.uk/2001/06/02/ballmer_linux_is_a_cancer/); Mike Ricciuti, *Gates Wades Into Open-source Debate*, <http://news.com.com/2100-1001-268667.html> (last visited Jan. 24, 2006).

236. David S. Evans & Bernard J. Reddy, *Government Preferences for Promoting Open-Source Software: A Solution in Search of a Problem*, 9 MICH. TELECOMM. & TECH. L. REV. 313, 340 (2003); Klaus M. Schmidt & Monika Schnitzer, *Public Subsidies for Open Source? Some Economic Policy Issues of the Software Market*, 16 HARV. J.L. & TECH 473, 476-77 (2003); Ulrich Wuermeling & Thies Deike, *Open Source Software: Eine juristische Risikoanalyse*, 19 COMPUTER UND RECHT 87, 91 (2003); Gerald Spindler, *Vertragsrechtliche Fragestellungen der Open Source Software (Contractual Questions Concerning Open Source Software)*, Address Before the Committee for Contract Law of the German Society for Law and Informatics (Mar. 28, 2003) (transcript available at [http://www.dgri.de/ausschuesse/ver\\_030328.html](http://www.dgri.de/ausschuesse/ver_030328.html)).

any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.<sup>237</sup>

This requirement either has a “viral” or “immunizing”<sup>238</sup> effect on licensees, depending on the licensee’s objectives.

Section 4 of the GPL clarifies that all of its restrictions operate as conditions and that non-compliance will result in a termination of all rights. The GPL’s restriction mechanism, therefore, is a license condition—the most intrusive type of license restrictions discussed in Part VI.D:

You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License.

Section 7 of the GPL manifests what the Free Software Foundation refers to as the “liberty or death” principle:<sup>239</sup>

If . . . conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this Li-

---

237. GPL License, *supra* note 227, § 2(b).

238. “Viral” because code that is combined with GPLed code may become “infected.” *See, e.g.,* Nadan, *Open Source Licensing*, *supra* note 2, at 360; Andreas Guadamuz, *Viral Contracts or Unenforceable Documents? Contractual Validity of Copyleft Licenses*, 26 EUR. INTEL. PROP. REV. 331-39 (2004), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=56910](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=56910) [hereinafter Guadamuz, *Viral Contracts*]; Marc Kaufman et al., *Licensing Open Source Software*, Nixon Peabody Website, [http://www.nixonpeabody.com/copyright\\_article.asp?ID=1&PubType=A](http://www.nixonpeabody.com/copyright_article.asp?ID=1&PubType=A) (last visited Jan. 21, 2006); *see also* James Gibson, *Once and Future Copyright*, 81 NOTRE DAME L. REV. 167, 204 (2005); OpenSource.org, *The Open Source Definition*, <http://web.archive.org/web/20010330062316/www.opensource.org/docs/definition.html> (web-archive version). In version 1.9 of the Open Source Definition that is currently in effect, the OSI dropped the word “contamination.” For the latest version, see *The Open Source Definition*, Open Source Initiative, <http://www.opensource.org/docs/definition.php> (last visited Jan. 21, 2006). “Immunizing” because this clause prevents GPLed code from being incorporated into proprietary (i.e., non-free) derivative works. *Cf.* Kenneth J. Rodriguez, *Closing the Door on Open Source: Can the General Public License Save Linux and Other Open Source Software?*, 5 J. HIGH TECH. L. 403, 414 (2005); Greg R. Vetter, *The Collaborative Integrity of Open-Source Software*, UTAH L. REV. 563, 634 (2004); Michael Torrie, *New Samba License?*, <http://lists.ximian.com/pipermail/mono-list/2002-April/004787.html> (last visited Jan. 24, 2006).

239. Richard M. Stallman, *Microsoft’s New Monopoly*, <http://www.fsf.org/licensing/essays/microsoft-new-monopoly.html> (last visited Jan. 21, 2006) (“The General Public License forbid[s] publication of a modified version if it isn’t free software in the same way. (We call that the “liberty or death” clause, since it ensures the program will remain free or die.)”).

cense, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all.

With these two GPL clauses, copyright owners permit licensees to adapt and distribute copies of the GPLed software under the condition that licensees grant the copyright owner and everybody else a license under the terms of the GPL, at no charge and in object and source code form, to any works that are “derived from” the GPLed program.

The GPL does not prohibit licensees from charging money for selling software copies.<sup>240</sup> However, the GPL itself already provides everyone with a license free of charge. As a result, licensees have no realistic opportunity to commercialize their copyrights in works that they “derive” from GPLed code.<sup>241</sup> The GPL intentionally deprives copyright owners of the very economic incentive the Copyright Act has made available in the interest of furthering the development of original works of authorship—the right to exclude the public from copying, adapting, distributing, etc., and to charge a fee for individual licenses.

It is worth noting, however, that the GPL generally permits end users to execute GPLed code in any combination they want. According to Section 0 of the GPL, “[t]he act of running the Program is not restricted.” As a result, software companies do not have to be concerned about invoking the “viral” effect of the GPL based on a contributory liability theory if they distribute their add-on programs separately, or in other words, not in

---

240. Free Software Foundation, *Selling Free Software*, <http://www.fsf.org/licensing/essays/selling.html> (last visited Jan. 21, 2006).

241. The copyright owner to the original program does not have to comply with the GPL, so long as she does not accept any modifications from others, and thus can still charge fees for her software, e.g., under a dual licensing model, which may involve releasing software under the GPL free of charge (for marketing purposes) and for a license fee under a proprietary license (to licensees that would rather pay a fee than comply with the GPL for their own modifications). Also, an increasing number of companies have found other ways to generate revenue in the context of “free software,” e.g., by using “free software” as a means to sell proprietary software, services, or hardware. *See, e.g.*, John Koenig, *Seven Open Source Business Strategies for Competitive Advantage*, *IT MANAGER'S J.*, May 13, 2004, <http://www.itmanagersjournal.com/articles/314>; Jens Sieckmann, *Freie Software in der Wirtschaft*, *BRAVEHACK* § 9.2 (2001), <http://www.bravehack.de/html/node45.html> (last visited Jan. 21, 2006); Open Source Biotechnology Project, *Open Source as a Business Approach*, <http://rsss.anu.edu.au/~janeth/OSBusMod.html> (last visited Jan. 21, 2006).

context<sup>242</sup> with any GPLed code, even if the add-on programs are intended for combination with a particular version of GPLed code.<sup>243</sup> End users who run add-on programs with the GPLed code would not infringe because the GPL allows execution without any restrictions. Without direct infringement, there could not be any contributory liability. Yet, given implementation and integration difficulties, end users typically expect packages from their software suppliers, so separate distribution is often not a practical option.

Thus, an important question remains: What exactly is a “derived work”? A number of possibilities come to mind. “Derived work” could mean: (1) “derivative work,” as this term is defined by the Copyright Act; (2) “derivative work” or “collective work,” as these terms are defined by the Copyright Act; (3) “any combination of programs,” whether or not the combination qualifies as a “derivative work”; or (4) something else. What the term “derived work” means determines how much a licensee has to give up when it distributes GPLed code in combination with improve-

---

242. If a company offers copies of GPLed code and add-ons for such code to a customer, it would need a license to distribute the GPLed code—and such a license is granted per Section 2(b). In this respect, it should not make a difference whether the vendor delivers the GPLed program and the add-on product at different times or on different CDs.

243. This conclusion is also reached in what seems to be the first German commentary on the GPL, TILL JAEGER, ET AL., *DIE GPL KOMMENTIERT UND ERKLÄRT* 4, 6 (2005). However, in its FAQ, the Free Software Foundation takes the contrary position that distributors of add-on products for GPLed programs have to comply with the GPL merely because they make a product that is designed to interact with GPLed products (even though it does not contain any copied code). See Free Software Foundation, GNU Project, Frequently Asked Questions About the GNU GPL, <http://www.gnu.org/licenses/gpl-faq.html#GPLAndPlugins> (last visited Oct. 20, 2006) [hereinafter GPL FAQ]. This position could only be viable in reliance on a contributory liability theory (which can be ruled out, given the expressive broad grant of end user license rights under the GPL) or if—as a matter of copyright law—the distributor of an add-on product needed a distribution license, which is usually not the case, based on the view explained in this Article. See *supra* Part V. The opposite view could try to emphasize that the courts in *Micro Star v. Formgen*, 154 F.3d 1107, 1112 (9th Cir. 1998) and *Midway Manufacturing v. Artic International, Inc.*, 704 F.2d 1009, 1013-14 (7th Cir. 1983), did not expressly rely on a theory of contributory infringement and thus, arguably, must have negated the requirement that a derivative work must incorporate significant expression from the original work. However, such arguments would have to overcome the fact that (1) the *Micro Star* and *Midway* courts were addressing screen outputs and “story lines” with a high level of non-functional creativity that is unusual in the software context, and (2) the vast majority of references in legislative history, cases, and commentaries emphasize the definitional requirement that derivative works must actually incorporate copyrighted expression from the adapted work. See *supra* Section III.A and accompanying references.

ments or other programs—and whether a licensee may distribute GPLed code in combination with third party code at all.

## **B. GPL Terminology and Interpretation**

### *1. “Works Based on the Program”*

The first operative Section of the GPL (Section 0) reads as follows:

This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The “Program,” below, refers to any such program or work, and a “work based on the Program” means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term “modification.”) Each licensee is addressed as “you.”

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

As is common in commercial contracting practice, the first Section of the GPL contains a number of definitions and specifications that apply to the document as a whole. Less common, however, are the explanatory notes that the GPL drafters interwove with the legally binding definitions.

For example, the last sentence of the section quoted above acknowledges that the conditions in the half-sentence preceding it may not always be met in practice. The cause for this anomaly seems to lie in the genesis of the document: it was written by programmers for programmers. In order to make the document useful for non-lawyers and projects without a budget for legal advice, and to establish the GPL as a standard, the GPL drafters tried to draft the license as user-friendly and accessible to programmers as possible.

Along these lines, the second sentence of Section 0 defines “works based on the Program” as the Program itself or “any derivative work under copyright law” followed by an interpretive explanation, albeit one not entirely accurate, regarding what the term “derivative works” means under copyright law. This explanation, introduced with the phrase “that is to

say,” gives an indication of what the GPL drafters thought, hoped, or may argue in a dispute is the meaning of the term “derivative works” under copyright law. Section 2 of the GPL contains additional explanations and declarations of intent of what “works based on the Program” should mean, which even include references to “collective works,” i.e., a term defined by the Copyright Act in contrast to the term “derivative work.”<sup>244</sup> In order to resolve these text-internal contradictions, it would seem appropriate to rely on the “operative” portion of the definition in Section 0 (which contains the reference to the Copyright Act) and treat the “explanatory notes” as statements of opinion that have been added for convenience purposes only.<sup>245</sup> Accordingly, the GPL would be interpreted to define “work based on the Program” to mean “derivative work as defined by the Copyright Act.”<sup>246</sup>

## 2. *Derived Works*

The first sentence of Section 2 of the GPL permits modifications to the GPLed program in reference to the defined term “work based on the Program.” The following sentences of Section 2 contain a number of license conditions and explanations and use a number of other terms to describe the result of modifications besides “work based on the Program,” including “modified files,”<sup>247</sup> “modified program,”<sup>248</sup> and “modified work.”<sup>249</sup> The critical Subsection (b) refers to “any work . . . that in whole or in part contains or is derived from the Program or any part thereof.”

Taken out of context, each of these terms seems to go well beyond the statutory definition of derivative works in the Copyright Act, because the statutory definition is not satisfied by every modification, or by any work that contains any part of another work or that is derived from any part of another work. As discussed in Section IV.C, under the Copyright Act, a combination of code with a GPLed program constitutes a derivative work of the GPLed program only if the combination (i) is sufficiently permanent, (ii) contains significant and creative portions of the GPLed program, (iii) is creative in its own right, and (iv) involves significant and creative internal changes to the GPLed program.<sup>250</sup>

---

244. *See supra* Part III.

245. If the GPL drafters had wanted to apply their own definition, independent of the Copyright Act, they could have omitted the reference to the Copyright Act concept of derivative works or added it as an explanation of what the GPL definition would include.

246. Implicitly assumed by Stoltz, *supra* note 117, at 1457 n.18.

247. GPL License, *supra* note 227, § 2(a).

248. *Id.* § 2(c).

249. *Id.* § 2 after (c).

250. *See supra* Section IV.C.

In context, however, the drafters of the GPL appear to have randomly chosen substitutes to the somewhat awkward term “work based on the Program” and used the substitute terms synonymously to improve sentence flow and readability. This impression is confirmed throughout the document, which also uses other substitutes, including the “derivative or collective works based on the Program”<sup>251</sup> and “derivative works.”<sup>252</sup>

Some of the explanations throughout the GPL, as well as the Free Software Foundation’s FAQ<sup>253</sup> and “Lesser General Public License,”<sup>254</sup> imply that the drafters of the GPL intended to cover software combinations that would not qualify as derivative works under the Copyright Act according to the test developed in this Article.<sup>255</sup> These examples primarily evidence a difference of opinion in the application of copyright law and do not necessarily indicate that the condition in Section 2(b) of the GPL covers more than derivative works as defined by the Copyright Act. Yet un-

---

251. GPL License, *supra* note 227, § 2.

252. *Id.* § 5.

253. For the answer to the Question: “What is the difference between ‘mere aggregation’ and ‘combining two modules into one program?’”, see GPL FAQ, *supra* note 243.

254. Free Software Foundation, GNU Project, GNU Lesser General Public License, <http://www.gnu.org/copyleft/lesser.html> (last visited Jan. 21, 2006) [hereinafter LGPL License].

255. Most notably, the LGPL contains specific exceptions for dynamic linking of libraries, which—according to this article—would not qualify as derivative works and thus “works based on the Program” under the GPL regardless. On this topic, see also Brian W. Carver, *Share and Share Alike: Understanding and Enforcing Open Source and Free Software Licenses*, 20 BERKELEY TECH. L.J. 443, 459 (2005). The preamble of the LGPL explains the relationship between the two licenses as follows:

This license, the GNU Lesser General Public License, applies to certain designated libraries, and is quite different from the ordinary General Public License. We use this license for certain libraries in order to permit linking those libraries into non-free programs. When a program is linked with a library, whether statically or using a shared library, the combination of the two is legally speaking a combined work, a derivative of the original library. The ordinary General Public License therefore permits such linking only if the entire combination fits its criteria of freedom. The Lesser General Public License permits more lax criteria for linking other code with the library.

LGPL License, *supra* note 254. The preamble is based on the assumption that a combination of dynamically linked programs constitutes a derivative work of such programs, which is contrary to the findings of this Article. Yet, it is worth noting that even this paragraph of the LGPL acknowledges that both GPL and LGPL ultimately defer to the definition of “derivative work” under copyright law—and the explanations constitute merely explanations regarding the drafters’ understanding of what kinds of software combinations constitute “derivative works.”

certainties remain, given the fact that the “explanations” appear within the license text.

### 3. *Rules of Contract Interpretation*

Given the prevailing controversies and uncertainties regarding the exact scope of Section 2(b) of the GPL, it seems worth exploring whether presumptive rules of contract interpretation would favor one interpretation over another. In this context it is important to note that quite different rules could apply depending on the context of a particular licensing relationship. The GPL contains neither a contractual choice of law nor a forum selection clause. Under statutory and common law conflicts of law principles, which vary from jurisdiction to jurisdiction, the governing law of a licensing relationship subject to the GPL will be determined by the residency of the licensor and licensee, and various other factors. Thus, in practice, there is not one GPL that applies to all free software globally. Instead, thousands of different versions provide for slightly different rights and obligations of the licensing parties based on peculiarities of the governing contract law.<sup>256</sup>

Nevertheless, two principles of contract interpretation are likely to apply in most jurisdictions in one form or another: courts try to (i) determine the parties’ intent and (ii) interpret ambiguous clauses against the party who caused the ambiguity.

#### a) Parties’ Intent

Courts typically try to determine the intent of the contracting parties as objectively evident to each other at the time of contract formation.<sup>257</sup> Where the contract language is clear and unambiguous, courts will not usually look to extrinsic evidence of intent.<sup>258</sup> Given the uncertainties around Section 2(b) of the GPL, however, it seems likely that courts would feel tempted to look beyond the four corners of the document. Even though the GPL emphasizes that it constitutes a license as opposed to a

---

256. See, e.g., JAEGER, *supra* note 243. The Free Software Foundation acknowledges the risk of providing official translations of the GPL, see GPL FAQ, *supra* note 243, <http://www.gnu.org/licenses/gpl-faq.html#MereAggregation>), but the variations introduced by different applicable bodies of contract law seem far greater.

257. See, e.g., CAL. CIV. CODE § 1636 (West 1985); Bürgerliches Gesetzbuch [BGB] [Civil Code] Aug. 18, 1896, Reichsgesetzblatt [RFBl] 195, §§ 133, 157. *But see* Trident Ctr. v. Conn. Life Ins., 847 F.2d 564, 568 (9th Cir. 1988); Pac. Gas & Elec. Co. v. G.W. Thomas Drayage & Rigging Co., 442 P.2d 641, 645 (Cal. 1967).

258. See, e.g., CAL. CIV. CODE § 1639.

contract,<sup>259</sup> courts would likely apply contract interpretation rules and try to determine the intent of the copyright owner who selected the GPL and the licensee who selected the program. In many cases, courts will probably find that neither party really had a choice—the GPL came to apply because a developer of a previous program version had opted for the GPL. Where the Free Software Foundation itself is involved as a party, it may be appropriate to take the various examples, explanations, and programmatic and ideological statements on its website into consideration. Where the Free Software Foundation is not involved, however, it will often be difficult to confirm that parties to a dispute were familiar with these materials at the time the licensing relationship was formed.

b) Interpretation Against the Drafter

Another common principle of contract interpretation is that in case of uncertainties, courts should interpret the contract against the party who caused the uncertainty to exist.<sup>260</sup> This could help licensees in cases against the Free Software Foundation. In cases where neither party has selected the GPL to apply for a particular modified program, however, it is not clear that either party is to blame for the GPL's uncertainties.

4. Summary

The GPL permits end users to combine software and execute software in combination without any restrictions, even if the combinations constitute derivative works of the GPLed programs. As a consequence, contrary to the views expressed by the Free Software Foundation, distributors can separately distribute add-on products intended for combination with GPLed code without fear of incurring contributory liability. Thus, the distribution of add-ons to computer games discussed in Section V.C of this Article should be unproblematic in the GPL context, since the suppliers of the add-on products do not also sell the actual games.

The GPL strictly prohibits the distribution of program combinations that qualify as “derived works” of GPLed programs unless the entire combination can be subjected to the license terms of the GPL. This has two serious consequences for licensees. First, licensees are prohibited from distributing GPLed programs in combination with proprietary third party

---

259. Eben Moglen, *Enforcing the GNU GPL*, <http://www.gnu.org/philosophy/enforcing-gpl.html> (last visited Jan. 21, 2006) (“Licenses are not contracts: the work’s user is obliged to remain within the bounds of the license not because she voluntarily promised, but because she doesn’t have any right to act at all except as the license permits.”).

260. *See, e.g.*, CAL. CIV. CODE § 1654.

programs whose copyright owners do not agree to the GPL terms. Second, if licensees create “derived works” of GPLed programs, they cannot commercialize such “derived works” through proprietary license models as contemplated by the Copyright Act.

Despite remaining uncertainties, the context of the GPL favors an interpretation of the term “derived work” to mean a “derivative work as defined by the Copyright Act.” Consequently, the combinations discussed in Section V.B and V.D should also be permissible under the GPL because they do not involve a creation of derivative works. However, the language of the GPL also allows broader interpretations, and its drafters take the position—in documents that do not seem determinative for contract interpretation purposes—that dynamically linked programs fall under Section 2(b) if distributed in combination with GPLed code.

### C. First Sale Doctrine

The drafters of the GPL did not take the same precautions to avoid the First Sale Doctrine that drafters of commercial licenses do. To the contrary, the Free Software Foundation expressly states that the GPL allows licensees to sell copies.<sup>261</sup> As a result, it is conceivable that a company could try to circumvent Section 2(b) of the GPL by arranging to buy copies of GPLed code from a third party.<sup>262</sup> This would seem to allow the company to resell those copies, and its customers to execute them, without regard to the GPL,<sup>263</sup> even if the company also sells its own proprietary add-on programs to its customers simultaneously.<sup>264</sup>

### D. Statutory Limitations on License Restrictions

The same legal principles that can affect the validity of commercial licenses can also affect the validity of the GPL as it applies in a particular licensing relationship. A few particular considerations apply, however, given the somewhat unusual characteristics of the GPL.

---

261. GPL FAQ, *supra* note 243, <http://www.fsf.org/licensing/licenses/gpl-faq.html#DoesTheGPLAllowMoney> (last visited Jan. 21, 2006) (“Does the GPL allow me to sell copies of the program for money? Yes, the GPL allows everyone to do this. The right to sell copies is part of the definition of free software.”).

262. For instance, from a development company to a distribution company within a group of wholly owned subsidiaries.

263. 17 U.S.C. §§ 109, 117 (2000).

264. Without such a “first sale” scheme, the distribution company would need to obtain a license for the distribution right to the GPLed code, and such licenses would only be available under the GPL terms. Of course, this scheme would not change the situation for scenarios involving derivative works under the Copyright Act, because the adaptation right does not become exhausted through a first sale. Also, it should be noted that courts often try to sanction circumvention schemes under various legal theories.

### 1. *Competition Law*

As previously observed, the applicability and effects of competition law depend largely on the situation (i.e., on the affected markets and the parties' market power).<sup>265</sup> Thus, competition laws would probably play an insignificant role with respect to a relationship between two individual developers, but they could well come into play if a number of dominant suppliers<sup>266</sup> or purchasers<sup>267</sup> pushed to establish the GPL as a standard with the intent to drive "software only" companies from the market. Independent of such case-by-case considerations, however, one observation seems to apply generally: Section 2(b) of the GPL does not have the same anti-competitive effect that grant-back clauses typically have, because it does not require an exclusive license or assignment of ownership rights and the license is granted to anyone. Thus, Section 2(b) of the GPL does not result in a concentration of intellectual property rights or market power in the hands of one particular licensor.

### 2. *Copyright Misuse*

By imposing GPL § 2(b) on licensees, copyright owners try to magnify their rights beyond those sanctioned by the Copyright Act in two different ways. First, Section 103 of the Copyright Act allocates ownership rights to authorized derivative works to the author to incentivize further investment in additional creativity.<sup>268</sup> In contrast, Section 2(b) of the GPL requires creators of derivative works to forfeit their exclusion rights and any chance to generate licensing revenue.<sup>269</sup> Second, if the term "derived work" were found to encompass more than "derivative works" and included, for example, compilations and other forms of software combinations, Section 2(b) of the GPL would seek to prohibit activities that Section 106 of the Copyright Act has not reserved for copyright owners and thus exponentially increase the impact caused by the first copyright magnifying mechanism.<sup>270</sup>

---

265. *See supra* Section VI.C.1.

266. For example, technology conglomerates that have traditionally sold hardware and/or services in addition to software could decide to "give away" software in order to drive "software only"-companies from the market.

267. For example, government units or large purchasing cooperatives could unite to insist on GPLed code in order to drive software prices down or distort competition by favoring "solution providers" who can afford to "give away" software so long as they can raise hardware and/or services prices.

268. *See* 17 U.S.C. § 103.

269. GPL License, *supra* note 227, § 2.

270. 17 U.S.C. § 106.

Given the fact that copyright misuse is an equitable concept under U.S. law, it is difficult to predict if and how a court would apply this doctrine in the context of the GPL. On one hand, the non-profit status and idealistic goals pursued by the proponents and original adopters of the GPL may sway courts in favor of the GPL. On the other hand, the “copyleft” policy manifested in the GPL seems a more direct attack on the delicate balance between access and protection in the Copyright Act<sup>271</sup> than any other licensing practice that has so far caused courts to find copyright misuse.<sup>272</sup> In fact, the intended objective behind Section 2(b) of the GPL is to eliminate the effects of copyright protection for computer programs and generally replace it by the rules of the GPL.<sup>273</sup> This flies in the face of the many decisions by U.S. courts that found it necessary to protect the economic interests of software copyright owners who pursued proprietary licensing models.<sup>274</sup> Also, more and more companies use the GPL for purposes other than idealism. If courts enforce clauses like Section 2(b) of the GPL, they will probably also have to accept it if proprietary software companies start prohibiting combinations of their programs with other software beyond the boundaries of the Copyright Act. This could have potentially significant implications for interoperability.

Thus, for purposes of U.S. copyright law, Section 2(b) of the GPL seems to raise significant issues under the doctrine of copyright misuse, particularly if it were interpreted to cover more than derivative works as defined by the Copyright Act.

### 3. *Unfair Contract Terms*

As in the commercial licensing context, whether Section 2(b) of the GPL could be found invalid as an unreasonable or unconscionable contract term will depend on the situation of the particular parties to a licensing relationship. In general, unilateral software licenses easily meet most of the elements required for procedural unconscionability. In addition to licensees not having an opportunity to negotiate, if they refuse to accept the license terms they will also be expressly prevented from exercising the rights conveyed by the license.<sup>275</sup> The bargaining power differential de-

---

271. *See supra* Section IV.A.

272. Stoltz, *supra* note 117, at 1477 (referring to “copyleft” as a “hack on the copyright system” and discussing the regime established under the GPL as an alternative to the statutory regime established in the Copyright Act.)

273. *Id.* at 1444-46.

274. *See generally supra* Part IV.

275. *See, e.g.,* GPL License, *supra* note 227, § 5 (“You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to

depends on the exact circumstances of a particular licensing relationship. Consumers will typically not be affected by the GPL, because it allows combinations on the end user level without any restrictions.<sup>276</sup> As between businesses, on the other hand, courts might well find enough bargaining power differential to allow licensees to invoke doctrines such as contracts of adhesion or unconscionability under U.S. laws. Such situations might include those where large businesses offer code under the GPL, where smaller vendors are required to comply with the GPL by large software purchasers (such as government units), or simply where the terms of the GPL are generally non-negotiable due to the ideology behind it.

Substantively, however, it will often appear far-fetched for licensees to argue that Section 2(b) of the GPL is unreasonable as a commercial matter, given the fact that the licensee does not have to pay for its license.<sup>277</sup> Nevertheless, the “no charge” aspect would have the opposite effect where customers/licensees demand that suppliers/licensors use the GPL. Also, under less flexible contract laws in other jurisdictions, Section 2(b) of the GPL seems to run a substantial risk of invalidity merely because it derogates so sharply from statutory law to the disadvantage of the licensee.<sup>278</sup>

## E. Summary

Combinations of programs with GPLed software are somewhat dangerous for companies with proprietary license models if those companies also want to distribute the GPLed code itself. Distribution of separate add-on programs, however, should be permissible, even where the combination

---

modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License.”).

276. See *supra* Section VII.A.

277. But see Guadamuz, *Viral Contracts*, *supra* note 238, at 336.

278. See generally Guadamuz, *Viral Contracts*, *supra* note 238. LG München, 21 O 6123/04 (2004) expressed approval for Section 2 of the GPL in general, but this occurred in dictum and in reference to the unofficial, non-binding German translation of the GPL (even though the English, official version applied in the case at hand). See Case Note by Hoeren, 20 COMPUTER UND RECHT 776, 777 (2004). It is far from clear, however, how a German court would approach the “derived works” definition if and when confronted with a case actually raising the “viral/immunizing” effect of this clause. Based on general principles of German contracts law, it would seem more likely that a German court would invalidate Section 2(b) on the basis that it is surprising, overbroad, or ambiguous, or interpret “derived works” to mean “derivative work under German copyright law,” which would likely result in a relatively narrow application. Like U.S. law, German copyright law requires derivative works to incorporate significant expression from the adapted work and would not, for example, classify software combinations created through dynamic links as derivative works. See Spindler, *supra* note 177, at 51; Wuermeling & Deike, *supra* note 236, at 90.

results in the creation of a derivative work as defined by copyright law, because the GPL's broad EULAs stand in the way of contributory liability theories.

Another risk is that courts could interpret Section 2(b) of the GPL to extend not only to derivative works, as this term is defined under applicable copyright law, but also to compilations and other types of combinations that are not otherwise subject to the copyright owner's exclusive rights. Any court that interprets Section 2(b) so broadly would have to seriously consider a licensee's challenges and defenses under the doctrine of copyright misuse, competition law, or unfair contract terms laws. In light of these potential legal limitations and the context of the GPL, a narrower interpretation of Section 2(b), limited to derivative works as defined by copyright law, seems more appropriate and likely. Consequently, dynamic linking to GPLed programs would not normally trigger the application of the GPL to the linking program, even if both programs are distributed together.

#### **F. Recommendations for GPL Version 3<sup>279</sup>**

It seems unrealistic to ask of the Free Software Foundation to give up the "viral" and "immunizing" concepts of Section 2(b), according to which licensees have to make derivative works available under the GPL, because this concept is the centerpiece of the Free Software Foundation's strategy. Yet, going forward, the Free Software Foundation should consider clarifying in the actual GPL text that the various and varying references to "derived works" in the GPL are meant to refer to derivative works as defined by applicable copyright law.<sup>280</sup> This would go a long way towards minimizing the risk of diverging GPL standards around the world due to varied local laws, because, by "dynamically linking" to "applicable copyright law," the GPL would largely adopt the statutory boundaries set by copyright law in the various jurisdictions where code may be created or used.<sup>281</sup> This in turn should significantly reduce any risks of offending

---

279. This Article was completed just prior to the release of the draft GPL Version 3 text. See Robert Gomulkiewicz, *General Public License 3.0: Hacking the Free Software Movement's Constitution*, 42 HOUS. L. REV. 1015 (2005); Free Software Foundation, GPLv3 Draft, <http://gplv3.fsf.org/gpl-draft> (last visited Jan. 21, 2006).

280. The author acknowledges that the legal community still has to play its own part by reaching a clear consensus regarding the definitional scope of "derivative works" in the context of software combinations, and may not deserve any help with reducing FUD after applying copyright protection to software in the first place, at least not in the eyes of the programmer community.

281. Such a "dynamic link" to "applicable copyright law" would work well even in situations and jurisdictions where it is not yet clearly defined which types of software

competition law, unfair contract terms law, or doctrines similar to the copyright misuse doctrine in other jurisdictions.<sup>282</sup> The risks arising from unfair contract terms laws could be minimized further by adding a governing law clause referring to the laws of a jurisdiction that respects freedom of contract in general and severability clauses in particular.

## VIII. CONCLUSION

Software combinations are less dangerous liaisons than some have recently argued, particularly in the context of the GPL. Under the U.S. Copyright Act, a combination of a computer program with other software results in the preparation of a derivative work only if the combination (a) is sufficiently permanent, (b) contains significant and creative portions of the other software, (c) is creative in its own right, and (d) involves significant and creative internal changes to the other software. Most software combinations fail to meet one or more of these requirements and thus constitute compilations, collective works, or non-copyrightable arrangements, none of which implicate copyright owners' adaptation rights under Section 106 of the U.S. Copyright Act.

Software combinations involving dynamic links usually lack permanency, combination creativity, and internal changes. Even software combinations through static links do not necessarily affect adaptation rights, because such linking often results in the creation of a compilation or non-creative aggregation of programs or sub-programs. Nevertheless, under the U.S. Copyright Act, software developers typically have to obtain a license before they may combine programs through static linking because such linking affects the duplication rights of the linked program's copyright owner. Also, adaptation rights may be affected where software combinations, regardless of the code linking method, result in significant and creative changes to original screen output, such as in the context of computer games.

Under common commercial licensing conditions, end users typically receive an express or implied license to execute proprietary software in

---

combinations might infringe adaptation rights, because a court in such jurisdiction would decide the copyright law question at the same time that it would apply the GPL under such circumstances.

282. An express contractual choice of law would further support and strengthen uniformity and reliability of the GPL also with respect to other clauses, such as warranty disclaimers, and limitations of liability, in the interest of all licensors worldwide. Ironically, however, a selection of one jurisdiction's contract law, such as the laws of a state in the United States, would probably be perceived in the developer community as too local and not global enough.

combination with other software, regardless of whether the combination would qualify as a derivative work. Under the GPL, end users are free to combine GPLed code with any other code. Developers and distributors do not have to be concerned about contributory liability, so long as they distribute add-on software separately and the end users are not legally restricted in combining the intended programs with the add-on software.

Any person who wants to distribute programs in combination with and alongside GPLed code, however, will have to closely examine the reach and consequences of the various conditions and restrictions in the GPL. The term “derived work” in the GPL should be interpreted to mean “derivative works as defined by copyright law,” and as a consequence, most programs should be able to be distributed in combination with dynamically linked GPLed code without the necessity of subjecting the linking programs to the GPL.

It seems possible, however, that courts may interpret the GPL in a broader way, which would increase concerns regarding the validity of the GPL under copyright misuse doctrines, competition laws, and unfair contract terms laws. Such concerns can be greater or lesser depending on the circumstances of the licensing parties and jurisdictions involved. If such broad interpretations prevail, the software industry might move more generally to GPL-like restrictive licensing practices that permit and prohibit certain software combinations. Such interpretations and ensuing practices would potentially have a serious impact on interoperability and cause software combinations to become dangerous liaisons.

# STRIKING OUT “COMPETITIVE BALANCE” IN SPORTS, ANTITRUST, AND INTELLECTUAL PROPERTY

by Salil K. Mehra<sup>†</sup> and T. Joel Zuercher<sup>‡</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1500
II. THE COMPETITIVE BALANCE ARGUMENT.....	1502
A. SPORTS LEAGUES AND ANTITRUST: THE STATE OF PLAY.....	1502
B. COMPETITIVE BALANCE: A SPLIT AND A CURVE BALL .....	1505
1. <i>Courts of Law: Playing by Different Rules?</i> .....	1506
2. <i>Courting the Fans—Selling the Public on Competitive Balance</i> .....	1511
III. RECONSIDERING THE COMPETITIVE BALANCE ARGUMENT .....	1514
A. THE “COMPETING COMPETITIONS” CRITIQUE: GIVE IT SOME ENGLISH.....	1514
B. THE “TWO-SIDED MARKET” CRITIQUE: A BAD TRADE? .....	1524
1. <i>Monopsony</i> .....	1525
2. <i>Two-Sided Markets</i> .....	1527
C. THE “WEAK LINK” CRITIQUE: LETTING FANS CALL THEIR OWN SHOTS.....	1532
1. <i>Competitive Balance and Fan Interest</i> .....	1533
2. <i>Competitive Balance and League Policies</i> .....	1534
3. <i>Major League Baseball v. New Competing Competitions</i> .....	1537
IV. HOW TO DEAL WITH THE COMPETITIVE BALANCE ARGUMENT .....	1541
A. REJECT COMPETITIVE BALANCE UNDER THE RULE OF REASON.....	1541
B. EXERCISE SKEPTICISM ABOUT ARGUMENTS THAT MAY BE AESTHETICS .....	1542
C. RECONSIDER SIMILAR ARGUMENTS BEYOND ANTITRUST.....	1544

---

© 2006 Salil K. Mehra and T. Joel Zuercher

† Visiting Professor, Keio University School of Law, Tokyo, Japan. Associate Professor, Temple University, Beasley School of Law. Prof. Mehra would like to thank Sarah Beth Mehra for her loving support and editing skills.

‡ Associate, Pepper Hamilton, LLP.

V. CONCLUSION .....1545

I. INTRODUCTION

On November 21, 2000, with Congress in recess for Thanksgiving and an unresolved Presidential election, the Senate Judiciary Committee held a hearing in Washington, DC. While the voters might have expected a meeting at such a time, they might not have expected the topic: the lack of competitive balance in Major League Baseball (“MLB”).<sup>1</sup>

“Competitive balance” has been a focus of sports antitrust cases for three decades; now it appears not only in Senate hearings, but also has entered the lexicon of the sports pages.<sup>2</sup> This Article uses the term “competitive balance” as sports antitrust cases have generally done since it first appeared in a 1976 federal antitrust case:

Competitive balance means in essence that all of the league’s teams are of sufficiently comparable playing strength that . . . fans will be in enough doubt about the probable outcome of each game and of the various division races that they will be interested in watching the games, thus supporting the teams’ television and gate revenues.<sup>3</sup>

The problem is that, in light of new experience and economic research, competitive balance should be thrown out of the ballgame.

The “competitive balance argument” maintains that, because predictable outcomes will reduce fan interest and therefore profitability, professional sports leagues require special treatment under the antitrust laws that recognizes their “strong and unique interest in maintaining competitive balance.”<sup>4</sup> An antitrust exception, proponents argue, is necessary to create the on-the-field competition that draws fans. Under this theory, competi-

---

1. See Irving Molotsky, *Congress Puts Baseball Economics in Play*, N.Y. TIMES, Nov. 22, 2000, at D5.

2. Andrew Zimbalist, *BackTalk; Monopoly’s Money*, N.Y. TIMES, Nov. 6, 2005, at H11 (“Brian France, NASCAR’s chief executive, said competitive balance was his concern when he proposed limiting the number of cars team owners could control.”); Buster Olney & Steven Greenhouse, *With Labor Woes, Baseball Throws Fans a Brushback*, N.Y. TIMES, July 14, 2002, at A1 (“The owners are also intent on restoring some level of payroll parity among teams; by improving competitive balance, they say, small-revenue teams would become more attractive to fans.”).

3. *Smith v. Pro-Football*, 420 F. Supp. 738, 745 (D.C. Cir. 1976).

4. *Mackey v. Nat’l Football League*, 543 F.2d 606, 621 (8th Cir. 1976) (considering maintenance of competitive balance as a factor under rule of reason examination of joint restraint on player movement between teams).

tive balance helps a sports league compete with other forms of entertainment, including other sports leagues and “American Idol”-style reality television programs that may themselves provide “competing competitions.”

However, the competitive balance argument is fundamentally flawed in that it relies upon three imperfect assumptions: (1) there is and must be only one championship competition per sports league, (2) leagues can and will successfully engineer balance in that competition, and (3) fan interest is directly related to a championship structure. This Article shows, with comparative data and research from the newly burgeoning field of sports economics, that each of these assumptions is doubtful. One possible solution lies in reconceiving the “competing competitions” envisioned by the competitive balance argument. In particular, competing competitions need not only be *between* different leagues of teams. Rather, a sports league can itself incorporate several different competing competitions among its constituent teams and thus maintain fan interest even in the absence of competitive balance in that league.

A circuit split currently exists on whether competitive balance is relevant to antitrust analysis of professional sports leagues, with the overwhelming majority of courts saying that it is.<sup>5</sup> Those judges whom have endorsed the competitive balance argument have done so based on an outdated paradigm of monolithic competition, where winning the singular league championship is “the only thing.”<sup>6</sup> While this may accurately reflect historical experience, where World Series or Super Bowl victories have been the sole endgames, such lucrative enterprises as English soccer and “fantasy” sports leagues demonstrate that there can be more than one kind of competition that draws fans’ attention and money. Thus, continued acceptance of the competitive balance argument may represent an aesthetic judgment about what an attractive sports league looks like, but does so unsupported by empirical study.

This Article advocates the rejection of the competitive balance argument in antitrust and the recognition of the value of innovative “competing competitions” beyond antitrust into intellectual property.<sup>7</sup> Previous commentators have expressed doubts about the means-ends connection between particular sports league restraints and the competitive balance ar-

---

5. See *infra* notes 27-31 and accompanying text.

6. See Christopher Lehmann-Haupt, *Is Pro Football Bad for Us?*, N.Y. TIMES, Jan. 11, 1971, at 29 (quoting legendary coach Vince Lombardi as saying “Winning is not everything; it is the only thing”).

7. The basic notion is that competing competitions can generate fan interest in a way that antitrust and intellectual property should consider. See *infra* Part III.

gument.<sup>8</sup> In contrast, this Article makes a frontal attack on the argument itself by surveying antitrust case law on competitive balance in professional sports<sup>9</sup> and by applying new economic findings to the current sports “arena.”<sup>10</sup> Indeed, lawyers, legislatures, and public opinion all seem to lag behind economists’ increasingly prevalent doubts about competitive balance.<sup>11</sup>

This Article proceeds in three parts. Part II summarizes the current treatment of competitive balance in antitrust law and in public discussions of sports. Part III explains why the competitive balance argument fails on its face by drawing on comparative data from English soccer, by applying to sports leagues the economic theories of monopsony and two-sided markets,<sup>12</sup> and by analyzing MLB’s current dispute with “fantasy” leagues over intellectual property rights. Part IV proposes that the competitive balance justification be summarily rejected, explaining how this relates to areas beyond antitrust. A brief conclusion follows.

## II. THE COMPETITIVE BALANCE ARGUMENT

### A. Sports Leagues and Antitrust: The State of Play

No one wants to pay money to see one team appear without an opponent. Few want to pay money to see two teams bicker about what the rules

---

8. See Roger Noll, *Buyer Power and Economic Policy*, 72 ANTITRUST L.J. 589, 615-16 (2005) [hereinafter Noll, *Buying Power*] (observing that “virtually all economic studies of professional sports” reject the claim that collusion in the market for players as a means aids competitive balance as an end); Stephen F. Ross, *The Misunderstood Alliance Between Sports Fans, Players and the Antitrust Laws*, 1997 U. ILL. L. REV. 519 (1997) [hereinafter Ross, *Misunderstood Alliance*] (arguing for application of a “least restrictive means” test to filter out restraints lacking good ends-means connections to competitive balance); see also Gary R. Roberts, *The NCAA, Antitrust and Consumer Welfare*, 70 TUL. L. REV. 2631, 2667-68 n.69 (1996) (questioning, in the context of amateur athletics, whether a means-ends link exists between athletic association restrictions and competitive balance and observing an absence of empirical data at the time on that point).

9. See *infra* notes 25-42 and accompanying text.

10. See *infra* notes 111-145 and accompanying text.

11. See, e.g., DAVID BERRI, MARTIN SCHMIDT & STACEY BROOKS, *THE WAGES OF WINS* 63 (2006) (arguing that “baseball does not have a competitive balance problem”); Allen Sanderson & John Siegfried, *Thinking About Competitive Balance*, 4 J. SPORTS ECON. 255, 273 (2003) (questioning whether “competitive imbalance in baseball deserve[s] so much attention” given empirical data suggesting it is not a problem).

12. Two-sided markets are situations in which one or several platforms enable interactions between different classes of end-users, such as a TV station trying to attract both viewers and advertisers. The owner of the platform tries to court both sides of the market at prices that allow a profit. See *infra* Section III.B.2.

of the game ought to be.<sup>13</sup> Thus, individual teams who are competitors on the field of play—and who may be economic competitors—must nonetheless cooperate to ensure there will even be a product at all. As a result, sports leagues pose an inherent dilemma in antitrust.<sup>14</sup>

The joint restrictions teams put in place through their leagues go beyond simple matters of time, place, and manner of game play. The restrictions extend to competition over investment and inputs, such as the skilled labor of professional athletes.<sup>15</sup> To analyze such restrictions, courts have fit sports leagues into their existing antitrust framework. In *NCAA v. Board of Regents of the University of Oklahoma*,<sup>16</sup> the Supreme Court set the modern standard for judging restraints of trade among rival teams. There, agreements limiting the ability of NCAA member universities to negotiate and enter into their own television contracts appeared in isolation to be anti-competitive.<sup>17</sup> The Court noted, though, that competing sports teams may have pro-competitive justifications for such collusion, since they must agree on a host of issues for the product to exist at all.<sup>18</sup> Accordingly, the Court found sports industry regulations well-suited to

---

13. The first intercollegiate American football game took place between Harvard and Canada's McGill, rather than between Harvard and Yale, because the two American universities could not agree on the rules to use. See *Epilogue: McGill and the Birth of Football*, MCGILL NEWS ALUMNI Q., Summer 2005, <http://www.mcgill.ca/news/2005/summer/epilogue/> (last visited Dec. 5, 2006); see also Gary R. Roberts, *Sports Leagues and the Sherman Act: The Use and Abuse of Section 1 to Regulate Restraints on Intra-league Rivalry*, 32 UCLA L. REV. 219, 229 (1984) [hereinafter Roberts, *Sherman Act*]. Roberts writes:

Just as a single game inherently requires a complete integration of the two coproducing teams, the league product also requires complete integration of all the member clubs, none of which is by itself able to produce anything, and no two of which are able to produce a pennant race or a league champion. This total integration is not by the member clubs' choice; it is an absolute requirement.

*Id.*

14. See ROBERT H. BORK, *THE ANTITRUST PARADOX* 332 (1978) (claiming that “[a]ll league sports . . . rest entirely upon the right to boycott . . . [a]nd [w]ere the leagues denied the power to enforce such [arguably anticompetitive] agreements . . . [they] would be destroyed”); Roberts, *Sherman Act*, *supra* note 13, at 295 (“[T]here is a legitimate concern that the structure of a league, unlike that of other business organizations, may cause, albeit infrequently, individual club economic interest to be contrary to the interests of the league as a whole.”).

15. See BORK, *supra* note 14, at 332.

16. 468 U.S. 85 (1984).

17. *Id.* at 98-99.

18. *Id.* at 100-01.

rule of reason analysis rather than per se condemnation.<sup>19</sup> This result extended to sports leagues antitrust treatment familiar to other industries, but with a couple of notable exceptions.<sup>20</sup>

Baseball has an historical judge-made antitrust exception that has been narrowed by Congress and lower courts several times over the past dozen years.<sup>21</sup> Additionally, the non-statutory labor exemption from antitrust applies in sports as in other unionized industries, though the exemption only narrowly applies to restraints that affect labor and management and not to those that cause antitrust injury to third parties.<sup>22</sup> Despite these idiosyncrasies, courts continue to see sports-related antitrust litigation.<sup>23</sup>

---

19. *Id.* Rule of reason analysis involves judicial consideration of not just the nature of the conduct at issue, but also issues such as its effects and the context of its application. *See* *Bd. of Trade v. United States*, 246 U.S. 231, 238-40 (1918).

20. *See infra* notes 21-22.

21. Three different Supreme Court opinions concern baseball's antitrust exemption. *See* *Flood v. Kuhn*, 407 U.S. 258 (1972); *Toolson v. N.Y. Yankees*, 346 U.S. 356 (1953); *Fed. Baseball Club of Balt. v. Nat'l League*, 259 U.S. 200 (1922) (holding that baseball is exempt from the Sherman Act because it is "not a subject of commerce"). One might wonder whether sports leagues are, in fact, generally exempt from antitrust laws. The short answer is "no." The slightly longer answer is that they do play by somewhat different rules. Perhaps most famously, under a longstanding judicially-created blanket exemption, the antitrust laws did not apply to Major League Baseball; since a 1998 legislative repeal, the antitrust laws now apply to baseball's dealings with its player labor union. *See* Curt Flood Act of 1998, Pub. L. No. 105-297, 112 Stat. 2824 (codified as amended at 15 U.S.C. § 26b (2002)) (repealing the categorical exemption with respect to anticompetitive restraints involving players). Additionally, some lower courts have ruled that the exemption itself applies only to baseball's reserve clause limiting the mobility of its players between teams. *See* *Piazza v. Major League Baseball*, 831 F. Supp. 420, 438 (E.D. Pa. 1993) (concluding that the exemption does not apply to restraints involving rejection of investors seeking to relocate a team); *Butterworth v. Nat'l League of Prof'l Baseball Clubs*, 644 So. 2d 1021, 1025 (Fla. 1994) (concluding that the exemption does not apply to restraints against team relocation). There is some disagreement over the current importance of baseball's exemption. *Compare* Mitchell Nathanson, *The Irrelevance of Baseball's Antitrust Exemption: A Historical Review*, 58 RUTGERS L. REV. 1 (2005), with Thomas Ostertag, *Baseball's Antitrust Exemption: Its History and Continuing Importance*, 4 VA. SPORTS & ENTMT. L.J. 54 (2004). There have also been calls for repeal of baseball's antitrust exemption. *See, e.g.,* J. Philip Calabrese, *Antitrust and Baseball*, 36 HARV. J. ON LEGIS. 531 (1999) (suggesting the Curt Flood Act of 1998 did not go far enough toward repealing the antitrust exemption); Larry Smith, *Beyond Peanuts and Cracker Jack: The Implications of Lifting Baseball's Antitrust Exemption*, 67 U. COLO. L. REV. 113, 138 (1996) (arguing for limiting a complete repeal of the MLB antitrust exemption as necessary or terminating altogether the MLB exemption to protect the future health of MLB).

22. As its name suggests, the non-statutory exemption is judge-made law that exempts certain union-employer agreements from antitrust law. There is a statutory labor exemption under the Clayton Act and the Norris-LaGuardia Act essentially declaring that

## B. Competitive Balance: A Split and a Curve Ball

The competitive balance argument is the main pro-competitive justification that sports leagues offer to defend agreements otherwise prohibited by antitrust laws. The leagues have continually argued to courts of both law and public opinion that a lack of competitive balance over time would lead to predictable outcomes that would reduce fan interest and profitabil-

---

labor unions, despite their general tendency to agree to fix prices, are not “combinations or conspiracies in restraint of trade” that offend the antitrust laws; the exemption specifically immunizes certain union activities such as group boycotts from the antitrust laws. 15 U.S.C. § 17 (2006); 29 U.S.C. §§ 52, 104, 105, 113 (2006). The non-statutory exemption applies not to union activities but to union-employer agreements. *See Connell Co. v. Plumbers & Steamfitters*, 421 U.S. 616, 622 (1975) (“[T]he non-statutory exemption has its source in the strong labor policy favoring the association of employees to eliminate competition over wages and working conditions.”). The courts of appeals have applied the non-statutory labor exemption to professional sports leagues; the dominant rule exempts such agreements when they concern a mandatory subject of bargaining (such as wages, hours, and conditions of employment), when they are the subject of bona fide arm’s length agreement and when the alleged restraint on trade affects only the parties to the agreement seeking to be exempted. *Mackey v. Nat’l Football League*, 543 F.2d 606, 614 (8th Cir. 1976); *see Powell v. Nat’l Football League*, 930 F.2d 1293, 1297 (8th Cir. 1989); *McCourt v. Cal. Sports, Inc.*, 600 F.2d 1193, 1198 (6th Cir. 1979); *see also Powell v. Nat’l Football League*, 678 F. Supp. 777, 783-84 (D. Minn. 1988); *Bridgeman v. Nat’l Basketball Ass’n*, 675 F. Supp. 960, 964 (D.N.J. 1987); *Zimmerman v. Nat’l Basketball Ass’n*, 632 F. Supp. 398, 403-04 (D.D.C. 1986); *Wood v. Nat’l Basketball Ass’n*, 602 F. Supp. 525, 528 (S.D.N.Y. 1984). *But see Claret v. Nat’l Football League*, 369 F.3d 124 (2d Cir. 2004) (adopting a more open-ended standard without the bona fide bargaining requirement to uphold a minimum age requirement in the NFL player entry draft); *Case Comment, Antitrust Law—Nonstatutory Labor Exemption—Second Circuit Exempts NFL Eligibility Rules from Antitrust Scrutiny—Claret v. National Football League*, 118 HARV. L. REV. 1379 (2005) (endorsing *Claret* test). The general rule that emerges from the case law is that the non-statutory labor exemption applies specifically to union labor market transactions. *See Brown v. Pro Football, Inc.*, 50 F.3d 1041, 1051 (D.C. Cir. 1995) (observing that “the case for applying the exemption is strongest where a restraint on competition operates primarily in the labor market and has no anti-competitive effect on the product market”). It follows that if a players’ union were to decertify, or if a restraint were to affect other markets—as some important league restraints do—this exemption should not apply. *Id.*

23. Indeed, even start-up leagues in “sports” that are relatively new to sports cable channels are the subject of lawsuits. *See Ryan Nakashima, Seven Poker Players Sue WPT over Use of Likeness, Name, SAN JOSE MERCURY NEWS*, July 19, 2006 (reporting “suit accus[ing] WPT Enterprises of ‘price fixing’ and ‘group boycotts’ by colluding with 12 member casinos to prevent players from entering tournaments unless they forfeit their rights” to their own likenesses).

ity.<sup>24</sup> This argument has achieved significant traction in antitrust cases; its impact in the public and political consciousness may be even bigger.

1. *Courts of Law: Playing by Different Rules?*

Over three decades, the federal courts have many times dealt with the competitive balance argument across the gamut of American team sports.<sup>25</sup> While the total number of cases is not huge, such cases are not rare either.<sup>26</sup> Regardless, their influence most certainly goes beyond their mere number, since many cases involve issues that garner much media attention, such as franchise relocation, interleague competition, and labor unrest. The competitive balance argument has been used to justify restraints that include joint restrictions on the entry of new investors into the league,<sup>27</sup> on the geographic territories in which sports teams may conduct

---

24. See, e.g., *Silverman v. Major League Baseball*, 67 F.3d 1054, 1061 (2d Cir. 1994) (mentioning briefly that “[i]n antitrust litigation, the leagues perennially argue that some form of reserve system [limiting the movement of players] is necessary for competitive balance”).

25. The argument has hoary antecedents, its logic having been favorably received by the D.C. Circuit 86 years ago in the slightly different context of the *Federal Baseball* case. See *Nat’l League of Prof’l Baseball Clubs v. Fed. Baseball Club of Balt., Inc.*, 269 F. 681, 687 (D.C. Cir. 1920) (ruling baseball exempt from antitrust). The court observed that:

If the reserve clause did not exist, the highly skillful players would be absorbed by the more wealthy clubs, and thus some clubs in the league would so far outstrip others in playing ability that the contests between the superior and inferior clubs would be uninteresting, and the public would refuse to patronize them. By means of the reserve clause and provisions in the rules and regulations, said one witness, the clubs in the National and American Leagues are more evenly balanced, the contests between them are made attractive to the patrons of the game, and the success of the clubs more certain.

*Id.*

26. Since the District Court for the District of Columbia first discussed competitive balance in the context of a rule of reason antitrust case in 1976, the federal courts have discussed competitive balance in sports antitrust cases in 38 opinions, discussed *infra* at notes 27-30 and accompanying and following text. Additionally, the federal courts have also addressed the competitive balance argument in the context of antitrust claims involving amateur collegiate athletics. See *NCAA v. Bd. of Regents of the Univ. of Okla.*, 468 U.S. 85, 117 (1984); *Law v. NCAA*, 134 F.3d 1010 (10th Cir. 1998); *Bd. of Regents of the Univ. of Okla. v. NCAA*, 707 F.2d 1147 (10th Cir. 1983); *In re NCAA I-A Walk-On Football Players Lit.*, 2006 U.S. Dist. LEXIS 28824 (W.D. Wash. 2006); *Law v. NCAA*, 902 F. Supp. 1394 (D. Kan. 1995); *Bd. of Regents of Univ. of Okla. v. NCAA*, 546 F. Supp. 1276 (W.D. Okla. 1982).

27. See *Sullivan v. Nat’l Football League*, 34 F.3d 1091, 1113 (1st Cir. 1994).

operations,<sup>28</sup> on the entry of players into a sports league,<sup>29</sup> on the movement of current players between teams, on the terms of player employment and wages,<sup>30</sup> and on the televised broadcasts of games to fans.<sup>31</sup> Were it

---

28. See *L.A. Mem'l Coliseum Comm'n v. Nat'l Football League*, 468 F. Supp. 154, 167 (C.D. Cal. 1979).

29. *Smith v. Pro Football*, 593 F.2d 1173 (D.C. Cir. 1978) (considering competitive balance and joint restraints involved in draft system of new players); *Robertson v. Nat'l Basketball Ass'n.*, 389 F. Supp. 867, 892 (S.D.N.Y. 1975) (discussing competitive balance in the context of a challenge to the NBA draft).

30. See *Mackey v. Nat'l Football League*, 543 F.2d 606, 621 (8th Cir. 1976) (considering competitive balance and joint restraints on movement of players within a sports league); *McCourt v. Cal. Sports*, 600 F.2d 1193, 1215 (6th Cir. 1979) (mentioning competitive balance argument in passing in the context of challenge to National Hockey League restraints on player movement between teams); *Fraser v. Major League Soccer*, 7 F. Supp. 2d 73, 78 (D. Mass. 1998); *Silverman v. Major League Baseball*, 67 F.3d 1054, 1061 (2d Cir. 1994) (mentioning briefly that “[i]n antitrust litigation, the leagues perennially argue that some form of reserve system [limiting the movement of players] is necessary for competitive balance”); *Nat'l Basketball Ass'n v. Williams*, 857 F. Supp. 1069, 1079 (S.D.N.Y. 1994) (considering competitive balance argument); *White v. Nat'l Football League*, 836 F. Supp. 1458, 1487 (D. Minn. 1993); *White v. Nat'l Football League*, 822 F. Supp. 1389, 1408 (D. Minn. 1993); *Jackson v. Nat'l Football League*, 802 F. Supp. 226, 232 (D. Minn. 1992); *McNeil v. Nat'l Football League*, 1992 U.S. Dist. LEXIS 21561, at \*1 (D. Minn. 1992); *McNeil v. Nat'l Football League*, 790 F. Supp. 871, 876 (D. Minn. 1992); *Powell v. Nat'l Football League*, 690 F. Supp. 812, 818 (D. Minn. 1988); *Mackey v. Nat'l Football League*, 407 F. Supp. 1000, 1003 (D. Minn. 1975); *Phila. World Hockey Club v. Nat'l Hockey League*, 351 F. Supp. 462, 486 (E.D. Pa. 1972); see also *Brown v. Nat'l Football League*, 50 F.3d 1041, 1060 (D.C. Cir. 1995) (Wald, J., dissenting) (concluding that competitive balance should be a consideration under the rule of reason in judging restraints on player movement within a league); cf. *Finley v. Kuhn*, 569 F.2d 527, 539-41 (7th Cir. 1978) (discussing competitive balance in the context of commissioner veto of player transfers between teams, but applying the at-the-time categorical exemption of “the business of baseball” from the antitrust laws). *But see* *Brown v. Pro Football*, 812 F. Supp. 237, 239 (D.D.C. 1992) (concluding that the D.C. Circuit’s pre-NCAA holding in *Smith* rejecting the competitive balance argument was still good law because NCAA was inapplicable to professional sports leagues due to unique character of amateur college athletics); *Brown v. Pro Football*, 1992 U.S. Dist. LEXIS 2903, at \*31 (D.D.C. 1992) (“This Circuit has expressly found defendants’ professed ‘competitive balance’ and ‘better product’ purposes to be irrelevant to the rule of reason analysis.”). The majority in the D.C. Circuit’s *Brown* opinion did not reach this question.

31. See *Chi. Prof. Sports Ltd. P’ship v. Nat'l Basketball Ass'n (Chi. Prof'l Sports II)*, 95 F.3d 593, 604 (7th Cir. 1996) (Cudahy, J., concurring) (observing that competitive balance helps “ensure that the league provides high quality entertainment throughout the season so as to optimize competition with other forms of entertainment”); *Kingray v. Nat'l Basketball Ass'n*, 188 F. Supp. 2d 1177, 1185 (S.D. Cal. 2002) (mentioning possible relevance of competitive balance argument in the context of league restrictions on broadcasting); see also *U.S. Football League v. Nat'l Football League*, 634 F. Supp. 1155, 1162 (S.D.N.Y. 1986).

not for purported ancillary benefits, these restraints might be considered per se antitrust violations.

Though the Supreme Court has not directly considered how antitrust should treat competitive balance in professional sports, the Court has given at least tacit approval to the competitive balance theory in the *NCAA* case discussed above.<sup>32</sup> In addition to considering the applicability of the rule of reason, the Court also considered whether NCAA rules that limited the frequency of televised college football games violated the Sherman Act. Although the Court condemned the NCAA restraints, it agreed generally that “the interest in maintaining a competitive balance among amateur athletic teams is legitimate and important.”<sup>33</sup> Particularly, the Court endorsed “justifiable means of fostering competition among amateur athletic teams,” concluding that such means of enhancing competitive balance would be pro-competitive because they would “enhance public interest in intercollegiate athletics.”<sup>34</sup>

Not surprisingly, professional sports leagues have sought cover under the competitive balance argument for their own practices.<sup>35</sup> However, the *NCAA* case concerned amateur college athletics; the circuits are split as to whether competitive balance in professional sports serves as an appropriate consideration under a rule of reason analysis.<sup>36</sup> Of the three Courts of

---

32. *NCAA v. Bd. of Regents of the Univ. of Okla.*, 468 U.S. 85 (1984).

33. *Id.* at 117. The Court concluded that goal of competitive balance was not causally linked with the restraints at issue. Lower courts have at times suggested that the Supreme Court’s logic in *Continental T.V., Inc. v. GTE Sylvania*, 433 U.S. 36 (1977), authorizing tradeoffs between inter-brand and intra-brand competition might apply. *See Sullivan v. Nat’l Football League*, 34 F.3d 1091, 1112 (1st Cir. 1994) (citing *Sylvania*, 433 U.S. 36 (1977)). However, *Sylvania* differed in important respects, since both harms and benefits accrued in the same market (televisions) and the *Sylvania* brand comprised a very small portion of the overall market. The First Circuit in *Sullivan* appeared to recognize this, requiring fact-finding on remand of benefits in a separate market, to the extent that they fed back into the harmed market. *Sullivan*, 34 F.3d at 1113; *see also infra* notes 117-123 and accompanying text.

34. *NCAA*, 468 U.S. at 117.

35. *See Silverman v. Major League Baseball*, 67 F.3d 1054, 1061 (2d Cir. 1995) (observing that “[i]n antitrust litigation, the leagues perennially argue that some form of reserve system [limiting the movement of players] is necessary for competitive balance”).

36. *Compare Smith v. Pro Football*, 593 F.2d 1173, 1186 (D.C. Cir. 1978) (rejecting competitive balance as offsetting pro-competitive benefit), *with Sullivan*, 34 F.3d 1091 (1994) (remanding for consideration of pro-competitive effects of competitive balance as offsetting anti-competitive effects of joint restraints by member teams on public ownership of teams), *and Mackey v. Nat’l Football League*, 543 F.2d 606, 621 (8th Cir. 1976) (recognizing that “the NFL has a strong and unique interest in maintaining competitive balance among its teams”); *see also Chi. Prof’l Sports II*, 95 F.3d at 604 (Cudahy, J.,

Appeals that have directly addressed whether competitive balance is a pro-competitive benefit or an anti-competitive restraint, only the D.C. Circuit has rejected the argument as anti-competitive. Specifically, in *Smith v. Pro Football*, in considering whether a league-wide player-entry draft was anti-competitive, the D.C. Circuit held that competitive balance, even if it “produc[es] better entertainment for the public, higher salaries for the players overall, and increased financial security for the clubs,” cannot be used to offset anti-competitive effects in the more narrow market for players entering the league.<sup>37</sup> While one might see this ban on “intermarket” tradeoffs as a relatively narrow ruling, it is largely fatal to the competitive balance argument. While leagues attribute restraints to an alleged need to generate and maintain fan interest, most antitrust cases seldom concern the market for fan interest. Rather, most cases concern the market for players, ownership interests, and broadcast rights.<sup>38</sup>

Since *Smith*, though, no other Circuit has held maintenance of competitive balance to be an impermissible consideration under the rule of reason.<sup>39</sup> Rather, subsequent Courts of Appeals have recognized that a professional sport league has “a strong and unique interest in maintaining competitive balance among its teams.”<sup>40</sup> Judges have recognized that competitive balance “is needed to ensure that” professional sports leagues “provide high quality entertainment throughout the season so as to optimize competition with other forms of entertainment.”<sup>41</sup> As a result, courts have concluded that the leagues’ “interest in maintaining competitive balance” should be weighed as an offsetting pro-competitive justification for

---

concurring) (suggesting that “competitive balance . . . is needed” but “is not the only contributor to the entertainment value of NBA basketball”).

37. *Smith*, 593 F.2d at 1186. Although *Smith* predates *NCAA*, it has continued to be followed in the D.C. Circuit after *NCAA* due to the amateur-professional distinction—particularly, the NCAA’s interest in avoiding professionalism and maintaining an image of scholasticism and amateurism. See *Brown v. Pro Football*, 812 F. Supp. 237 (D.D.C. 1992); *Brown v. Pro Football*, 1992 U.S. Dist. LEXIS 2903 (D.D.C. 1992).

38. The split is essentially a sports-specific version of the broader question of whether pro-competitive benefits that offset anti-competitive harms must accrue to the same relevant market under antitrust analysis. See *infra* notes 117-123 and accompanying text.

39. Courts have, however, often concluded that the particular restraints at issue before them are only tenuously linked with the asserted goal of competitive balance. See *infra* note 43.

40. *Mackey*, 543 F.2d at 621 (considering maintenance of competitive balance as a factor under rule of reason examination of joint restraint on player movement between teams).

41. See *Chi. Prof'l Sports II*, 95 F.3d at 604 (Cudahy, J., concurring).

otherwise anti-competitive restraints.<sup>42</sup> While the competitive balance argument is not a trump card that bests all plaintiffs,<sup>43</sup> it does provide professional sports leagues with a unique antitrust defense not available to other industries.<sup>44</sup>

Given this largely favorable reception, professional sports leagues continue to promote the competitive balance argument.<sup>45</sup> This seems logical, since courts have rejected the leagues' alternative justifications for their seemingly anti-competitive restraints. For example, courts have refused to allow leagues to justify player restrictions on the grounds that requiring teams to compete with each other for player services would lead to "ruinous competition," or that a team needs to have monopsony power over a player for a certain amount of time in order to recoup its investment in developing the player's skills.<sup>46</sup> Some commentators have argued that horizontal agreements among competing teams in a sports league should receive completely unique treatment based on the proposition that the law should treat a league as a "single entity," rather than as a group of bargaining competitors.<sup>47</sup> Such treatment would shift the analysis from relatively

---

42. See *Sullivan v. Nat'l Football League*, 34 F.3d 1091, 1111, 1113 (1st Cir. 1994); see also *Mackey*, 543 F.2d at 621.

43. Some courts that have recognized the leagues' interest in competitive balance have nevertheless found the restraints at issue to be inappropriate means of achieving the asserted end. See, e.g., *Mackey*, 543 F.2d at 622 (stating that the court "need not decide whether a system of inter-team compensation for free agents moving to other teams is essential to the maintenance of competitive balance in the NFL" because the rule at issue was "significantly more restrictive than necessary to serve any legitimate purposes"); cf. *NCAA v. Bd. of Regents of the Univ. of Okla.*, 468 U.S. 85, 117 (1984) (concluding that broadcast restrictions did not have a good means-ends link with maintaining competitive balance).

44. See Ross, *Misunderstood Alliance*, *supra* note 8, at 542, 544 (claiming that maintaining "competitive balance" as opposed to rejected defenses, such as the need to avoid "ruinous competition" or "recoup" investment in new players, is a "peculiar" "specialized" need for professional sports).

45. See also *infra* Section II.B.2.

46. Ross, *Misunderstood Alliance*, *supra* note 8, at 538; see generally *U.S. v. Soco-Vacuum Oil Co.*, 310 U.S. 150, 221 (1940) (holding that fear of ruinous competition cannot be a justification for price fixing); *Mackey*, 543 F.2d at 621 (concluding that there is nothing unique about sports leagues to justify such defenses).

47. See Myron C. Grauer, *Recognition of the National Football League as a Single Entity Under Section 1 of the Sherman Act: Implications of the Consumer Welfare Model*, 82 MICH. L. REV. 1 (1983); Gary R. Roberts, *The Single Entity Status of Sports Leagues Under Section 1 of the Sherman Act: An Alternative View*, 60 TUL. L. REV. 562 (1986); John C. Weistart, *League Control of Market Opportunities: A Perspective on Competition and Cooperation in the Sports Industry*, 1984 DUKE L. J. 1013 (1984). But see Lee Goldman, *Sports, Antitrust and the Single Entity Theory*, 63 TUL. L. REV. 751, 761 (1989) (concluding that the single entity theory is "ultimately unpersuasive").

strict prohibitions on horizontal restraints of trade<sup>48</sup> to more ambiguous restrictions of monopolization.<sup>49</sup> Newer sports leagues, such as Major League Soccer, have attempted to structure themselves as single corporate entities with the constituent teams as subunits, in order to garner the relatively favorable antitrust treatment in the manner that the single-entity argument proposes.<sup>50</sup> However, the single entity argument has been largely rejected by courts.<sup>51</sup>

In sum, courts have generally accepted the competitive balance argument’s basic premise that leagues must restrain economically competing teams in order to promote league interests as a whole. As the next Section will describe, sports leagues tout the argument not just in courts of law, but in the court of public opinion.

## 2. *Courting the Fans—Selling the Public on Competitive Balance*

American professional sports leagues have actively promoted public awareness of competitive balance. NFL, NHL, and MLB officials have all stressed the need for competitive balance in their respective leagues.<sup>52</sup>

---

48. 15 U.S.C. § 1 (2006).

49. 15 U.S.C. § 2 (2006).

50. *See* *Fraser v. Major League Soccer*, 284 F.3d 47, 58 (1st Cir. 2002) (calling it “doubtful” whether Major League Soccer—organized with intent to be a single entity rather than a league of separate corporate and investment entities—possessed the unity of interests required for single entity treatment under antitrust precedent).

51. *Sullivan v. Nat’l Football League*, 34 F.3d 1091, 1099 (1st Cir. 1994) (rejecting single entity treatment for league of separate corporate and investment entities); *L.A. Mem’l Coliseum Comm’n v. Nat’l Football League*, 726 F.2d 1381, 1388-90 (9th Cir. 1984) (same); *N. Am. Soccer League v. Nat’l Football League*, 670 F.2d 1249, 1256-58 (2d Cir. 1982) (same); *Smith v. Pro Football, Inc.*, 593 F.2d 1173, 1179-82 (D.C. Cir. 1978) (same); *Mackey v. Nat’l Football League*, 543 F.2d 606, 620 (8th Cir. 1976) (same); *see* Pepper Brill, Note, *Major League Soccer or Major League Sham? Players Bring Suit to Bite the Hand that Feeds Them*, 1999 COLUM. BUS. L. REV. 585, 610 (1999) (observing that “the courts have repeatedly rejected the single entity defense with respect to sports leagues.”). *But see* *Nat’l Football League v. N. Am. Soccer League*, 459 U.S. 1074 (1982) (Rehnquist, J., dissenting) (endorsing single entity theory).

52. *See, e.g.*, Judy Battista, *Distant Rumbles Heard in N.F.L.’s Labor Peace*, N.Y. TIMES, Jan. 25, 2006, at D3 (quoting NFL commissioner as supporting cost controls and revenue sharing to preserve “competitive balance” described as “the bedrock of the NFL’s success”); Murray Chass, *On Any Given Day: Parity, or at Least Something Like It, Has Arrived*, N.Y. TIMES, July 18, 2004, at H2 (quoting MLB commissioner as citing “competitive balance” as logical outcome of revenue sharing and salary cost controls); Joe Lapointe, *Lockout is First Shot in Hockey’s Labor War*, N.Y. TIMES, Sept. 16, 2004, at D1 (quoting NHL commissioner as citing “competitive balance” as a reason to try to control costs by capping players salaries, despite the fact that in the previous 3 seasons 12 different teams—the maximum number possible—had played in the league’s championship semifinals).

Baseball has perhaps been the most active in introducing the phrase into public discussion, having charged its Commissioner in 2000 to protect the “integrity of . . . the national game of Baseball” and the public perception that “there is an appropriate level of long-term competitive balance among [the] clubs.”<sup>53</sup> In this regard, MLB also formed a “Blue Ribbon Panel” of independent experts to study whether baseball has a “competitive balance” problem and to recommend possible reforms.<sup>54</sup> The “Panel Report” suggested a number of reforms that would, for the most part, assist teams with smaller revenues. These suggestions include “taxing” excessively high payrolls, allowing teams to relocate to cities where they could garner higher revenues, and subjecting foreign players to a player-entry draft rather than allowing them to negotiate as free agents.<sup>55</sup> Of course, reforms such as these would tend to aid lower-revenue teams at the expense of players and fans,<sup>56</sup> perhaps requiring cutbacks in player salaries and enabling teams to leave fans in their home cities for greener pastures. Consequently, reviewers of the Panel Report have disagreed about its overall value.<sup>57</sup>

---

53. MAJOR LEAGUE BASEBALL CONST., art. II, § 4; see Stephen F. Ross, *Light, Less-Filling, It's Blue Ribbon*, 23 CARDOZO L. REV. 1675, 1675-76 (2002) [hereinafter Ross, *Light, Less-Filling*] (reviewing report of Commissioner's committee on competitive balance).

54. Richard C. Levin et al., *The Report of the Independent Members of the Commissioner's Blue Ribbon Panel on Baseball Economics July 2000*, [http://www.mlb.com/mlb/downloads/blue\\_ribbon.pdf](http://www.mlb.com/mlb/downloads/blue_ribbon.pdf) [hereinafter Panel Report].

55. *Id.*

56. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 32-33 (noting composition of Blue Ribbon Panel and suggesting that it was biased towards owners in small markets as opposed to other interests); Ross, *Light, Less-Filling*, *supra* note 53, at 1688 (asking “[w]hy are almost all solutions[proposed by the Panel] at the expense of non-owner constituencies?”).

57. Compare Ross, *Light, Less-Filling*, *supra* note 53, at 1690 (stating that “[o]n balance . . . the Report remains a positive contribution”), with BERRI, SCHMIDT & BROOKS, *supra* note 11, at 32, 36-50 (questioning whether “this was truly a panel worthy of a blue ribbon” and doubting whether MLB actually has a competitive balance problem and whether the Panel's proposals are worthwhile). See also ANDREW ZIMBALIST, *MAY THE BEST TEAM WIN: BASEBALL ECONOMICS AND PUBLIC POLICY* 50 (2003) (approving of proposals redistributing revenue from richer teams in larger cities to poorer ones in smaller cities, but proposing a team minimum salary level to prevent owners from just pocketing redistributive payments).

Beyond its proposals, several of which were enacted,<sup>58</sup> the Panel Report also served to focus public discussion on MLB’s lack of competitive balance, linking baseball’s problems particularly to the issue of controlling player salaries. As a result, both the public and its elected officials have come to view MLB’s problems through a lens favorable to the league’s team owners. Other professional sports leagues, including the NHL and the NFL, have taken note of and have followed MLB’s lead in making similar arguments.<sup>59</sup> Similarly, the sports media have increasingly cited a lack of competitive balance in sports as a serious problem that league policy must address.<sup>60</sup> As a result, Congressional committees have held hearings to generate possible solutions to the alleged “problem” of lack of competitive balance. One outcome of these hearings has been to suggest a solution of which has been considered for decades: the legislative grant of favorable antitrust treatment for professional sports leagues.<sup>61</sup>

In general, public discussion has not focused on antitrust law. Unfortunately, it has not particularly focused on data, either. Thus, as in the courts of law, the existing state of play in the court of public opinion may be due for a reappraisal. As the next Part describes, the recognition the competitive balance argument has received may well be the legal equivalent of an unearned run.<sup>62</sup>

---

58. An August 2002 agreement between the baseball team owners and the players’ union included the adoption of revenue sharing and a “luxury” or “competitive balance” tax on teams with extremely high payrolls. *See* Lawrence Ritter, *Show Me the Money*, N.Y. TIMES, May 25, 2003, at G7.

59. *See* Lapointe, *supra* note 52, at D1; Battista, *supra* note 52, at D3.

60. For example, a search of The New York Times’ online database shows 38 articles discussing competitive balance and Major League Baseball in the almost 7 years since January 1, 2000, but only 29 such articles in the 14 years ending on December 31, 1999.

61. Congress has, at various times over decades, considered the extension of baseball’s exemption to other sports. For example, in 1951 four separate bills were introduced to exempt organized professional sports from the antitrust laws. None of them was enacted. *See* H.R. 4229, H.R. 4230, H.R. 4231, & S. 1526, 82d Cong. (1st Sess. 1951). More recently, bills have been introduced, but failed to pass, that would provide additional exemptions to other leagues. *See, e.g.*, Kenneth Silverstein, *Bill May Halt NFL’s Sack of Cities*, AM. CITY & COUNTY, Apr. 1, 1996, available at [http://www.americancityandcounty.com/mag/government\\_bill\\_may\\_halt/index.html](http://www.americancityandcounty.com/mag/government_bill_may_halt/index.html) (describing bill introduced by Senator John Glenn that would exempt NFL from antitrust liability for blocking franchise relocations).

62. In other words, the recognition competitive balance has received by courts is the product of “field errors” (such as misguided public opinion and misinformed legislative response) rather than the product of hard-hitting judicial analysis.

### III. RECONSIDERING THE COMPETITIVE BALANCE ARGUMENT

Judges have joined the media and Congress in looking upon the asserted link between increased competitive balance and increased fan interest with a relatively uncritical eye. Courts have endorsed the idea that if an anti-competitive restraint can be linked to competitive balance, the pro-competitive payoff of making the league more attractive to fans may outweigh the restraint's negative effects.<sup>63</sup>

However, recent innovations throw into question the idea that leagues need competitive balance to attract fans. Newer models of competition demonstrate alternatives that do not require focus upon a single league championship. In particular, examples from English soccer and from American baseball and basketball suggest that the time has come to re-evaluate the competitive balance argument. Because such examples are attractive but may fail to persuade due to their essentially anecdotal nature, this Part also provides support through empirical economic research, much of which is fairly recent.

Thus, this Part proceeds with three different critiques that challenge the need for competitive balance. First, competing competitions within a league can maintain fan interest, notwithstanding a lack of competitive balance for purposes of a traditional, singular league championship. Second, a study of two-sided markets and monopsony, or buyer market power, reveals rejections of the asserted welfare tradeoff between fans, players, and leagues implicit in the competitive balance argument. Third, antitrust should consider the possibility that maintaining room for alternative competitions, including fan innovation, may be welfare-enhancing overall.

#### A. The “Competing Competitions” Critique: Give It Some English

In considering the competitive balance argument, American courts implicitly assume that only a singular league championship will take place. Courts should not be faulted for assuming this must be the case, for each of the major American professional sports leagues currently holds a singular championship competition. However, English soccer<sup>64</sup> shows that such an arrangement does not have to be the case. American courts should not

---

63. See Mackey, 543 F.2d at 621.

64. This Article uses the word “soccer” for what the English would call football, unless as part of a proper name (i.e., the “Football League” of England).

allow the picture of a single championship competition to hold them captive when comparative data shows an alternative.

The English Premier League provides a meaningful comparison with North American professional sports leagues for several reasons. For one, it is among only a few leagues outside of North America that are truly on similar financial footing, in terms of both club valuations and revenue. According to media-published estimated valuations, the top Premier League clubs have similar valuations as the top teams in the NFL and MLB, and higher valuations than top NBA franchises<sup>65</sup> (see Table 1). Second, the Premier League and the North American leagues are considered global position leaders in their respective sports. Here, the Premier League is a particularly interesting comparison because although many of its teams are quite old, it only established itself as an independent entity in its current form in 1992 (partly in response to declining interest in English soccer during the 1980s).<sup>66</sup> As a result, the Premier League represents an innovation in league structure in a century-old professional sport.

---

65. See Roger G. Noll, *The Economics of Promotion and Relegation in Sports Leagues*, 3 J. SPORTS ECON. 179 (2002) [hereinafter Noll, *Promotion and Relegation*] (stating that the Premier League is roughly comparable to MLB and the NBA, and behind only the NFL, in profits per team and average franchise value).

66. See Premierleague.com, *The History of the F.A. Premier League*, <http://www.premierleague.com/fapl.rac?command=forwardOnly&nextPage=enHistory&categoryCode=History> (last visited Oct. 23, 2006) (describing Premier League's formation at a time when "The English game was at possibly its lowest ebb" and explaining how the new league used satellite television and new "competition in Europe" to prosper). In part, the Premier League's creation represented independence and commercial rebranding, rather than all-out restructuring, of what used to be called the "First Division" of an umbrella organization in England called the Football League. The Football League comprised (and continues to comprise) several different levels of soccer leagues. However, this move occurred in the same year that a preexisting pan-European tournament was reconfigured to match the top several teams from a variety of nations' top soccer leagues. Thus, together, these changes altered the form of competition at the top level of English soccer.

**Table 1<sup>67</sup>**  
**Highest-Valued Teams by League**

Premier League	MLB
1. Manchester United—\$1.37B	1. New York Yankees—\$1.03B
2. Arsenal—\$841M	2. Boston Red Sox—\$617M
3. Chelsea—\$508M	3. New York Mets—\$604M
4. Liverpool—\$370M	4. Los Angeles Dodgers—\$482M
NBA	NFL
1. New York Knicks—\$543M	1. Washington Redskins—\$1.3B
2. Los Angeles Lakers—\$529M	2. Dallas Cowboys—\$1.1B
3. Houston Rockets—\$422M	3. New England Patriots—\$1B
4. Chicago Bulls—\$409M	4. Philadelphia Eagles—\$952M

According to *Forbes*, Manchester United of the Premier League is the most valuable soccer club in the Premier League (and the world) with a valuation assessed at \$1.37 billion. As demonstrated by Table 1, that value compares favorably with any of the North American league franchises. The Washington Redskins are the most valuable NFL club at \$1.3 billion, the New York Yankees lead MLB with a valuation of \$1.03 billion, and the New York Knicks lead the NBA at \$543 million.<sup>68</sup>

The Premier League also generates revenue comparable to North American sports leagues. In 2004, the Premier League generated total revenue estimated at \$2.4 billion,<sup>69</sup> which places it in the same ballpark as

67. *Forbes.com*, Special Report: The Most Valuable Soccer Teams, <http://www.forbes.com/lists/> (follow "Richest Soccer Teams" hyperlink) (last visited Apr. 24, 2006).

68. *Id.*

69. DAN JONES ET AL., DELOITTE & TOUCHE, 2005 ANNUAL REVIEW OF FOOTBALL FINANCE: A CHANGING LANDSCAPE 26, Chart 2.3 (2005) [hereinafter 2005 ANNUAL REVIEW] (estimating total aggregate revenues for Premier League teams rose from £464 million in 1996-97 to £1.326 billion in 2003-04). The latter figure was converted to dollars at the average 2004 exchange rate of 1.83 dollars per pound. Lawrence H. Officer,

the NFL (estimated at \$5.3 billion),<sup>70</sup> MLB (\$4.5 billion),<sup>71</sup> the NBA (\$3 billion),<sup>72</sup> and the NHL (\$2.2 billion).<sup>73</sup> Additional facts are also worth noting. First, there are only 20 teams generating the Premier League’s revenue while no North American sports league has fewer than 30 teams. Second, although the Premier League, like some American sports, has global appeal, the United Kingdom, its primary market, has one-fifth the population of the United States and a 25 percent lower Gross Domestic Product per capita.<sup>74</sup> Third, the Premier League’s aggregate revenues nearly tripled from 1997 to 2004.

This revenue growth has in part been the result of the Premier League solving a chicken-and-egg problem between attracting the best players and attracting more fans. The Premier League profited from the explosion of cable and satellite television over the last decade, leading to huge increases in foreign and domestic broadcast revenue. Those revenues have allowed clubs to pay more for top international talent, which in turn has further driven demand for the Premier League on television.<sup>75</sup> These changes make the Premier League one of a handful of soccer leagues, along with those of Spain and Italy, that are truly international entertainment products.<sup>76</sup> In North America, the NBA and MLB have enjoyed an influx of international talent in the last decade—the NBA from Europe and South America, and MLB from Asia and Latin America—transforming once exclusively North American leagues into global show-

---

*Exchange rate between the United States dollar and the British pound, 1791-2004*, ECONOMIC HISTORY SERVICES, <http://eh.net/hmit/exchangerates/pound.php> (last visited Aug. 11, 2006).

70. Michael K. Ozanian, *The Business of Football*, FORBES.COM, Jan. 27, 2005, <http://www.forbes.com/lists/2004/09/01/04nfland.html> (estimating 2004 NFL revenue at \$5.3 billion).

71. Michael K. Ozanian, *The Business of Baseball*, FORBES.COM, <http://www.forbes.com/business/2005/04/06/05mlbland.html> (last visited Dec. 13, 2006) (follow “revenue” hyperlink) (estimating 2004 MLB revenue at \$4.5 billion).

72. *Necessary Roughness*, FORBES, Dec. 27, 2004, at 131, available at [http://www.forbes.com/free\\_forbes/2004/1227/131.html](http://www.forbes.com/free_forbes/2004/1227/131.html) (estimating NBA revenue at \$3 billion).

73. Michael K. Ozanian, *Ice Capades*, FORBES, Nov. 29, 2004, at 124, available at [http://www.forbes.com/forbes/2004/1129/124\\_print.html](http://www.forbes.com/forbes/2004/1129/124_print.html) (estimating 2003-04 NHL revenue at \$2.2 billion).

74. *Compare* CIA, World Factbook, United Kingdom, <https://www.cia.gov/cia/publications/factbook/geos/uk.html> (last visited Aug. 5, 2006) (listing population at 60.3 million and per capita GDP at \$30,300), with CIA, World Factbook, United States, <https://www.cia.gov/cia/publications/factbook/geos/us.html> (last visited Aug. 6, 2006) (listing population at 298.4 million and per capita GDP of \$41,800).

75. 2005 ANNUAL REVIEW, *supra* note 69, at 26.

76. *Id.*

cases.<sup>77</sup> While the NFL cannot attract global talent because the sport is played sparsely elsewhere, it, along with the NBA and MLB, has aggressively marketed itself internationally.<sup>78</sup>

Interestingly, the Premier League has been successful despite the fact it appears laughably imbalanced by American standards. Following the logic of the competitive balance argument, this fact is puzzling. Like the D.C. Circuit in *Smith*, economists examine balance at the season-level (whether teams can reasonably compete for the league championship), and to a lesser extent, on a game-by-game basis (whether a team could reasonably win any particular game).<sup>79</sup> Though analysts tend to focus more often on season-level balance, the two measures are largely related since a league full of “.500” teams will probably be evenly-matched enough at any given time to produce competitive games.<sup>80</sup>

Additional league competitive balance data are discussed in the paragraphs below, but one might more quickly gather how competitively imbalanced the Premier League is by observing sports gambling odds. Sports wagering is legal in the United Kingdom, and the largest retail gambling operation in the world is the UK-based Ladbrokes—long a FTSE 100 index<sup>81</sup> corporation. In August 2006, approximately one month before the start of the season for both the 32-team NFL and 20-team English Premiership, Ladbrokes naturally was taking bets on the odds to win each league’s respective titles. Given that professional odds-makers try to set and maintain odds in order to balance the flows of money on both sides of a bet, the odds come to reflect the expectations of those who “put their money where their mouth is.” The Indianapolis Colts, the team deemed mostly likely to win the NFL championship (the Super Bowl), were offered at 5:1—meaning for every dollar (well, pound sterling) wagered, the bettor would win five more if the Colts won the league title. These odds imply that the market believes the Colts have about a 17% chance of win-

---

77. For example, at the 2005 MLB All-Star game, the annual home run derby of the game’s top sluggers featured eight different competitors representing eight different countries. ESPN 2006 SPORTS ALMANAC 71 (Gerry Brown & Michael Morrison eds., 2006) [hereinafter ESPN ALMANAC].

78. Super Bowl XXXIX between New England and Philadelphia in February, 2005 was broadcast in 222 countries and translated into thirty-two languages. *Id.* at 246.

79. See *Smith v. Pro Football*, 593 F.2d 1173, 1185 (D.C. Cir. 1978).

80. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 45 (proposing to measure competitive balance based on the difference between the observed level of balance in team records over a season and an ideal level based on the statistically expected standard of deviation if the teams were all equally matched).

81. This is a leading index of the London Exchange, comparable to the Dow 30 in the U.S.

ning the Super Bowl. Meanwhile, the Premier League favorite was Chelsea, offered at 4:9—for every dollar you bet, you would win about 44 cents if Chelsea won the league championship, implying that the market thinks Chelsea has roughly a 70% chance of winning the Premier League Championship.<sup>82</sup> Thus, if you believe the gamblers, the Premier League pales in comparison to the NFL when it comes to season-level competitive balance (and indeed, not just gamblers, but the English media are aware of this contrast).<sup>83</sup> Nonetheless, predictable results have not thwarted the Premier League in rivaling North American sports leagues when it comes to money.

The previous decade’s results in the Premier League show striking imbalance relative to the NFL, NBA, and MLB. In the last ten years only three clubs have won the twenty-team Premier League.<sup>84</sup> Manchester United has won five titles, Arsenal has three and Chelsea two.<sup>85</sup> Meanwhile, in North America, the NBA has boasted four different champions over that spell, six different MLB clubs have won the World Series, and there have been seven different Super Bowl winners (see Table 2).

---

82. See Ladbrokes.com, Online Sports Betting, Poker Games, Casino and Games, <http://www.ladbrokes.com> (last visited Aug. 8, 2006) (printout on file with author). The difference in odds is in part due to the lopsidedness of talent in the Premier League, but also due to the NFL’s championship being the result of a season-ending, single-elimination tournament between top teams. The Premiership’s championship, on the other hand, is awarded to the team with the best overall records in wins, ties, and losses, based on points earned for wins and ties. A single-elimination tournament increases the overall “randomness” that a given team will win the championship by forcing the “best” team to face higher odds of being eliminated due to a poor performance in a single-elimination playoff game.

83. See *In a League of Its Own*, THE ECONOMIST, Apr. 29, 2006, at 63 (writing enviously about how the NFL “giv[es] all 32 owners a chance to field teams that are both financially viable and athletically competitive” in “striking contrast with English football’s Premier League,” which “is dominated by the same teams, year after year”). Oddly for a publication entitled, “The Economist,” the article fails to discuss how the Premier League nonetheless compares well with the NFL with respect to financial investment and revenues. *Id.*

84. Due to relegation and promotion, 36 different teams have, at one season or another, participated in the 20-team Premier League. See *infra* notes 91-93 and accompanying text.

85. See SKY SPORTS 2005-2006 FOOTBALL YEARBOOK 590-99 (Glenda Rollin & Jack Rollin eds., 2005) [hereinafter SKY SPORTS YEARBOOK]. Even this short list of champions needs context. Arsenal and Manchester United alternated as champion for nine years before Chelsea won in 2004-05, and Chelsea’s breakthrough only occurred after Chelsea owner Roman Abramovich boosted its payroll to the “\$200 million mark, roughly the same level as the New York Yankees.” See David Moore, *Constellation of Stars Lifts Chelsea*, USA TODAY, Jul. 19, 2006, at 1C. Chelsea is also the prohibitive favorite to win in 2006-07. See *supra* notes 81-83 and accompanying text.

**Table 2<sup>86</sup>**  
**Season-over-Season Competitive Results 1996-2005**

No. of Different Clubs that:	Premier League (36 diff. teams)	MLB (30 teams)	NBA (30 teams)	NFL (32 teams)
Won League Championship	3 (8%)	6 (20%)	4 (13%)	7 (22%)
Finished in Top Eight (Eng.)/Qualified for Play- offs (N. Am.)	22 (61%)	20 (67%)	28 (93%)	31 (97%)

Winning the title is not the only measure of balance, though. Indeed, winning a league championship is special because it is rare (more so for Cubs fans); but something that is rare may well provide too small a sample to be reliable. Thus, one might also compare the number of teams that qualify for a postseason tournament. For example, competitive imbalance also emerges upon comparing across the past decade the number of North American leagues to have qualified at least once for the playoffs to the number of different Premier League clubs to have finished in the top eight of the league table.<sup>87</sup> The results show that twenty-two of the thirty-six soccer clubs that spent a season or more in the Premier League over that span had at least one top eight finish,<sup>88</sup> while thirty-one of thirty-two NFL teams qualified for the postseason along with twenty-eight of thirty NBA clubs and twenty of thirty MLB teams.<sup>89</sup> The Premier League, as demon-

---

86. ESPN ALMANAC, *supra* note 77, at 95, 246, 371. The NBA Champions were San Antonio Spurs (3 titles), Los Angeles Lakers (3), Chicago Bulls (3), and Detroit Pistons (1). MLB World Series winners were New York Yankees (3 titles), Florida Marlins (2), Chicago White Sox, Boston Red Sox, Anaheim/Los Angeles Angels, and Arizona Diamondbacks (1 each). NFL Super Bowl Champions included the New England Patriots (3 titles), Denver Broncos (2), Pittsburgh Steelers, Tampa Bay Buccaneers, Baltimore Ravens, St. Louis Rams, and Green Bay Packers (1 each). *Id.* at 254-55.

87. Finishing in the top eight in the Premier League serves as a fair comparison with playoff qualification in the three North American leagues. Taken in the aggregate, 39% of teams in the three North American leagues (36 of 92) qualify for the playoffs annually. That correlates closely with the 40% of Premier League teams (8 of 20) that finish in the top eight annually.

88. SKY SPORTS YEARBOOK, *supra* note 85, at 590-99. The comparison is somewhat difficult because, although the Premier League has included 36 different teams during this span, it only includes 20 teams in any particular season. *Id.*

89. ESPN ALMANAC, *supra* note 77, at 99-102, 249-56, 371-75. The only NFL team to miss the playoffs was the expansion Houston Texans, who did not begin play until 2002. The two NBA teams that didn't qualify were the expansion Charlotte Bobcats, who began play only in 2004-05, and the Golden State Warriors. The Warriors were the only

strated by these statistics, represents a challenge to the idea that leagues need competitive balance to succeed financially.

Despite its lack of competitive balance, the Premier League maintains fan interest because of its *competing competitions*. Soccer in England is governed by the Football Association, an independent body not affiliated with any club or club ownership group, that sets the rules for competition.<sup>90</sup> Like major league sports in North America, the Premier League produces a league champion. Unlike North American leagues, the championship is decided via a thirty-eight game schedule with no playoffs, the winner being the team that achieves the best record.<sup>91</sup> However, there are also additional competitions at the top and bottom of the league. In addition to the games that count toward the league championship, teams in the Premier League compete in several tournaments during the season, sometimes with teams from other leagues. At the bottom of the standings is a second competition to avoid placing in the bottom three, who are relegated to the division below and must try to re-qualify for the Premier League the next year.<sup>92</sup>

Fans pay attention to the Premier League in part due to its members' participation in a complex system of tournaments separate from the league games. Two “cup” competitions are interspersed within the league schedule. The Football Association (“F.A.”) Cup is older and more prestigious than the other cup competition, the League Cup. The cups are both structured as single-elimination knockout tournaments in which there are no seedings and the draw for each successive round is done at random. As long as a given team meets basic standards of ability and stadium quality, the F.A. Cup is open to any team in England. Participants in the League Cup include not only the twenty Premier League teams, but also seventy-two other clubs that compete in the lower-echelons of the league struc-

---

team to play the full decade in either the NBA or NFL and not once qualify for the post-season.

90. See TheFA.com, History of the FA Cup, <http://www.thefa.com/TheFA/TheOrganisation> (last visited Apr. 24, 2006) [hereinafter History of the FA Cup].

91. Team records are translated into a total point score by awarding points for wins and ties, with the team holding the most points at the end of the season being awarded the championship. SKY SPORTS YEARBOOK, *supra* note 85, at 48; *see also* Premier-League.com, Rules of the Barclays Premiership Competition, <http://www.premierleague.com/fapl.rac?command=forwardOnly&nextPage=enCompIntro> (last visited Apr. 24, 2006).

92. SKY SPORTS YEARBOOK, *supra* note 85, at 595.

ture.<sup>93</sup> Twice a year, the Premier League and “minor” league teams participate together in single-elimination tournaments (see Table 3).

**Table 3<sup>94</sup>**  
**Summary of Annual Competitions Available to Premier League Clubs**

The Premier League	Every Premier League team plays each opponent twice in the thirty-six game regular season. At the end of the season, the team with the most points (based on wins and ties) is crowned champion.
The Relegation System	Every season the Premier League teams placing 18th-20th in competition are relegated to a lower league named (confusingly) the Championship, where they must attempt to requalify for the Premier League the following season.
The F.A. Cup	A single elimination tournament that takes place on designated weekends during the season. The competition is open to any club within England meeting a basic standard of ability and stadium quality. In 2005-06 a record 674 clubs entered the Cup. The competition was first played in 1872, and the winner receives an invitation to compete in the UEFA Cup.
The League Cup	A single elimination tournament similar in structure to the F.A. Cup, but open only to the 92 clubs in the four professional divisions of the league. Like the F.A. Cup, the winner receives an invitation to the UEFA Cup.
The Champions League	The most prestigious club competition in Europe. The name reflects the fact that it originally invited only the previous season’s Premier League champion to compete against the champions of other European Leagues.. It has since been expanded so that the top four League finishers are also invited to participate. Entrants play a round-robin group stage against other top European clubs, with the top finishers advancing to a single-elimination tournament stage that culminates in the crowning of a European Champion.
The UEFA Cup	A second European tournament structured in a similar manner to the Champions League. The UEFA Cup is typically open to the Premier League teams that finish fifth and sixth in the league standings, along with the winners of the two cup competitions.

Various Premier League teams also compete in European competitions during the season. The Champions League is a competition involving the best club teams from throughout Europe. The top four annual finishers in the Premier League qualify for the following year’s Champions League

---

93. The League Cup was first contested in 1962. *Id.* at 602-19. The FA Cup was first contested in the 1871-72 season and is now in its 125th season. History of the FA Cup, *supra* note 90.

94. SKY SPORTS YEARBOOK, *supra* note 85, at 793-827.

competition. The Champions League starts with a group round robin stage, from which the top teams advance to single-elimination knockout rounds that culminate in the crowning of the European club champion.<sup>95</sup> The UEFA Cup is a secondary European competition whose structure mirrors that of the Champions League with group stages leading to knockout rounds.<sup>96</sup>

In addition to these championship competitions, there is also a competition to avoid the bottom of the barrel. England’s soccer structure allows any club—no matter how small—to reach the top level of play through a system of merit.<sup>97</sup> Every year the bottom three finishers in the Premier League are dropped, or “relegated,” and replaced by the top three teams from the second-tier division, currently known as the Championship. The three relegated teams can earn a return to the Premier League if they finish at the top of the Championship the next season. Similarly the bottom three teams in the twenty-two team Championship are relegated to the division below, known as Division One. That process continues to Division Two, which is the fourth and bottom level of fully professional soccer in England. These four tiers of professional play comprise the “Football League.”<sup>98</sup>

The ladder continues below Division Two with the semi-professional Football Conference. Though the F.A. changes the process from time to time, usually the bottom team from Division Two is relegated to semi-professional status, while the champion of the Conference is promoted to full professional status in Division Two. This process of relegation and promotion continues beyond the Conference through an array of semi-professional conferences down to small regional and local leagues.<sup>99</sup>

This system thus allows any team to enter the league structure. Though initial entrance to the league structure is unhindered, the F.A. sets objec-

---

95. *Id.* at 793-809.

96. *Id.* at 811-27.

97. This Article focuses on professional sports in England, but the system of promotion and relegation is not unique to the country. According to a consultation paper issued by the European Union, the system of promotion and relegation “is one of the key features of European sport.’ In theory teams can start at the lowest rung of the ladder in a regional competition and by dint of sporting merit alone they reach the top.” Stephen Ross & Stefan Szymanski, *Open Competition in League Sports*, 2002 WIS. L. REV. 625, 635-36 (2002) [hereinafter Ross & Szymanski, *Open Competition*] (quoting European Commission Directorate-General X, *The European Model of Sport*, [http://ec.europa.eu/sport/action\\_sports/historique/docs/doc\\_consult\\_en.pdf](http://ec.europa.eu/sport/action_sports/historique/docs/doc_consult_en.pdf) (last visited Nov. 20, 2006)).

98. Noll, *Promotion and Relegation*, *supra* note 65, at 179.

99. *Id.* at 180; *see also* Ross & Szymanski, *Open Competition*, *supra* note 97, at 635-36.

tive requirements for each level, including stadium capacity, financial standing, and ability to pay players. Any club can progress up the ladder as long as it wins games and meets the progressively strict business guidelines set by the F.A. at the higher levels.<sup>100</sup>

The “open architecture”<sup>101</sup> of professional soccer in England helps the Premier League maintain fan interest through these various competing competitions. This is not to say that North American sports leagues should be compelled to follow such a model. Rather, this example demonstrates that competitive balance may not be the requisite that antitrust courts assume. Courts make this assumption based on the questionable paradigm that fan interest must focus on a single, unique league championship rather than multiple rewards and penalties on multiple levels of play. Instead, the experience of professional soccer in England shows that it is possible for fan interest to remain high even if courts were to strike down the anticompetitive restraints protecting competitive balance.<sup>102</sup>

## B. The “Two-Sided Market” Critique: A Bad Trade?

A closer examination of the explanations used by professional sports leagues to defend restraints in antitrust actions reveals several weaknesses

---

100. The open entry system also provides an outlet for disgruntled fans. Last year some Manchester United fans, furious over the sale of the club to American billionaire Malcolm Glazer, founded their own club and named it F.C. United of Manchester. The club has begun play in the North West Counties League, and could conceivably play in the Premier League in a decade if promoted yearly. John Cassidy, *The Red Devil*, THE NEW YORKER, Feb. 6, 2006, at 46.

101. The term “open architecture” has been used for more than a decade to describe systems that are characterized by “free access” and “decentralized” design, such as the internet, rather than the traditional “closed,” “channelized” systems characterized by scarcity, like traditional broadcast media. See Jerry Berman & David Weizner, *Abundance and User Control*, 104 YALE L.J. 1619, 1622 (1995). The term has an older, similar meaning when referring to the design of computer hardware. See Rochelle Dreyfus, *Dethroning Lear: Licensee Estoppel and the Incentive to Innovate*, 72 VA. L. REV. 677, 735 n.206 (1986) (describing IBM as having designed the PC with “open architecture” rather than a “proprietary” system).

102. It is possible that, taking existing monopolistic practices and anti-competitive restraints as a given, competitive balance could become more important than it would otherwise be. However, where such arrangements are not compelled or authorized by the government apart from the antitrust laws, giving weight to the competitive balance argument because of such anti-competitive circumstances would essentially give bonus points to law violators. In a sense, applying the competitive balance argument to monopolistic enterprises may be sports’ version of antitrust’s famous “Cellophane Fallacy.” See HERBERT HOVENKAMP, FEDERAL ANTITRUST POLICY § 3.4b-c (discussing error in assuming that currently monopolized industry is “normal” when considering consumer response).

in the competitive balance argument. While team owners and leagues cite the competitive balance ideal, an alternative motive for labor market and investment restraints lies in applying the concepts of monopsony and two-sided markets to sports leagues. Taken together, these concepts undercut the asserted validity of the economic tradeoff between anti-competitive costs and pro-competitive benefits that lies at the heart of the competitive balance argument. As this Section will discuss, monopsony (or buyer market power) creates significant problems for the competitive balance argument. The economic issues inherent in two-sided markets further exacerbate the flaws in the competitive balance argument.

### 1. *Monopsony*

Courts and commentators have readily applied the theory of monopsony to professional sports leagues.<sup>103</sup> If teams can jointly agree, through league rules or otherwise, to reduce the competition faced by the talented player labor they employ, they can reduce the cost of one of their most important inputs. As a result, they often institute restrictive policies—such as player-entry drafts, reserve clauses, and salary caps<sup>104</sup>—that reduce the ability of players to force teams to compete for their services. Additionally, in the same way that a monopoly cartel must worry about a new entrant undercutting their price, a monopsony cartel must worry about defectors or new competitors outbidding their low fixed price. The concern manifests itself in league policies designed to control new investment and stem the development of competing leagues.<sup>105</sup>

---

103. *See* Chi. Prof'l Sports Ltd. P'ship v. NBA, 95 F.3d 593, 599 (7th Cir. 1996) (Easterbrook, J.) (observing that, to its players, the NBA “looks more like a group of firms acting as a monopsony” than as a single source of a product called “NBA basketball”); *Brown v. Pro Football, Inc.*, 50 F.3d 1041, 1061 (D.C. Cir. 1995) (Wald, J., dissenting) (observing that “[a]thletic prowess is . . . a unique and highly specialized resource, of precisely the genre vulnerable to monopsony manipulation”). *See* Ross, *Misunderstood Alliance*, *supra* note 8, at 542 (observing and defending the use of professional sports league restraints to enhance monopsony power).

104. *See* Mackey v. Nat'l Football League, 543 F.2d 606, 621 (8th Cir. 1976) (considering restraints on movement of existing players within a sports league); *Smith v. Pro Football*, 593 F.2d 1173 (D.C. Cir. 1978) (considering player-entry draft system); *Brown*, 50 F.3d 1041 (considering fixed salary system for players).

105. *See, e.g.*, *Sullivan v. Nat'l Football League*, 34 F.3d 1091, 1111 (1st Cir. 1994) (considering challenge to league ownership policy); *U.S. Football League v. Nat'l Football League*, 634 F. Supp. 1155 (S.D.N.Y. 1986) (considering challenge to alleged exclusive deals to thwart entrant by locking up television broadcast market).

Monopsony has an often-ignored cost, though.<sup>106</sup> Like monopoly, it leads to lost efficiency through “deadweight loss.”<sup>107</sup> A monopolist causes its consumers to purchase less of its product when it raises prices. Thus, a rational monopolist makes up for the lost output by recouping revenue on the remaining units sold. Similarly, a monopsonist causes its suppliers to sell less of their product when it lowers prices. A rational monopsonist makes up for the lost supply by paying less on the remaining units it purchases. That is, the monopsonist benefits by paying less for what it buys, which represents a transfer of welfare from its suppliers due to the monopsonist’s market power. Furthermore, monopsony generates an additional overall loss to society, since the amount of output purchased is less than what it would be in a competitive market.

When applied in the context of major league sports, monopsony theory leads to a claim that may be hard to swallow for those repeatedly subjected to the rhetoric of “overpaid” athletes playing a “game” for a living. Monopsony theory predicts that the aggregate number of athletically talented people striving to become professional athletes will plummet if sports league wages are depressed and other fields look relatively more attractive. This theory should also make sense to law students, since typically the number of students applying to law schools—and consequently the test scores of those admitted—rises in a recession, when alternative fields look worse by comparison. To be sure, recession or not, law schools will fill their classes and still “look” and “feel” the same—just as the Yan-

---

106. Despite its efficiency loss, *see infra* notes 108-109 and accompanying text, courts have sometimes treated monopsony more leniently than monopoly on the grounds that while seller market power leads to higher prices, buyer market power leads to lower prices. These prices could then be passed on to consumers if the buyer were, for example, a chain store or intermediary. *See, e.g.,* *Balmoral Cinema v. Allied Artists Pictures Corp.*, 885 F.2d 313, 314 (6th Cir. 1989) (suggesting that an agreement among theater owners not to bid against each other for movies to show might “serve rather than undermine consumer welfare” because it could result in lower prices for consumers); *Kartell v. Blue Shield*, 749 F.2d 922, 930 (1st Cir. 1984) (refusing to condemn a health insurer’s policy of setting the maximum price it would pay for health care services used by its insureds and stating that “the prices at issue here are low prices, not high prices”). This appears to ignore the problem of deadweight loss, which is common to both monopoly and monopsony. *Compare* HOVENKAMP, *supra* note 102, § 1.2 (questioning differential treatment since monopsony reduces output in the monopsonized market just as monopoly does in the monopolized market, and observing that “[m]any federal judges have failed to see this”), and Noll, *Buying Power*, *supra* note 8, at 591 (concluding that “asymmetric treatment of monopoly and monopsony has no basis in economic analysis”), with Jonathan M. Jacobson and Gary J. Dorman, *Joint Purchasing, Monopsony and Antitrust*, 36 ANTITRUST BULL. 1 (1991) (arguing for more lenient treatment of joint-purchasing organizations by competitors).

107. *See* HOVENKAMP, *supra* note 102, § 1.2.

kees will still have 25 players on their roster. Some might even repeat the cliché that the most successful law students (such as law review editors), are, like major-league athletes, overpaid.<sup>108</sup> But if law firms acted in concert to downwardly fix the prices that they paid to law review editors, studying law would become a less attractive career choice for the potential students at the margin. Similarly, given the long odds against making it to the major leagues, a young person with athletic talent might skip taking the first step on the professional ladder and instead proceed into another field. As a result, the “deadweight loss” theory predicts that under monopsony, the overall talent level in a league (or law schools) will decline due to depressed wages.<sup>109</sup>

Courts and commentators who advocate the competitive balance argument often fail to note either the possibility of monopsony or the “deadweight loss of talent” that monopsony creates. A few commentators, like Professor Stephen Ross, defend the competitive balance argument despite recognizing leagues’ tendency to become monopsonies.<sup>110</sup> However, as discussed in the next Section, viewing professional sports leagues through the lens of two-sided market theory compels the conclusion that the nature of the leagues magnifies these harms.

## 2. *Two-Sided Markets*

Courts and commentators often try to fit professional sports leagues into the corporate and antitrust paradigm of “joint ventures” when evaluating the restraints employed by these leagues.<sup>111</sup> Comparing sports leagues to joint ventures makes sense if one assumes that teams in a league are separate entities cooperating to make a product.<sup>112</sup> But sports leagues are

---

108. See, e.g., Geoff Yuda, *Competition: Just Part of the “Business”?*, 23 PENN. LAWYER 28, 29 (2001) (reporting results of a survey in which 39 percent of large-law-firm managing partners in Pennsylvania thought first- and second-year associates were “grossly overpaid”); *Associate Salaries: Are Your Firm’s on Target with Market Rates?*, 96-2 LAW OFFICE MGMT. & ADMIN. REPORT 1 (Feb. 1996) (reporting that “many law firm administrators and managing partners fe[lt] that associates” were “overpaid” in 1996 despite static salaries for years). The author, on the other hand, believes that the market is just “clearing.”

109. See Noll, *Buying Power*, *supra* note 8, at 589.

110. See, e.g., Ross, *Misunderstood Alliance*, *supra* note 8, at 540-42.

111. See *Chi. Prof’l Sports Ltd. P’ship v. Nat’l Basketball Ass’n.*, 95 F.3d 593, 599 (7th Cir. 1996) (considering applicability of joint venture rubric to the NBA); see, e.g., Ross & Szymanski, *Open Competition*, *supra* note 97, at 627 (arguing efficiency would be enhanced during the entry of new teams into sports leagues if teams were viewed as joint ventures).

112. See Thomas A. Piraino, *A Proposal for the Antitrust Regulation of Professional Sports*, 79 B.U. L. REV. 889 (1999) [hereinafter, Piraino, *Proposal for Antitrust*] (arguing

more than simply joint ventures between teams; they are also intermediaries between players on one side, and fans on the other. Thus, they form a two-sided market.

Two-sided markets<sup>113</sup> are situations in which one or several platforms—typically a media outlet, a market, a network, or even computer software—facilitate interactions between different end-users. The owner of this platform will usually try to court both sides of the market at prices that allow a profit.<sup>114</sup> Typical examples of two-sided markets include newspapers that compete for advertisers as well as readers, computer operating systems that compete for software developers to write applications and consumers to use them, and credit card payment systems that try to attract both merchants and cardholders. Two-sided markets are characterized by what economists call “demand-side economies of scale,” more commonly known as the “chicken-and-egg” problem. These effects are sometimes referred to as “network effects.”<sup>115</sup> Specifically, an end-user in the market values the product more as more end-users adopt it. For example, Microsoft Windows becomes more valuable to a computer user as more users adopt it and as more software developers write applications for it. The shelves of law school libraries contain much discussion about two-sided markets and the network effects that characterize them. However,

---

that antitrust treatment of joint ventures is also appropriate for monopolistic professional sports leagues).

113. “Two-sided market” is a term of art borrowed from economics. Its market definition may not necessarily apply in antitrust law, since case law has sometimes discussed different sides of a two-sided market as, in fact, different markets. *See* Jean Charles Rochet & Jean Tirole, *Two-Sided Markets: An Overview* (IDEI-CEPR, Toulouse, France, January 23-24, 2004), available at [http://www.frbatlanta.org/filelegacydocs/ep\\_rochet\\_over.pdf](http://www.frbatlanta.org/filelegacydocs/ep_rochet_over.pdf); David Evans, *The Antitrust Economics of Multi-Sided Platform Markets*, 20 *YALE J. ON REG.* 325, 330 (2003) (defining a two-sided market, but observing that “[d]espite their economic importance, multi-sided markets have only recently received attention from economists and, with the exception of some recent work on payment cards, have received virtually no attention in the scholarly literature on antitrust”); Randal C. Picker, *Unbundling Scope-of-Permission Goods: When Should We Invest in Reducing Entry Barriers?*, 72 *U. CHI. L. REV.* 189, 202-03 (2005) (describing Microsoft’s Windows Media Player as a platform in a two-sided market, with consumer listeners on one side and content creators on the other).

114. *See* Rochet & Tirole, *supra* note 113, at 4.

115. “Network effects” have been defined as situations where consumers of a product benefit from other consumers’ use of it or a compatible product. *See* Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 *CALIF. L. REV.* 479, 483-84 (1998).

the application of the concept to the antitrust law of sports leagues appears to be a novel issue.<sup>116</sup>

A professional sports league can be seen as a two-sided market in which individual teams and leagues as a whole provide platforms to court both the fan- and player-sides of the market. They do so in an environment dominated by a chicken-and-egg problem between these two groups of end-users. That is, all else being equal, fans are attracted to teams and leagues with more athletic talent—witness the higher prices and greater attendance at major league professional sports versus minor leagues. On the other side of the market, talented players are attracted to teams and leagues with higher numbers of fans, which makes higher player salaries possible. Each end of the market feeds into the other.

Given this environment, it follows that teams in a league increase their profits if they coordinate their behavior to maximize the talent they can purchase while minimizing its cost, thus maximizing the fan interest they can translate into revenue. Monopsony thus encourages restraints to keep costs lower on the player-side of the market. Leagues might also be expected to apply restraints among teams to raise revenue on the fan-side of the market.

Defenders of the competitive balance argument argue that the benefit to one side of a two-sided market outweighs the harm to the other side of the market.<sup>117</sup> In particular, they argue that benefit accruing to leagues through competition against other leagues or other forms of entertainment outweigh the economic harms of monopsony in the player input market.<sup>118</sup> In arguing this, they endorse a distributional tradeoff between the economic welfare of one group (players) versus another (fans). While one can

---

116. One commentator has described sports leagues as benefiting from network effects, because they become more attractive to fans as they become “national in scope,” and has described nationwide leagues in the single sports context as “natural monopolies.” Piraino, *Proposal for Antitrust*, *supra* note 112, at 899. This conception of network effects differs from the two-sided market idea described in this Section, and may in fact simply be a form of economies of scale, rather than “true” network effects. Natural monopolies are often defined as industries characterized by such strong economies of scale that they exhibit “declining average costs”—meaning that having a single supplier can be more efficient than having competition. See John Duffy, *The Marginal Cost Controversy in Intellectual Property*, 71 U. CHI. L. REV. 37, 38 (2004).

117. See Ross, *Misunderstood Alliance*, *supra* note 8, at 539-41 (defending competitive balance argument based on “possibility” that harm caused by monopsony to competition in the player labor market may be offset by increased output through qualitatively better entertainment due to competitive balance).

118. *Id.*

defend such a tradeoff on non-economic grounds,<sup>119</sup> one can only defend such a tradeoff economically by claiming that the benefits to one group outweigh the costs to the other. This argument differs from efficiency defenses in which the benefits and harms accrue to the same parties. In a prominent example, regulators and courts considering major corporate mergers frequently face the argument that the reduction in competition that consumers may face when competitors merge will be more than offset by a reduction in costs due to increased efficiency as a result of the merger. In such an argument, the costs and benefits accrue to the same group: consumers. Despite this, antitrust has rejected this kind of tradeoff in the merger context.<sup>120</sup> Given this doubt, redistributive transfers between different parties have not surprisingly been viewed skeptically, with those skeptics advocating that they be tightly constrained.<sup>121</sup> In a further distillation of these views, the First Circuit has endorsed the position that, in the professional sports context, such a tradeoff is relevant only to the degree that the benefit in one market “feeds back” into the harmed market.<sup>122</sup>

---

119. *Id.* at 541 (citing “the need to democratically interpret the Sherman Act” and stating that “it is highly doubtful that the voters . . . would prefer such an anti-consumer result” as the invalidation of “an agreement necessary to enable a sports league to offer exciting championship races should nonetheless be invalidated because it has an adverse effect on some player[s]” salaries). It should be noted that this view implicitly assumes the necessity of a single unique championship competition, as critiqued in the previous Section.

120. *See* *United States v. Phila. Nat’l Bank*, 374 U.S. 321, 370 (1963) (rejecting defendants’ argument that they should be allowed to merge because, although competition in Philadelphia would be reduced, the banks would be better able to compete for large business accounts against New York-based banks); *FTC v. Tenet Healthcare Corp.*, 17 F. Supp. 2d 937, 948 (E.D. Mo. 1997) (concluding that balancing of pro-competitive effects in one market versus anti-competitive effects in another is not allowed in merger cases).

121. *See, e.g.*, Robert Pitofsky, *Efficiencies in Defense of Mergers*, 7 *GEO. MASON L. REV.* 485, 490-91 (1999) (advocating a “narrow approach” to such tradeoffs in the merger context due to “measurement issues” concerning the harms and benefits and the “fairness question” of whether it is “appropriate to deny to one group the guarantees of a competitive market in order to provide the benefits of efficiency to another group”). There are those who do advocate this kind of interparty transfer, although they have drawn some criticism. *See* Noll, *Buying Power*, *supra* note 8, at 591-92 (noting that antitrust tolerance of monopsony “must be based on an argument that it is socially desirable to redistribute income to a group of buyers even if doing so is costly to their members of society” and calling such a position “difficult to defend”).

122. *See* *Sullivan v. Nat’l Football League*, 34 F.3d 1091, 1113 (1st Cir. 1994) (concluding that considering pro-competitive effects on another market is legitimate “to the extent the NFL’s policy [that bans public ownership] strengthens and improves the league” as “a popular entertainment product unimpaired by the conflicting interests that public ownership would cause . . . resulting in increased competition in the market for ownership interests in NFL clubs”).

That is, the market aided by a restraint must itself generate a compensating benefit in the harmed market that exceeds the harm created by the restraint.<sup>123</sup>

The effects of monopsony and the characteristics of two-sided markets make the tradeoff at the core of the competitive balance argument difficult to defend. First, monopsony involves a reduction in output, and thus, a deadweight loss to overall market efficiency. If a buyer cartel holds prices down, fewer sellers will enter the market. Thus, even if restraints create a reduction in player labor costs that accrues to fans as consumers, there is still a net welfare loss as athletic talent forgoes entering the market at the lower price. While it may be difficult in practice to determine whether less athletic talent is entering a monopsonistic league than it otherwise would, this general effect of monopsony is fairly undisputed.<sup>124</sup>

Second, the demand-side economies of scale that exist in professional sports characterize it as a two-sided market. This categorization means that the output-reducing effects of monopsony should be magnified. In the same way that fans find a team or league more attractive as it attracts more talent, conversely the reduction in player talent that monopsony causes should tend to dampen fan interest. Additionally, if a player finds a team or league that draws more fans more attractive, then a reduction in fan interest should also feed back into the player market. The end result should be a spiral of dissatisfaction. Strong market power combined with this dynamic may explain why professional leagues remain very profitable at least temporarily, while fan interest wanes or shifts to competing forms of entertainment (e.g., NASCAR). In the same way that it is difficult to ascertain whether monopsony actually causes downward pressure on the amount of player talent, determining whether fan interest is lower than it might have been may be hard to measure. However, if sports leagues are properly understood as two-sided markets,<sup>125</sup> such effects ought to be real.

Sports leagues defend their right to engage in monopsony practices and reap the gains that come with being a successful platform in a two-

---

123. *Id.*

124. See Roger D. Blair & Jeffrey L. Harrison, *Antitrust Policy and Monopsony*, 76 CORNELL L. REV. 297, 316 (1991) (noting that monopsony “reduce[s] the producers’ profits, causing them to reduce supply in the future”); Noll, *Buying Power*, *supra* note 8, at 594 (arguing that monopsony’s “net effect on society is a loss of efficiency and welfare and a redistribution of wealth to consumers of the monopsonized product”); see also *supra* notes 106-107.

125. See, e.g., SangHoo Bae & Jiyoung Kwon, *Two-Sided Network Effects in the Market for Professional Sports* (Northeastern University Economics Dept. Seminar Series, Feb. 7, 2006), available at <http://www.economics.neu.edu/activities/seminars/documents/baepaper.pdf>.

sided market. They argue that they are either currently constrained by external competitors in the same field, or that they will have to compete in the future with new entrants “for the field” itself.<sup>126</sup> Defendants in other antitrust cases involving two-sided markets have also made these arguments.<sup>127</sup> Deserving of merit or not, these claims are distinct from the assertion that there is an inherent, internal need for competitive balance within a sports league. Indeed, a focus on leagues’ external competitors and on possible new competitors in order to excuse the leagues’ restraints is congruent with the approach taken by antitrust defendants in other areas.<sup>128</sup>

The competitive balance argument is flawed both in its assumption that a single unique competition is beneficial, and that the tradeoff it creates between fans and players is justified. However, the competitive balance argument also draws strength from a third justification: that it is tightly linked to fan interest.

### C. The “Weak Link” Critique: Letting Fans Call Their Own Shots

In addition to advocating for restraints in the market for player labor, the competitive balance argument also has a fan-side component. The argument goes that with a system of competitive balance, fans will show more interest in a sports league compared to other leagues or other forms of entertainment. However, empirical research from the fairly new field of sports economics refutes this alleged connection between fan interest and competitive balance. Furthermore, there are strong arguments that, even if they wanted to, some sports leagues would not or could not significantly improve competitive balance through their policies. Finally, the rejection of the competitive balance argument extends beyond the narrow confines of antitrust law. The example of major league baseball’s fight against fan-organized fantasy leagues suggests that this Article’s contentions bear on questions beyond the rule of reason in antitrust, into questions of intellectual property rights.

---

126. *United States v. Microsoft Corp.*, 253 F.3d 34, 49 (D.C. Cir. 2001). *See generally* Harold Demsetz, *Why Regulate Utilities?*, 11 J.L. & ECON. 55 (1968) (discussing the distinction between competition for a field and competition within it).

127. *Microsoft*, 253 F.3d at 49.

128. *Id.* at 51-55 (considering what other products might be able to play the intermediary role that Microsoft Windows does between software users and software developers).

### 1. *Competitive Balance and Fan Interest*

The basic, fan-side premise of the competitive balance argument is that fans find sports leagues more attractive when teams are more evenly-matched. On a broad level, this is almost certainly true; few would want to watch a game where the identity of the victor was already a certainty. Indeed, the fact that television replays of classic sports events draw abysmal ratings further proves this point.<sup>129</sup>

But the fact is that games with an absolute certainty of outcome are like Soviet elections: an extreme case. Decades ago, scholars found a link between fans' desire to attend games and the level of uncertainty surrounding the victor of a league championship race, but their findings assumed an extreme case in which the identity of the victor was either known or unknown.<sup>130</sup> Even if it were true that total absence of surprise repels fans' attention, this does not by itself prove that maintenance of fan interest requires league-wide competitive balance.<sup>131</sup> In fact, recent studies of the link between fan interest and the degree of competitive balance (rather than the simple binary choice of whether victors are known or un-

---

129. See Stefan Szymanski, *Tilting the Playing Field: Why a Sports League Planner Would Choose Less, Not More, Competitive Balance 2* (Tanaka Business School Discussion Papers, TBS/DP05/35, 2004) [hereinafter, Szymanski, *Tilting the Playing Field*], available at <http://www3.imperial.ac.uk/pls/portallive/docs/1/43004.PDF> (last visited July 10, 2006) (observing that “without uncertainty a game is relatively uninteresting” and noting the lack of “excitement generated by watching a re-run of last year’s World Series”); BERRI, SCHMIDT & BROOKS, *supra* note 11, at 26 (observing that “the fact that we do not know the outcome attracts us” and noting “the dismal ratings ESPN Classic earns from showing reruns of the greatest sporting events in history”). The comparison between watching a rerun and lack of competitive balance is not totally apt, however, since a fan with a good memory watching a rerun might recall not only the victor of the game, but could also anticipate the major events in the game as they were about to take place. To the extent that part of the viewer’s utility is the “adrenaline rush” while watching quick, unexpected events, that would be lost.

130. Two studies suggest that more fans attend a game when a championship race is closer. See, e.g., HENRY G. DEMMERT, *THE ECONOMICS OF PROFESSIONAL TEAM SPORTS* 11 (1973); Roger Noll, *Attendance and Price Setting*, in *GOVERNMENT AND THE SPORTS BUSINESS* 115, 156-57 (Roger Noll ed., 1974).

131. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 82-83. The authors point out that “[t]here is a disconnect between the theoretical and empirical treatment of competitive balance.” *Id.* The reason is that theories that discuss when uncertainty of outcome is “completely taken away” do not fit a world where “uncertainty of outcome” persists even when the level of competitive balance is “quite low.” *Id.* In other words, even during the Yankees’ dominant periods in baseball, fans could not be completely certain of the outcome of games or season championships.

known) have found little empirical evidence that such a link exists.<sup>132</sup> For example, in a 2006 book, economists David Berri, Martin Schmidt, and Stacey Brook examined the relationship between competitive balance and attendance in the NBA over time and found no statistically significant relationship.<sup>133</sup> In 2002, economist Brad Humphreys weighed several measures of competitive balance against baseball attendance over the course of the 20th century and found virtually no relationship.<sup>134</sup>

The lack of empirical support for the proposition that increased competitive balance attracts fans at the margin should not be as surprising as it sounds. There are disproportionately successful teams that opposing fans “love to hate”; revenues and attendance numbers do not discriminate between attendance out of admiration versus dislike. Furthermore, “anti-Yankees” fans, for example, do not live in a world in which outcomes are 100% foreordained. Even if they enjoy their schadenfreude relatively rarely, the intense feelings and emotions when they do could compensate for the lack of competitive balance.

## 2. *Competitive Balance and League Policies*

The competitive balance argument also falls short because of the lack of evidence that leagues seek, or are even capable of achieving, high levels of competitive balance. If leagues do not possess sufficient motive to create competitive balance, then it is reasonable to doubt whether they will design policies accordingly. This doubt applies also to those league policies that run afoul of the antitrust laws. Additionally, even if leagues pursued policies that promote competitive balance, there is reason to wonder whether they can practicably achieve it.

---

132. *Id.* at 216 (stating that “we do not have much evidence that fans really care about the level of competitive balance”); see Stefan Szymanski, *Income Inequality, Competitive Balance, and the Attractiveness of Team Sports: Some Evidence and a Natural Experiment from English Soccer*, 111 *ECON. J.* 69, 69 (2001) (concluding, with respect to English professional league soccer, that “match attendance appears unrelated to competitive balance”).

133. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 64-82.

134. Brad Humphreys, *Alternative Measures of Competitive Balance in Sports Leagues*, 3 *J. SPORTS ECON.* 133-48 (2002). Humphreys found that there was no statistically significant relationship between attendance and competitive balance as measured by either the Herfindahl-Hirschman Index familiar to antitrust merger analysis or by an index based on the relationship of observed standard deviation of wins in a league compared to the standard deviation statistically expected in a league full of games where wins and losses are as even as a coin-toss. *Id.* Using a third, newly-introduced eponymous measure, Humphreys examined baseball throughout the 20th century and found a very weak relationship: moving from the lowest level of observed competitive balance to the highest level would draw an average of only four more fans per game. *Id.*

Some recent economic studies have suggested that, at the margin, increased competitive balance could actually decrease fan attendance.<sup>135</sup> In one example, Professor Stefan Szymanski analyzed the relationship between attendance and wins for the 30 MLB teams from 2001 and 2003, and discovered that a situation of perfect competitive balance would actually generate *lower* league-wide attendance.<sup>136</sup> Similarly, Professors Berri, Schmidt, and Brook confirmed that, based on data from the NBA’s 2003-04 season, increasing competitive balance would have negatively impacted league-wide attendance.<sup>137</sup> Indeed, both studies point out that markets differ in size and economic response to wins. The authors infer that because wins appear to generate more revenue in some markets than others, leagues to some degree actually have a strong interest in avoiding competitive balance.<sup>138</sup>

Given the evidence that competitive balance may actually decrease fan attendance, a rational sports league might choose not to pursue competitive balance as a matter of policy.<sup>139</sup> Economists make sense of these findings with the intuition that not all teams in a league are created equal. At the margin, some teams gain more fan interest than others for every win. This could be for a variety of reasons, including the fact that some teams play in larger markets with more people.<sup>140</sup> However, there are only so many wins to go around.<sup>141</sup> As a result, a league that is not yet at full attendance capacity could sell additional seats to its games by instituting

---

135. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 83 (finding that empirical “analysis of gate revenue [in the NBA] suggests that [moving closer to] perfect balance would lower league revenues”); See Szymanski, *Tilting the Playing Field*, *supra* note 129, at 28 (concluding, based on empirical study, that increased competitive balance may not be in the interest of maximizing attendance in Major League Baseball).

136. See Szymanski, *Tilting the Playing Field*, *supra* note 129, at 28.

137. See *id.* at 127.

138. See *id.* at 8, 28.

139. One might argue that professional sports leagues could maximize their competitiveness versus other forms of entertainment by pursuing an optimal level of competitive imbalance. Such a standard seems utterly unworkable for courts under the rule of reason. At any rate, this is not the competitive balance argument as courts have so far approached it.

140. See BERRI, SCHMIDT & BROOKS, *supra* note 11, at 83 (finding that “an additional win had greater value for the Chicago Bulls, New York Knicks and Los Angeles Lakers than one more win had for the Minnesota Timberwolves, Indiana Pacers and San Antonio Spurs” for the 2003-04 NBA season).

141. In the absence of ties, the number of wins must equal the number of losses, league-wide.

policies that yield more wins for teams that generate comparatively more marginal revenue per win.<sup>142</sup>

Even if professional sports leagues want to achieve competitive balance, they may not be able to do so. Recent economic studies have found that one of the most important determinants of competitive balance is the size of the underlying population capable of playing the sport.<sup>143</sup> The scope of the eligible population depends on factors beyond a sports league's control plus factors, like the inherent nature of the sport, which the league cannot easily alter. For example, economists have found, perhaps unsurprisingly, that the key factor limiting competitive balance in the NBA is that there is a short supply of very tall people, and an even shorter supply of those with NBA-caliber skills.<sup>144</sup> The NBA cannot simply make more Shaquille O'Neals and Major League Baseball cannot easily engineer another Barry Bonds.<sup>145</sup> To the extent that the nature of a sport subjects an already small minority of athletically gifted people to an additional filter—e.g., “are you 7' tall?” or “can you hit a slider?”—the variance among those who succeed will increase, and achieving competitive balance may become difficult, if not impossible.

Given the difficulties of increasing competitive balance and the questionable motives to do so, it seems overly charitable for courts to allow sports leagues to claim that their anti-competitive restraints aim to achieve such a result. After all, sports leagues have not attempted to create or en-

---

142. See Szymanski, *Tilting the Playing Field*, *supra* note 129, at 32 (observing applicability of this insight, but noting that a league like the NFL, where “nearly all” games are already sold out, would not face the same incentives).

143. See Sangit Chatterjee & Mustafa R. Yilmaz, *Parity in Baseball: Stability of Evolving Systems?*, 4 CHANCE 37 (1991); see also ANDREW ZIMBALIST, *BASEBALL AND BILLIONS* 97 (1992); Martin B. Schmidt & David J. Berri, *On the Evolution of Competitive Balance: The Impact of an Increasing Global Search*, 4 ECON. INQUIRY 693 (2003).

144. David J. Berri, Stacey J. Brook, Aju Fenn, Bernd Frick & Roberto Vicente-Mayoral, *The Short Supply of Tall People: Competitive Imbalance in the National Basketball Association*, 39 J. ECON. ISSUES 1029 (2005).

145. In fact, many restrictive policies found in sports leagues (such as minimum-age requirements and limitations on the flow of “backup” players on one team to another team where such players would be on-field starting players) actually reduce the pool of talent available to draw upon. Furthermore, Congress has threatened to close off pharmaceutical solutions to this problem. Clifton Brown, *Uniform Steroid Rule is Proposed in House*, N.Y. TIMES, Apr. 28, 2005, at D1 (“While Commissioner Paul Tagliabue defended the National Football League’s steroid policy before a House committee Wednesday, some members of Congress were planning to support a uniform testing policy for all professional sports leagues in the country.”); Murray Chass, *Congress Blows Smoke and Ignores Real Killers*, N.Y. TIMES, Nov. 15, 2005, at D4 (describing proposed legislation that sets forth penalties for steroids use in professional sports).

force rules requiring that particularly outstanding players be shared among teams as an “essential facility” for winning a championship. But, as the next Section discusses, fans have innovated to create a system for sharing such players and for maintaining interest by creating fantasy leagues that allow them to “reallocate” key talent and create their own competitive balance.

### 3. *Major League Baseball v. New Competing Competitions*

The competitive balance argument presumes that the league itself must produce the competition that generates fan interest (and thus revenue). However, recent experience suggests that that is not necessarily the case. Fan innovation and globalization are capable of creating new competitions outside of the traditional league system.

Fans create interest-generating competition through so-called “fantasy” leagues in which they reallocate (or “draft”) players onto fantasy teams of their own creation. Despite the term, “fantasy” baseball does not involve sending Batman, Cinderella and Don Quixote up to bat. Instead, fantasy players form a league by “drafting” real-life professional players to be members of their respective fantasy teams. The statistics that these players generate in real games are collected, and fantasy points are awarded based on these statistics. The better a fantasy team owner’s real-life players perform, the more fantasy points his or her team accrues. The points tallied ultimately determine a fantasy league champion.<sup>146</sup> Invented in 1979 by writer and editor Daniel Okrent as “roisserie baseball,” such fantasy leagues have spread beyond baseball to other sports.<sup>147</sup> As a result, they have become widespread, with 15 million people spending approximately \$1.5 billion annually to create and participate in such leagues.<sup>148</sup> The advent of computers and the internet aided the gain in popularity, as they have made tracking and crunching player data much easier.<sup>149</sup> These

---

146. For an authoritative history and description of “fantasy” or “roisserie” baseball, see SAM WALKER, *FANTASYLAND: A SEASON ON BASEBALL’S LUNATIC FRINGE* (2006).

147. Okrent is credited with inventing roisserie league baseball. See, e.g., Matthew Purdy, *Who’s on First? Wonder No More*, N.Y. TIMES, June 7, 2001, at G1 (describing Okrent as “the Abner Doubleday of Roisserie baseball” after the claimed inventor of actual real-world baseball). However, board games based on the principle of simulating professional baseball by drawing on real world statistical data for its players existed for decades before Okrent’s invention. See Lorne Manly, *Strat-o-matic A Throwback, Endures Era of the Xbox*, N.Y. TIMES, Jan. 13, 2006, at D1.

148. See Alan Schwarz, *Baseball Is a Game of Numbers, But Whose Numbers Are They?* N.Y. TIMES, May 16, 2006, at A1.

149. See Jim Hu, *Let the Fantasy Games Begin—At Work*, CNETNEWS.COM, Sept. 27, 2004, [http://news.com.com/Let+the+fantasy+sports+games+begin--at+work/2100-1038\\_3-5381539.html](http://news.com.com/Let+the+fantasy+sports+games+begin--at+work/2100-1038_3-5381539.html) (describing the online industry and observing that “[b]efore the Web,

“virtual reality” fan-operated leagues rely on the “real world” leagues for the player performances and statistics that the fantasy leagues use to generate their own games.<sup>150</sup> As a result, fans produce an “improved” league that contains the same level of unpredictability as the underlying real-world sports league. Because fantasy leagues rely on actual events in the real-world league, they help maintain fan interest in the latter.<sup>151</sup>

Unfortunately for the owners of fantasy baseball teams, Major League Baseball has taken a very real stance against these leagues. By litigating, MLB’s internet arm seeks to compel the entities that provide fantasy baseball player statistics to purchase licenses from them.<sup>152</sup> In a case currently on appeal, MLB has so far been unable to garner such intellectual property protection.<sup>153</sup> The basis of MLB’s argument does not appear to be that they themselves own the statistics, an argument that courts have previously rejected.<sup>154</sup> Rather, MLB, which purchased the players’ internet and

---

fantasy players would gather after work and pour through daily papers and sports magazines to compile their scores . . . [b]ut now the Internet does all the grunt work in real time”).

150. *Id.*

151. *See, e.g., Has Major League Baseball Hit a Foul in Its Recent Skirmish with Online Fantasy Leagues?*, KNOWLEDGE@WHARTON, June 14, 2006, <http://knowledge.wharton.upenn.edu/article.cfm?articleid=1495> (free registration required) [hereinafter *MLB Skirmish*] (quoting *Baseball America* reporter stating that “[t]here is no question that fantasy leagues have helped baseball maintain and cultivate its popularity” and that they “have given people a great reason to pay attention every day”). The fantasy leagues generate interest among those who otherwise might ignore baseball, though they also alter the interest of existing fans. *Id.* (quoting fantasy-league magazine editor asserting that he “wouldn’t care at all about Major League Baseball but for fantasy baseball”); WALKER, *supra* note 146, at 63 (author describing history of and own participation in a fantasy baseball league and observing that “[f]or as much baseball as I’m watching in these early days, I don’t have the slightest idea what the [real] standings are, nor do I care”).

152. *See* Complaint at 5, C.B.C. Distrib. & Mktg. v. Major League Baseball Advanced Mktg., 443 F. Supp. 2d. 1077 (E.D. Mo. 2006), *available at* <http://www.pacer.psc.uscourts.gov> (alleging that MLB has attempted to preclude non-licensed entity from providing player statistics for use in fantasy baseball).

153. *See* C.B.C. Distrib. & Mktg. v. Major League Baseball Advanced Mktg., 443 F. Supp. 2d. 1077, 1081, 1095 (E.D. Mo. 2006) (concluding that use of player names with statistics without license does not violate right of publicity and, at any rate, would be protected First Amendment speech even if it were a violation). Major League Baseball is currently appealing this ruling. *See MLB Says It Will Appeal Fantasy Ruling*, ESPN.COM, Aug. 9, 2006, <http://sports.espn.go.com/mlb/news/story?id=2544949>.

154. *See* Nat’l Basketball Ass’n. v. Motorola, Inc., 105 F.3d 841, 843 (2d Cir. 1997) (rejecting the NBA’s misappropriation claim against Motorola and concluding that the NBA lacked a property right in factual descriptions of its games).

wireless rights from the players’ union in 2005, contends that the players’ rights of publicity are being infringed.<sup>155</sup>

MLB’s challenge to fantasy league data providers raises familiar questions concerning the tension between intellectual property and the First Amendment,<sup>156</sup> as well as whether the right of publicity has become unmoored.<sup>157</sup> In addition to addressing these arguments, courts should also allow the logic of competing competitions to inform their considerations of whether to allow the right of publicity to fetter fantasy leagues.<sup>158</sup> Intellectual property issues often revolve around questions of how to best foster innovation while also avoiding misappropriation. Courts should take into account the value of innovation created by fantasy league fans when weighing the rights of fantasy leagues versus players’ rights of publicity. Additionally, they should consider the competitive benefits fantasy leagues generate as “competing competitions.”<sup>159</sup> While MLB’s intellectual property rights may be entitled to protection, the ability to foreclose

---

155. See Memorandum in Support of MLBAM Motion for Summary Judgment at 1, *C.B.C. Distrib. & Mktg. v. Major League Baseball Advanced Mktg.*, 443 F. Supp. 2d 1077 (E.D. Mo. 2006), available at [www.pacer.psc.uscourts.gov](http://www.pacer.psc.uscourts.gov) (asserting that “the right of publicity prevents [player statistics provider] from using players’ names and identities” without a license).

156. See, e.g., Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 NYU L. REV. 354 (1999); James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33 (2003); Mark A. Lemley & Lawrence Lessig, *The End of End-To-End: Preserving the Architecture of the Internet in the Broadband Era*, 48 UCLA L. REV. 925 (2001); Jessica Litman, *The Sony Paradox*, 55 CASE W. RES. L. REV. 917 (2005).

157. See Stacey L. Dogan & Mark Lemley, *What the Right of Publicity Can Learn from Trademark Law*, 58 STAN. L. REV. 1161 (2006) (arguing that expansion of the right of publicity should be tempered by grounding the right in trademark law principles of source identification and limited by attaching the right mainly to the traditional trademark law harms of confusion and dilution).

158. It is possible that MLB is exempt from the areas of antitrust law concerning this aspect of its business. Professional baseball is exempt from antitrust laws under its judicially-created exemption (save for labor relations, the exception for which Congress has repealed). See *Flood v. Kuhn*, 407 U.S. 258 (1972); *Toolson v. N.Y. Yankees*, 346 U.S. 356 (1953); *Fed. Baseball Club of Balt. v. Nat’l League*, 259 U.S. 200 (1922) (holding that baseball is exempt from the Sherman Act because it is “not a subject of commerce”). However, no case has discussed whether the scope of baseball’s exemption extends to internet distribution of statistics. Furthermore, the judge-made law of baseball’s antitrust exemption should not prevent a court from considering the efficiency consequences of fan innovation when dealing with intellectual property rights.

159. See *MLB Skirmish*, *supra* note 151 (stating that the League’s legal stance “will have unexpected and unwelcome ramifications” and “might . . . discourage the sort of innovation that has made some segments of the economy so vital”).

unlicensed fantasy leagues may fall outside the reasonable scope of such rights. In a reversal of the typical intellectual property dispute, MLB, the intellectual property rights-holder, is arguably using its rights to appropriate the profits from another's innovation.<sup>160</sup>

In addition to the innovation of fantasy leagues, globalization provides another means for externally-driven competing competitions to take place. In an attempt to emulate the success of World Cup soccer, American professional sports leagues in recent years have sought to engage professional hockey and baseball players in competitions against their global counterparts. While the NHL and MLB did not endorse these tournaments enthusiastically at first,<sup>161</sup> both leagues have since relented and now permit national teams to be comprised of players from both their leagues and foreign leagues.<sup>162</sup> In the future, there may be tournaments that match existing professional teams from different nations' leagues against each other, with rosters grouped by team affiliation rather than citizenship.<sup>163</sup> Such tournaments already take place among teams from different national soccer leagues in Europe. The most notable is the Champions League, which matches the top teams from a variety of leagues.<sup>164</sup> Should American pro-

---

160. *Id.* (contrasting the fact that, in the past decade of music industry litigation against online file-sharing, the music industry was trying to protect its core business of selling music, while Major League Baseball's litigation is aimed at seeking revenue from a source that it had not been engaged in and others had worked to develop—selling statistics).

161. See Murray Chass, *Finally a Chance to Find the Real World Champion Past October*, N.Y. TIMES, May 12, 2005, at D4 (describing formation of the World Cup-style "World Baseball Classic" and noting that "[t]he players union for years proposed it, but Major League Baseball, until recent years, was not interested"). A fear of competing competitions may explain the reluctance of league and baseball owners to foster a tournament that could involve external competitors; see also Harvey Araton, *These Games Deserve Better From N.H.L.*, N.Y. TIMES, Feb. 15, 2002, at D2 (describing NHL's grudging allowance of its players to represent their countries for the first time at the 2002 Winter Olympics and noting NHL commissioner's stance that future Olympic participation "is no sure thing").

162. There actually has been an amateur "Baseball World Cup" run by the International Baseball Federation, for 68 years, but it does not include professional players active in Major League Baseball. See International Baseball Federation, *Baseball World Cup*, <http://www.baseball.ch/2003/t/wc/wc.htm> (last visited Oct. 23, 2006).

163. Such games occur now, but only as single-match non-recurring exhibitions that do not seriously threaten existing leagues. Additionally, they appear to be done only with league approval. Tyler Kepner, *Yanks Rediscover Japan 70 Years After First Visit*, N.Y. TIMES, Mar. 28, 2004, at H1 ("On Sunday night, the Yankees play the Giants in an exhibition game that may feel much more important. It is the first of two exhibitions; the other is against the Hanshin Tigers on Monday and is expected to attract more attention here than the series with the Devil Rays.").

164. See SKY SPORTS YEARBOOK, *supra* note 95.

professional sports leagues attempt to wall themselves off from such proposals through exclusionary agreements, the logic of competing competitions suggests that, at minimum, they should not be allowed to justify their actions based on poorly-defined and weakly-supported competitive balance concerns. Furthermore, the value inherent in having competitions compete against each other should compel courts to provide the same regard for future international proposals that antitrust courts provide to entrants to an existing market.

#### **IV. HOW TO DEAL WITH THE COMPETITIVE BALANCE ARGUMENT**

Given the preceding discussion, this Article’s proposal is a simple one: Reject the “competitive balance” justification under the rule of reason. As simple as they are, this prescription and the reasoning supporting it yield three new levels of understanding. First, at the narrowest, the competitive balance argument should no longer avail defendants in sports antitrust cases under the rule of reason. Second, more broadly, future antitrust courts should cast a more skeptical eye toward “aesthetic” arguments like competitive balance, given that there is scant proof of a hard nexus with financial results. Finally, and most broadly, the empirical weakness of the competitive balance argument warrants judicial skepticism beyond antitrust criticisms.

##### **A. Reject Competitive Balance Under the Rule of Reason**

“Competitive balance” is a thirty-year-old argument that should fail. Given the serious doubts that empirical evidence casts on the argument, courts should no longer allow professional sports leagues to argue that their practices should be excused due to their allegedly positive effects on competitive balance. Instead, courts should treat professional sports leagues as they treat other industries involving two-sided markets. As Prof. Stefan Szymanski, a prominent figure in sports economics, points out, “the study of sports economics as a distinct area of research rests on a single issue, the ‘competitive balance problem.’”<sup>165</sup> Similarly, the competitive balance argument explains in large part why antitrust has handled sports with kid gloves. There is no reason why sports leagues should be treated any differently from other industries that require cooperation be-

---

165. Szymanski, *Tilting the Playing Field*, *supra* note 129, at 1.

tween economically competing entities to produce a product,<sup>166</sup> given that these other industries also deal with organized labor.

Antitrust should treat leagues as businesses that offer one product (access to on-the-field competition) in a two-sided market. For example, should a league be a monopoly in that regard, it should enjoy the benefits, such as supra-competitive pricing, that being a monopoly provides. However, one should approach actions to entrench or extend that monopoly in the same manner that antitrust law would analyze similar efforts to take advantage of monopoly status.<sup>167</sup> That is, should fan innovation, international tournaments, or other investors arise to provide competing competitions, predatory or exclusionary acts by professional sports leagues aimed at preventing such competition should be examined under the same rubric that antitrust applies to other industries. That may not be easy; there is certainly disagreement as to how and when a monopolist may thwart a competitor.<sup>168</sup> However, in the absence of a tested argument for treating pro sports leagues otherwise, applying the antitrust treatment afforded other industries seems to be efficient, fair, and logical.

## **B. Exercise Skepticism About Arguments that May Be Aesthetics**

The evidence suggests that competitive balance does not drive fan interests, that it may not be achievable as a result of league policies, and that leagues may not even rationally want to achieve competitive balance, even if they could. How, then, can we explain judges' overwhelming fondness for competitive balance in sports as a "legitimate concern" in antitrust

---

166. Consider the consumer electronics industry, where competing manufacturers often face a dynamic where they must agree on a dominant standard, such as DVD or VHS, or face a public war to invest in a standard that might fail, such as Betamax or Digital Audio Tape. David Pogue, *For DVD Watchers, High-Def Disarray*, N.Y. TIMES, Jan. 1, 2006, at D3 ("The obvious losers are movie fans. They risk buying a \$1,000 player that can play only half of available movies. Worse, when one format finally wins, some customers would have bet on the wrong horse.").

167. See *United States v. Microsoft*, 253 F.3d 34, 70-71 (D.C. Cir. 2001) (affirming conclusion that by foreclosing a "substantial percentage of the available opportunities for browser distribution, Microsoft managed to preserve its monopoly in the market for operating systems" and thus violated Section 2 of the Sherman Act).

168. See Thomas Piraino, *Identifying Monopolists' Illegal Conduct Under the Sherman Act*, 75 N.Y.U. L. REV. 809, 857 (2000) (concluding that Microsoft's design of its operating system to not work with potential competitors' products was an antitrust violation because the only possible reason for such practices is "to extend its monopoly power from the operating system market to the applications markets"). *But see* Patrick Ahern, *Refusals to Deal After Aspen*, 63 ANTITRUST L. J. 153, 154 (1994) (describing a Supreme Court case that appeared to impose on a monopolist a duty to deal with competitors and explaining that among commentators, "most agreed that it was a troublesome case").

law? Of course, one argument could be simply that the data are wrong and judges have the discretion to determine what is right. However, one would hesitate to press that point against the author, subjects, and fans of the non-fiction bestseller *Moneyball*.<sup>169</sup>

Rather, “competitive balance” may simply be a matter of aesthetics. Judges may rely on an intuitive feel for what seems to “make” for a better sports league. The problem with this is that it exhibits a relatively strong “status quo” bias.<sup>170</sup> If one assumes that leagues in the future will exist as they currently do, then for them to consist of evenly-matched teams may make sense. However, this assumption may itself be an error. Judges are no better situated to decide whether league rules promote better sports competition than they are to decide whether blanket licenses for collections of musical compositions promote better-sounding music.<sup>171</sup> Additionally, one should expect that the leagues would make the competitive balance argument so long as it accords with their own rational self-interest. Thus, the assertion that competitive balance makes a league more entertaining may be nothing more than an aesthetic judgment, with no basis in logic or empirical evidence. Though one can “feel” that a league in which the Yankees win five straight World Series is a less optimal one,

---

169. *MONEYBALL* is a 2003 non-fiction bestseller by Michael Lewis that describes how Billy Beane and the Oakland Athletics manage to outguess the rest of MLB in scouting talented players by depending more on data and less on the conventional wisdom of baseball that depends on how a player “looks” or what “tools” he appears to have. See MICHAEL LEWIS, *MONEYBALL: THE ART OF WINNING AN UNFAIR GAME* 31 (2003) (quoting baseball manager focused on data rather than impressionistic scouting based on how a player “looks” as saying “we’re not selling jeans here”). The book itself has spawned a small but prominent literature in American law reviews for its explanation of how data trumped eyewitness observation and how a multimillion dollar enterprise like MLB can evolve a conventional wisdom that appears incorrect and refutable by data and empirical analysis. Richard H. Thaler & Cass R. Sunstein, *Market Efficiency and Rationality: The Peculiar Case of Baseball*, 102 MICH. L. REV. 1390 (2004); Note, *Losing Sight of Hindsight: The Unrealized Tradition of Law and Sabermetrics*, 117 HARV. L. REV. 1703 (2004); see also Paul Caron & Rafael Gely, *What Law Schools Can Learn from Billy Beane and the Oakland Athletics*, 82 TEX. L. REV. 1483 (2003).

170. This bias is akin to the availability heuristic, in which people answer a question of probability or frequency by thinking of salient examples that come to mind. See Cass Sunstein, *Hazardous Heuristics*, 70 U. CHI. L. REV. 751, 752 (2003) (discussing the “availability heuristic” as “probably . . . the most well-known [heuristic] in law”).

171. See *Broadcast Music, Inc. v. Columbia Broad. Sys., Inc.*, 441 U.S. 1, 21-22 (1979) (concluding that a non-exclusive blanket license to the music of all composers handled by an intellectual property rights clearinghouse was not price-fixing but rather a pro-competitive practice and effectively, a “different product” than simply the aggregation of underlying rights).

and though it may be a popular stance, the evidence does not support this view.

### C. Reconsider Similar Arguments Beyond Antitrust

While not free from controversy, antitrust already possesses a body of law dealing with practices aimed at blocking new entrants to a market.<sup>172</sup> It may thus be worthwhile to extend the idea of “competing competitions,” and the superior efficiency and innovation they provide, beyond antitrust.

In particular, new forms of competition within existing sports leagues can generate new consumer welfare through technological changes or changes in the environment in which leagues exist. Specifically, fantasy leagues, international tournaments between teams from different national leagues, or even entirely new competitions yet unimagined, all have the potential to bring structurally static—and possibly monopolistic<sup>173</sup>—professional sports leagues “up to date” with consumer preferences.

It is important to note, though, that one cannot accurately weigh new competing competitions unless one recognizes that, despite their lack of longevity and established doctrinal intellectual property rights, competing competitions can nevertheless represent valuable innovation. Commentators have recognized that intellectual property rights cannot capture the full scope of innovation.<sup>174</sup> Thus, courts should take the claims of new competing competitions seriously if they complain about exclusionary practices that block them.

---

172. See *Lorain Journal v. United States*, 342 U.S. 143, 150 (1951) (condemning incumbent newspaper’s policy forbidding advertisers from buying advertising time from a radio station new to the market). *But see* Hovenkamp, *supra* note 102, § 7.6c (commenting that the Supreme Court in *Lorain Journal* never considered what percentage of the market was actually foreclosed by the newspaper’s practices). See also *Microsoft*, 253 F.3d at 70 (concluding that “a monopolist’s use of exclusive contracts may, in certain circumstances, give rise to [an antitrust] violation even though the contracts foreclose” a relatively small share of the available market).

173. See generally Thomas A. Piraino, *Proposal for Antitrust*, *supra* note 112 (arguing that professional sports leagues are monopolies and their restrictive policies should be scrutinized for a tighter ends-means connection); Stephen F. Ross, *Monopoly Sports Leagues*, 73 MINN. L. REV. 643 (1989) [hereinafter Ross, *Monopoly*] (arguing that American professional sports leagues are monopolies that should be broken up by the government).

174. See, e.g., R. Polk Wagner, *Information Wants to Be Free: Intellectual Property and the Mythologies of Control*, 103 COLUM. L. REV. 995, 1004-05 (2003) (describing phenomena which can generate valuable information that is neither appropriable nor subject to intellectual property rights). While some may use this argument to bolster existing intellectual property rights by noting their positive effects on such unappropriable innovation, *id.*, such innovation is valuable in and of itself, whether it is generated by incumbents or new entrants.

## V. CONCLUSION

In light of the evidence, the competitive balance argument is seriously flawed. The proposal advocated here to remedy the situation is a relatively modest one. Others have suggested more ambitious steps, such as forcibly restructuring American sports leagues.<sup>175</sup> Due to their relatively intrusive nature, these suggestions are unlikely to be implemented; it is difficult to imagine Congress or the federal judiciary changing the landscape of pro sports by fiat. The approach advocated in this Article has the virtue of appearing low-key despite its potential for far-reaching change. By depriving the leagues of one argument in their defense under the rule of reason, it publicly spotlights the fact that little evidence exists to support the competitive balance argument.

Additionally, this proposal seeks to perform “intellectual jiu-jitsu” by channeling the forces of innovation and globalization against the dominance of American professional sports leagues. The rule and reasoning advocated in this Article defends and empowers those who privately find ways to improve the entertainment that the leagues currently provide. Preventing existing leagues from foreclosing new competitions may well promote their continued existence in substantially the same form as today. But, it also forces them to improve their product, rather than depend on restrictive policies that they can no longer justify with vague allusions to “competitive balance.”

---

175. See, e.g., Piraino, *Proposal for Antitrust*, *supra* note 112, at 889 (advocating application of antitrust’s controversial essential facilities doctrine to compel sports teams to admit new teams); Ross, *Monopoly*, *supra* note 173 (advocating AT&T-style government breakup of leagues); Ross & Szymanski, *Open Competition*, *supra* note 97 (advocating implementation of a European soccer-style relegation and promotion system in American professional sports).



# REVITALIZING THE DOCTRINE OF TRADEMARK MISUSE

By William E. Ridgway<sup>†</sup>

## TABLE OF CONTENTS

I. OVERVIEW OF MISUSE.....	1550
A. MISUSE WITHIN PATENT AND COPYRIGHT LAW .....	1550
B. MISUSE WITHIN TRADEMARK LAW.....	1553
1. <i>Early Cases and Unclean Hands</i> .....	1553
2. <i>Modern Decline</i> .....	1554
II. TRANSLATING MISUSE TO TRADEMARK LAW.....	1557
A. OVERVIEW OF TRADEMARK LAW.....	1557
B. DISTINGUISHING TRADEMARK LAW FROM PATENT AND COPYRIGHT LAW .....	1558
1. <i>Trademarks Principally Protect Society Rather than Creators</i> ..	1558
2. <i>Differences in Constitutional Basis for Protection</i> .....	1561
3. <i>Trademark Law's Vague, Fairness-Based Boundaries</i> .....	1562
III. WHAT CONDUCT SHOULD CONSTITUTE TRADEMARK MISUSE?.....	1564
A. ANTICOMPETITIVE AND PUBLIC POLICY APPROACH.....	1565
B. THE PROPOSED APPROACH.....	1566
1. <i>Incorrect Assertion of Rights</i> .....	1567
2. <i>Improper Purpose</i> .....	1571
C. APPLYING THE TEST .....	1576
1. <i>Presumption of an Improper Purpose</i> .....	1576
2. <i>No Presumption of Improper Purpose</i> .....	1581
IV. WHAT IS THE APPROPRIATE REMEDY?.....	1583
A. COURT-BASED EQUITABLE REMEDIES .....	1583
B. STATUTORY-BASED AFFIRMATIVE CLAIMS .....	1585
V. CONCLUSION .....	1588

---

© 2006 William E. Ridgway

<sup>†</sup> Law clerk to Hon. Vaughn R. Walker, United States District Court for the Northern District of California. J.D., Stanford Law School (2006); B.S., Stanford University (2003). The author thanks Mark Lemley, David Olson, and the editors of the *Berkeley Technology Law Journal* for valuable comments. An earlier version of this Article received the International Trademark Association's 2006 Ladas Memorial Award for best student paper on trademark law.

The internet has diversified the public sphere by eliminating many of the barriers to speaking across associational boundaries. But for some categories of speech—namely, criticism and parody—companies have invoked trademark law to restore those barriers. This occurs when a company responds to critique with cease-and-desist letters alleging trademark infringement. Such letters tend to compel compliance from non-infringing speakers because ready access to a lawyer is a barrier the internet has failed to eliminate. Even though such companies' accusations rely on the color of law, few of these disputes make their way into case reporters. Instead, their fate is an obscure demise in trademark law's shadow—a realm where traditional First Amendment defenses are of little use. One could view this state of affairs as an inevitable consequence of granting a trademark right because trademark holders will make mistakes in enforcing their rights. But in tolerating mistakes trademark law should not immunize abuse. To broker these competing interests, this Article urges revitalizing trademark misuse—a doctrine with early roots but without contemporary recognition.

Trademark law traditionally relied on a clear line between trademark users (sellers of goods and services) and its beneficiaries (consumers). But the internet muddies this distinction by rendering pithy, descriptive words—or in trademark parlance, source identifiers—essential to organizing and disseminating information.<sup>1</sup> Indeed, trademark law bears on most of the internet's core features, such as domain names, hyperlinks, term-based searches, and meta-tags.

As a consequence, companies turn to trademark law to bend the internet toward their interests.<sup>2</sup> Much of this behavior is not troublesome. A company's interest in preventing confusion about its product generally aligns with trademark law's underlying purpose to minimize consumer search costs. And yet, as this Article highlights, some of the control over marks wrested by companies under trademark law gives reason for concern, especially for conduct that passes under the court's radar. In short, law that worked, and continues to work, for sellers of goods and services spills into places where it does not, and trademark law's limiting doctrines cannot adequately respond.

---

1. For a similar argument with respect to copyright, see Jessica Litman, *Revising Copyright for the Information Age*, 75 OR. L. REV. 19 (1996) (asserting that copyright law, due to its emphasis on reproduction, affects, and indeed prohibits, most of the activity that takes place on the internet).

2. See, e.g., Dan L. Burk, *Trademark Doctrines for Global Electronic Commerce*, 49 S.C. L. REV. 695, 696 (1998) (discussing a surprising "explosion of Internet-related litigation arising out of trademark disputes").

This Article argues that the doctrine of trademark misuse would provide the needed response to the aforementioned problems, and would do so without unduly impacting the parts of trademark law that function well. Trademark law currently imposes one-way liability, obliging infringers to pay trademark holders for their violations. A doctrine of trademark misuse would revise this regime by permitting liability to run both ways, regulating the behavior of both the potential infringer *and* the trademark holder. Such a regime would better safeguard those who find themselves the target of an abusive trademark holder. Acknowledging that the word “misuse” lacks clear definition—let alone talismanic power—this Article aims to construct a framework for trademark misuse with the intention of clarifying a doctrine often fraught with confusion throughout the law of intellectual property.

This Article proceeds in five parts. Part I briefly describes the history and development of the misuse doctrine by first focusing on its divergent trajectories within patent and copyright law, and then discussing its general failure in the trademark context. Part II identifies the relevant differences between trademark law on the one hand and patent and copyright law on the other. It argues that two core differences—that trademarks protect society rather than creators and that the boundaries of the right depend more upon fairness-based factors—render trademark law more amenable to an equitable doctrine of misuse than patent and copyright law, especially if the doctrine monitors enforcement decisions and curbs overprotection.

Guided by these insights, in Part III the Article maintains that a trademark misuse doctrine should prohibit a trademark holder’s (1) incorrect assertion of rights (2) with an improper purpose, i.e., a purpose unrelated to the adjudication of the claim upon which the assertion of rights is based. Under this test, this Article proposes that factual situations in which a trademark holder’s purpose is unlikely to be related to consumer confusion should raise a presumption of an improper purpose. One situation meriting such a presumption is the case of a trademark holder’s attack on a defendant’s critique or parody, as free speech is the value most conspicuously threatened by trademark law’s newfound relevance in the internet area.

Next, Part III applies this general test to clarify the doctrine’s contours. In particular, conduct raising a presumption of improper purpose under this test includes sending cease-and-desist letters misrepresenting that some activity is infringing to internet users or intermediaries, or using domain name dispute resolution in bad faith to deprive internet users of their domain name, a tactic often called “reverse domain-name hijacking.” Of equal importance, Part III proposes ways to cabin the misuse doctrine in

order to avoid impacting the vast majority of trademark holders who lawfully enforce their rights.

Part IV analyzes potential remedial schemes. After finding misuse's current options generally inadequate, it proposes alternative remedies, including an affirmative cause of action modeled after provisions from the Digital Millennium Copyright Act (DMCA). Finally, Part IV addresses the potential barriers to realizing a doctrine of trademark misuse.

## I. OVERVIEW OF MISUSE

### A. Misuse Within Patent and Copyright Law

What role the doctrine of misuse should play in contemporary intellectual property is still an open question—a status largely attributable to the doctrine's haphazard development and conflicting rationales, first in patent law<sup>3</sup> and later in the area of copyright.<sup>4</sup> Notwithstanding this confusion, the doctrine's basic structure remains fairly constant. Conduct found to constitute misuse typically falls into one of two categories: (1) improper attempts to extend the scope of the right, and (2) violations of antitrust laws.<sup>5</sup> Under this affirmative defense, property holders shown to have misused their right lose the power to enforce their right until the misuse is “purged.”<sup>6</sup> A finding of misuse thus permits an alleged infringer—as well as any third party—to use freely the protected right until the trademark holder discontinues the misuse.

In early decisions, courts consistently rejected a patent misuse defense because it conflicted with their views of property as a natural right.<sup>7</sup> With

---

3. See 6 DONALD S. CHISUM, CHISUM ON PATENTS § 19.04[f][3] (2005) [hereinafter CHISUM] (noting that the predictability of misuse is hindered because “courts have failed to adopt a general theory as to the proper limitations on the exploitation of the patent monopoly”); see also *USM Corp. v. SPS Techs., Inc.* 694 F.2d 505, 512 (7th Cir. 1982) (noting that outside of conventional antitrust principles, the “law is not rich in alternative concepts of monopolistic abuse; and it is rather late in the day to try to develop one without in the process subjecting the rights of patent holders to debilitating uncertainty”).

4. See *infra* text accompanying notes 17-26.

5. See CHISUM, *supra* note 3, § 19.04 (“A patent owner may exploit a patent in an improper manner by violating the antitrust laws or extending the patent beyond its lawful scope.”).

6. *Id.*

7. See George Gordon & Robert J. Hoerner, *Overview and Historical Development of the Misuse Doctrine*, in ABA SECTION OF ANTITRUST LAW, INTELLECTUAL PROPERTY MISUSE: LICENSING AND LITIGATION 4 (2000) (citing for example, *Strait v. National Harrow Co.*, 51 F. 819 (C.C.N.D.Y. 1892), in which the court rejected a misuse-like argument and noted that “[e]ven a gambler, or the keeper of the brothel, cannot be deprived

time, however, courts warmed up to the defense as a way to limit the then-unwieldy doctrine of contributory infringement.<sup>8</sup> But the Supreme Court's seminal pronouncement of an intellectual property misuse doctrine did not happen until the case of *Morton Salt Co. v. G.S. Suppinger Co.*<sup>9</sup> The plaintiff in that case owned the patent on a machine for depositing salt tablets in food cans and leased such machines to commercial canners on the condition that the lessees only buy salt from their subsidiary.<sup>10</sup> The Court ruled it per se illegal for a patent holder to tie the sale of a patented good to the sale of a nonpatented good, and held that such conduct, until purged, precluded enforcement of the patent.<sup>11</sup> To support this drastic result, the Court drew on the sweeping principle that it "may appropriately withhold [its] aid where the plaintiff is using the right asserted contrary to the public interest."<sup>12</sup>

In the two subsequent *Mercoid* decisions,<sup>13</sup> the Supreme Court extended its new doctrine in a way that effectively eliminated a patent holder's ability to prevent a competitor from making and selling a component that had no use other than as part of a patented invention. In response, Congress intervened to resuscitate the doctrine of contributory infringement and to restrict the scope of patent misuse.<sup>14</sup> This response set off a trend: since *Mercoid*, courts and Congress have continued to pare down the doctrine of misuse within patent law.<sup>15</sup> Academic commentary largely encourages this narrowing as well, as it finds misuse too non-economic for

---

of his property because he is an obnoxious person or a criminal"); see also Adam Mossoff, *Who Cares What Thomas Jefferson Thought About Patents? Reevaluating the Patent 'Privilege' in Historical Context*, 92 CORNELL L. REV. \_\_ (forthcoming 2007), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=892062](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=892062).

8. Gordon & Hoerner, *supra* note 7, at 6.

9. 314 U.S. 488 (1942). Earlier, the Court applied misuse-like logic in *Carbice Corp. of America v. American Patents Development Corp.*, 283 U.S. 27, 33 (1931) (recognizing that "[c]ontrol over the supply of such unpatented material is beyond the scope of the patentee's monopoly" and refusing to grant relief to a patent holder who engages in such a practice).

10. *Morton Salt*, 314 U.S. at 490-91.

11. *Id.* at 493.

12. *Id.* at 492.

13. *Mercoid Corp. v. Mid-Continent Co.*, 320 U.S. 661, 665 (1944); *Mercoid Corp. v. Minneapolis-Honeywell*, 320 U.S. 680, 684 (1944).

14. See Gordon & Hoerner, *supra* note 7, at 15.

15. See Patent Misuse Reform Act of 1988, Pub. L. No. 100-703 § 201, 102 Stat. 4674 (codified as amended at 35 U.S.C. § 271(d) (2003)); see also *Ill. Tool Works, Inc., v. Indep. Ink, Inc.*, 126 S.Ct. 1281 (2006) (holding that, in antitrust cases, courts should not presume that a patent confers market power upon the patentee).

antitrust and too unpredictable for patent law—a veritable black sheep of two doctrinal families.<sup>16</sup>

Yet, misuse's decline within patent law contrasts its revival within copyright, led by the Fourth Circuit in *Lasercomb America, Inc. v. Reynolds*.<sup>17</sup> In that case, the defendant, Reynolds, alleged misuse based on a provision in Lasercomb's standard licensing agreement that barred the licensee from creating a competing product for a period of one-hundred years.<sup>18</sup> While the court recognized that "uncertainty engulfs the 'misuse of copyright' defense," it was ultimately "persuaded . . . that a misuse of copyright defense is inherent in the law of copyright just as a misuse of patent defense is inherent in patent law."<sup>19</sup> Critically, the court stressed the equitable rather than anticompetitive basis for misuse: "[t]he question is not whether the copyright is being used in a manner violative of antitrust law . . . but whether the copyright is being used in a manner violative of the public policy embodied in the grant of a copyright."<sup>20</sup> As a result, *Lasercomb* recognized copyright misuse as a functioning doctrine, but only vaguely defined the doctrine's scope.

Other circuits have followed *Lasercomb*'s lead, and in doing so, have proffered unique interpretations of this nascent doctrine. In *Alcatel USA, Inc. v. DGI Technologies, Inc.*, the Fifth Circuit ruled that Alcatel had misused a software copyright by attempting to extend its copyright to cover uncopyrighted expansion cards.<sup>21</sup> The plaintiff copyright owner tried to distinguish *Lasercomb* by noting that its contractual restriction with the defendant did not "prohibit the independent development of compatible operating system software."<sup>22</sup> But the court rejected this argument, noting that any restriction of copying rendered the situation, in effect, identical to *Lasercomb*.<sup>23</sup>

The Ninth Circuit adopted an equitable doctrine of copyright misuse in *Practice Management Information Corp. v. American Medical Association*.<sup>24</sup> In this case, the defendant's copyright, which covered a guide to a

---

16. See, e.g., Mark A. Lemley, Comment, *The Economic Irrationality of the Patent Misuse Doctrine*, 78 CALIF. L. REV. 1599 (1990); Note, *Is the Patent Misuse Doctrine Obsolete?*, 110 HARV. L. REV. 1922, 1925 (1997) (summarizing the "[n]umerous authorities [that] contend that the misuse doctrine should be abolished").

17. 911 F.2d 970 (4th Cir. 1990).

18. *Id.* at 972-73.

19. *Id.*

20. *Id.* at 978.

21. 166 F.3d 772 (5th Cir. 1999).

22. *Id.* at 793.

23. *Id.*

24. 121 F.3d 516 (9th Cir. 1997).

coding system for medical professionals to identify medical procedures, was licensed on a non-exclusive, royalty-free basis in exchange for the agreement to use the guide exclusively.<sup>25</sup> In holding that the license provision requiring exclusive use constituted copyright misuse, the Ninth Circuit determined that any welfare-enhancing explanations did not overcome the “adverse effects of the licensing agreement,” which resulted in “substantial and unfair advantage over [defendant’s] competitors.”<sup>26</sup>

In sum, the recent growth of misuse within copyright law contrasts its decline in patent law. This trajectory reflects a shifting perspective on misuse’s proper role: courts realize that antitrust functions adequately without misuse’s assistance, so they instead use the doctrine to remedy other perceived problems within intellectual property. A doctrine of trademark misuse, at least of the kind argued for in this Article, comports with this shifting perspective.

## B. Misuse Within Trademark Law

### 1. *Early Cases and Unclean Hands*

Courts first applied the doctrine of misuse to a trademark dispute in 1883, long before its application to other intellectual property. Despite its vintage, trademark misuse garners little acceptance from courts or commentators today. In an 1883 case, *Manhattan Medicine Co. v. Wood*, the plaintiff sought an injunction against a competitor for imitating its popular medical product, “Atwood’s Vegetable Physical Jaundice Bitters.”<sup>27</sup> Despite the defendant’s nearly identical packaging, the Court declined to enforce plaintiff’s right because its products claimed falsely to be “manufactured by Moses Atwood in Georgetown, Massachusetts.”<sup>28</sup> According to the Court, because these statements misled the public, a court of equity could not lend its aid to this fraud by enforcing the action.<sup>29</sup>

The Supreme Court similarly withheld relief from a dishonest trademark owner several decades later in *Clinton E. Worden & Co. v. California Fig Syrup Co.*<sup>30</sup> The Ninth Circuit agreed with California Fig Syrup Company that the defendant’s Fig Syrup impermissibly imitated plaintiff’s core product, “Syrup of Figs.”<sup>31</sup> But the Supreme Court reversed because the product contained few, if any, figs; in fact, both products merely con-

---

25. *Id.* at 516.

26. *Id.* at 521.

27. 108 U.S. 218 (1883).

28. *Id.* at 222.

29. *Id.*

30. 187 U.S. 516 (1903).

31. *See id.* at 521.

sisted of the “ordinary and well-known laxative called senna.”<sup>32</sup> This analysis led the Court to conclude:

[W]hen the owner of a trade mark applies for an injunction to restrain the defendant from injuring his property by making false representations to the public, it is essential that the plaintiff should not in his trade mark, or in his advertisements and business, be himself guilty of any false or misleading representation . . . [I]f the plaintiff makes any material false statement in connection with the property which he seeks to protect, he loses his right to claim the assistance of a court of equity . . .<sup>33</sup>

Unfortunately, the Court never followed through on its promise to proscribe “any material false statement[s] in connection with [a mark].”<sup>34</sup> Instead of forming the basis for a trademark misuse doctrine of the kind recommended in this Article, this line of cases never extended beyond claims for mislabeling a trademarked product.<sup>35</sup>

## 2. *Modern Decline*

In contemporary cases, courts rarely apply trademark misuse to anti-competitive conduct, and even less frequently do they resort to equitable principles.<sup>36</sup> One arguable exception is *United States Jaycees v. Cedar Rapids Jaycees*,<sup>37</sup> in which the national Jaycees organization sued a local chapter for continuing its use of the “Jaycees” trademark after it had terminated the chapter for admitting women into membership. By the time the case reached the Eighth Circuit, the national Jaycees had already amended its by-laws to allow women as members, but it nonetheless persisted in the lawsuit.<sup>38</sup> The court characterized the national Jaycees’s case as essentially declaring: “It is our trademark; we have a right to choose who uses it, even if we do so arbitrarily and vindictively, and the court

---

32. *Id.* at 520.

33. *Id.* at 528.

34. *Id.*

35. This case law was eventually embodied in the Lanham Act, which provides an exception to incontestability that includes intentionally false and misleading designations of origin. 15 U.S.C. §1115(b) (2000).

36. Brett Frischmann & Dan Moylan, *The Evolving Common Law Doctrine of Copyright Misuse: A Unified Theory and Its Application to Software*, 15 BERKELEY TECH. L.J. 865, 867 n.5 (2000) (“Trademark misuse today resembles copyright misuse twenty years ago: it has been raised as a defense in trademark infringement cases but is not widely recognized.”).

37. 794 F.2d 379 (8th Cir. 1986).

38. *Id.* at 381.

must lend a hand.”<sup>39</sup> While the court acknowledged a trademark holder’s general right to choose who uses its trademark, the court declined to enforce those choices when “the only purpose to be served is punishment of an otherwise productive and conforming member simply because the member was on the prevailing side in a past internal policy dispute.”<sup>40</sup> But in its discussion, the court never mentioned the term “misuse”; rather, the court relied on its power to grant injunctions in accordance with “principles of equity.”<sup>41</sup> Hence, the *Jaycees* court drew upon the logic of misuse without explicitly invoking the doctrine.

Most assertions of trademark misuse by defendants instead derive from antitrust principles.<sup>42</sup> In *Carl Zeiss Stiftung v. V.E.B. Carl Zeiss, Jena*,<sup>43</sup> for example, plaintiffs allegedly misused their trademarks through price fixing and tying arrangements.<sup>44</sup> While the court recognized its powers to deny enforcement, it limited such a defense to antitrust violations in which the mark itself “has been the basic and fundamental vehicle required and used to accomplish the violations.”<sup>45</sup> The court reached this conclusion by distinguishing between patents and trademarks, emphasizing that the former entail limited monopolies whereas the latter provide mere rights to prevent confusion and thus pose little threat to free competition.<sup>46</sup>

A typical antitrust-based trademark misuse claim involves the so-called “tie in” of a trademark to other goods or services. For instance, in *Queen City Pizza, Inc. v. Domino’s Pizza, Inc.*,<sup>47</sup> franchisees charged that Domino’s Pizza misused its trademark when it required franchisees to purchase as much as 90 percent of their supplies from Domino’s at supra-competitive prices.<sup>48</sup> The court denied relief, though, because it defined the relevant aftermarket for antitrust purposes to include all franchises poten-

---

39. *Id.* at 383.

40. *Id.*

41. *See id.* at 382 (citing 15 U.S.C. § 1116(a)).

42. *See, e.g.,* *Estee Lauder, Inc v. The Fragrance Counter, Inc.*, 189 F.R.D. 269, 272 (S.D.N.Y. 1999) (referring to trademark misuse as “the antitrust misuse defense”).

43. 298 F. Supp. 1309 (S.D.N.Y. 1969).

44. *Id.* at 1311.

45. *Id.* at 1315.

46. *See id.* at 1314. Several courts have followed *Zeiss*. *See, e.g.,* *Union Carbide Corp. v. Ever-Ready Inc.*, 531 F.2d 366, 389 (7th Cir. 1976); *Helene Curtis Indus. v. Church & Dwight Co.*, 560 F.2d 1325 (7th Cir. 1977); *Coca-Cola Co. v. Howard Johnson Co.*, 386 F. Supp. 330 (N.D. Ga. 1974).

47. 124 F.3d 430 (3d Cir. 1997).

48. *Id.* at 434.

tially available to the franchisee, for which Domino's Pizza did not maintain market power.<sup>49</sup>

A district court in Georgia, however, distinguished the *Queen City Pizza* decision and protected franchisees against supposed trademark misuse. In *Collins v. International Dairy Queen, Inc.*,<sup>50</sup> the court scrutinized Dairy Queen's franchise agreements because they obliged Dairy Queen to consider requests to use other suppliers.<sup>51</sup> In view of evidence that Dairy Queen, in fact, never honored such requests, the court rejected Dairy Queen's contention that the franchises could simply compare costs ex ante.<sup>52</sup> Rather, such conduct was said to lock franchisees into a tying relationship.<sup>53</sup>

Finally, two recent cases extensively discuss the origin and rationale of trademark misuse before ultimately rejecting any doctrinal expansion. In *Juno Online Services, L.P. v. Juno Lighting, Inc.*,<sup>54</sup> the court examined the doctrine within the context of a dispute over the domain name "juno.com." Juno Online registered the name to provide free e-mail services.<sup>55</sup> Then, Juno Lighting, the owner of the Juno mark, attempted to cancel the domain name and registered "juno-lighting.com."<sup>56</sup> Juno Online brought suit claiming, in part, that Juno Lighting's tactics constituted trademark misuse.<sup>57</sup> The court examined trademark misuse's history with skepticism and rejected Juno Online's affirmative claim of misuse as lacking precedent.<sup>58</sup>

Similarly, in *Northwestern Corp. v. Gabriel Manufacturing, Co.*,<sup>59</sup> the court rebuffed a trademark misuse claim, this time asserted by the defendant, Ashland, who was allegedly selling inferior knock-offs of Northwestern's gumball machines. The court dismissed Ashland's argument by

---

49. *Id.* at 442.

50. 980 F. Supp. 1252 (M.D. Ga. 1997).

51. *Id.* at 1259-60.

52. *Id.*

53. *Id.* For a discussion of the economics of franchise cases, see Benjamin Klein & Lester F. Saft, *The Law and Economics of Franchise Tying Contracts*, 28 J. L. & ECON. 345 (1985).

54. 979 F. Supp. 684 (N.D. Ill. 1997).

55. *Id.* at 685-86.

56. *Id.* at 686.

57. *Id.*

58. *Id.* at 687-91. Other courts similarly conclude that trademark misuse is only an affirmative defense to an infringement claim; it is not an independent cause of action. See, e.g., *Ford Motor Co. v. Obsolete Ford Parts, Inc.*, 318 F. Supp. 2d 516, 521 (E.D. Mich. 2004) ("The court declines to announce or create an independent cause of action for trademark misuse"); *Dunn Computer Corp. v. Loudcloud, Inc.*, 133 F. Supp. 2d 823 (E.D. Va. 2001) (holding that plaintiff's declaratory misuse claim was unripe).

59. 48 U.S.P.Q.2d 1902 (N.D. Ill. 1998).

first remarking that “the defense of trademark misuse is a phantom defense,”<sup>60</sup> and then characterizing Ashland’s argument as requiring trademark holders to be “model citizens” or else risk “having [their] argument fall upon deaf ears.”<sup>61</sup> To the court, a broad conception of misuse “defie[d] common sense”; nonetheless, the court sanctioned the doctrine’s use if a trademark owner “somehow [does] violence to the public policy which establishes [the] trademark right,” namely, using a trademark to misrepresent a product.<sup>62</sup>

In sum, the doctrine of misuse is rarely applied in trademark law. For anticompetitive conduct, courts recognize the doctrine but seldom decide cases on this basis because trademarks infrequently generate market power. An equity-based doctrine is even less developed: courts occasionally exercise discretion to withhold injunctive relief, but have declined to recognize misuse as an independent doctrine.

## II. TRANSLATING MISUSE TO TRADEMARK LAW

Misuse’s checkered history and confused theoretical underpinning in patent and copyright law should caution against expanding its application, or at least suggest we not blindly transfer the doctrine to the trademark context. This Part examines the relevant differences between trademark law on the one hand and patent and copyright law on the other to help determine the appropriate contours of a trademark misuse doctrine.

### A. Overview of Trademark Law

A trademark is a word, name, symbol, or device that is used in trade with goods to indicate the source of the goods and to distinguish them from the goods of others.<sup>63</sup> Regulating such marks protects against consumer confusion about the source, sponsorship, or approval of goods or services or commercial activities of another. To do so, it makes actionable the use of another’s trademark in ways that lead to a “likelihood of confusion.”<sup>64</sup> Trademark law also protects owners of highly distinctive or famous marks against dilution or a tarnishing of their image even absent a likelihood of confusion.<sup>65</sup>

---

60. *Id.* at 1907.

61. *Id.*

62. *Id.* at 1909.

63. See JANE C. GINSBURG ET AL., TRADEMARK AND UNFAIR COMPETITION LAW: CASES AND MATERIALS 45 (3d ed. 2001).

64. *Id.* at 47.

65. *Id.*

Trademarks are intended to save time for consumers, who can rely on distinctive marks as proxies for quality and assess the goods and services they desire without repeatedly undertaking independent research and investigation. Granting limited property rights in a trademark additionally provides producers with economic incentives to create and maintain goodwill toward their marks because it ensures that this goodwill accrues exclusively to the trademark holder's benefit. Accordingly, trademark law primarily reduces consumer decision-making costs, and, as a byproduct, promotes the production of quality goods or services.<sup>66</sup>

## **B. Distinguishing Trademark Law from Patent and Copyright Law**

Trademarks differ from their intellectual property counterparts, patents and copyrights, in several key ways that render trademark law even more amenable to a misuse doctrine than patent and copyright law. These differences relate to the beneficiaries of the protection, the underlying constitutional basis for protection, and the nature of the right's boundaries.

### *1. Trademarks Principally Protect Society Rather than Creators*

Trademark law's underlying policy is critically distinct. Patents and copyrights remedy a public goods problem by providing limited exclusivities in return for the creation of new ideas and expression. Though these exclusivities encourage invention, they also impose societal costs in the form of supracompetitive pricing and potentially delayed subsequent creation.<sup>67</sup> Hence, these intellectual property systems strive to balance creator incentives with the resulting societal costs.<sup>68</sup> Trademarks broker a fundamentally distinct balance. On the cost side, trademarks are not unique: they impose exclusionary rights that narrow the public domain. On the benefit side, trademarks serve society more directly than other intellectual property regimes by protecting against consumer confusion.<sup>69</sup> But contrary

---

66. See, e.g., *Park-N-Fly, Inc. v. Dollar Park & Fly, Inc.*, 469 U.S. 189, 197-98 (1985).

67. Cf. Michael Heller & Rebecca Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 698-99 (1998) (stating that in order to encourage innovation with respect to biomedical technology, monopolies are granted despite the possibility of restricted use and increased costs); see also Michael Heller, *The Tragedy of the Anticommons: Property in Transition from Marx to Markets*, 111 HARV. L. REV. 621, 673-75 (1998) (describing cost balancing for the creation of effective rights of exclusion in scarce resources under anticommons property doctrines).

68. ROBERT P. MERGES ET. AL., *INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE* 15 (2003).

69. *Nw. Corp. v. Gabriel Mfg., Co.*, 48 U.S.P.Q.2d 1902, 1908 (N.D. Ill. 1998). In the legislative history of the statute, Congress expressed its intention to protect both the

to patents and copyrights, the property-holders' interests are only secondary because they do not face a public goods problem.<sup>70</sup>

Trademarks therefore impose costs similar to those that other forms of intellectual property impose for the sake of a distinct kind of benefit. This difference in the benefits ledger exacerbates the consequences of overprotection. For patents and copyrights, overprotection results when the marginal societal costs exceed the marginal benefits to the inventor. But at least overprotection yields benefits to the inventor, which tends to mitigate the costs. For trademarks, overprotection is virtually all costs: finding infringement for a non-confusing source does not advance trademark's purpose of protecting against consumer confusion whatsoever.<sup>71</sup> Although overprotection of a trademark may benefit the property holder, this windfall is immaterial to the underlying policy.

This difference in benefits appears to cut the other way with respect to underprotection, as withholding the intellectual property right for trademarks imposes costs on society at large rather than individual inventors or authors. For trademarks, if other societal interests trump the right, consumer confusion may result, whereas for copyrights and patents, the creator potentially loses supracompetitive profits. At first glance, this distinction may seem to counsel against extending misuse to trademarks because the costs spill to society at large. While this is a critical distinction that resurfaces in the Section discussing possible remedial schemes, the outcome may not be so clear-cut, at least from a dynamic perspective. For patents and copyrights, encouraging the creation of public goods requires an *ex ante* reliance that courts will uphold the rights *ex post*. Withholding the right undermines this reliance, as the Court noted in *Harper & Row*:

To propose that fair use be imposed whenever the social value [of dissemination] . . . outweighs any detriment to the artist, would be to propose depriving copyright owners of their right in

---

consumer and the trademark owner. Senate Report No. 79-1333 at 3 (1946). In protecting the consumer, Congress stated that it enacted this legislation "so [the public] may be confident that, in purchasing a product bearing a particular trademark which it favorably knows, it will get the product which it asks for and wants to get." *Id.* In addition, Congress stated that the trademark owner's expenditure of "energy, time, and money in presenting to the public the product" should be protected from "misappropriation by pirates and cheats." *Id.*

70. For a discussion of the public goods problem in intellectual property, see William E. Ridgway, Note, *Realizing Two-Tiered Innovation Policy Through Drug Regulation*, 58 STAN. L. REV. 1221, 1236-37 (2006).

71. Of course, the existence of consumer confusion is not entirely all-or-nothing, but the point remains that the relevant benefits quickly drop-off, especially when compared to the benefits for patents and copyrights.

the property precisely when they encounter those users who could afford to pay for it.<sup>72</sup>

From a dynamic perspective, withholding a trademark right may similarly destroy reliance interests of the consumer.<sup>73</sup> Yet consumers may not rely with the same level of vigilance: i.e., as opposed to prospective patent and copyright holders, consumers are less likely to understand the intellectual property regime and thereby alter their conduct in light of its inconsistent application.

Some scholars worry that withholding protection injures the trademark holder's reliance interest, eliminating incentives to develop valuable trademarks in the first place.<sup>74</sup> But this blurs trademark policy with patent and copyright policy. As mentioned previously, distinctive marks do not suffer any kind of public goods problem: their creation is nearly costless and, to a large extent, they are their own reward. Indeed, to reach a doomsday scenario of a world devoid of brands would require significant under-protection.

Finally, patents and copyrights better link enforcement incentives with their underlying policy. That is, patents and copyrights confer private benefits to creators, and these benefits form the basis of the property holders' decision whether to enforce. For these exclusionary rights, the system intends that property holders exploit whatever degree of market power they might gain thereby as an incentive.<sup>75</sup> Moreover, weaknesses in the courts' institutional competence suggest deferring to property-holders with respect to this decision. By contrast, trademark law disconnects this relationship: private property holders must enforce the rights associated with their marks to promote societal interests. This policy may cut both ways: as much as trademark holders over-enforce to extract private gains, they may also under-enforce in light of particularly high expected litigation costs.<sup>76</sup> But this disconnect renders judicial oversight of enforcement decisions more essential within the trademark context.<sup>77</sup>

---

72. *Harper & Row, Publishers, Inc. v. Nat'l Enters.*, 471 U.S. 539, 559 (1985) (internal quotations omitted).

73. Societal reliance only makes it more likely that confusion will result when source-identifiers are similar.

74. *See, e.g.*, WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* 270 (2003).

75. *Bonito Boats, Inc. v. Thunder Craft Boats, Inc.*, 489 U.S. 141, 150-51 (1989).

76. Trademark law partially solves the under-enforcement problem by adopting liberal standing requirements. *See* ROBERT D. BLAIR & THOMAS F. COTTER, *INTELLECTUAL PROPERTY: ECONOMIC AND LEGAL DIMENSIONS OF RIGHTS AND REMEDIES* 185 (2005) ("Because the ultimate beneficiaries of trademark rights are not trademark owners, a

## 2. *Differences in Constitutional Basis for Protection*

When litigants invoke comparisons to copyright and patent law, as enumerated in Section II.B.1 above, courts or commentators almost invariably distinguish trademark by citing patent and copyright law's explicit constitutional origin, in contrast to trademark law's indirect basis in the Commerce Clause.<sup>78</sup> Two inferences allegedly follow from this distinction: (1) that courts should not translate doctrines like misuse into trademark law,<sup>79</sup> and (2) that legislatures retain greater control over trademark law's contours.<sup>80</sup> Under the first inference, courts distinguish trademark law for the wrong reasons. As emphasized above, the policies underlying trademark law are unique within intellectual property law; but among those differences, constitutional origin seems the least relevant. The second inference ignores the fact that equitable principles have always been an important part of American common law, even trademark law.<sup>81</sup>

To the contrary, origin may militate in favor of an equitable misuse doctrine that emphasizes First Amendment values. Within First Amend-

---

somewhat looser application of standing rules may fit the law of trademarks and unfair competition better than it would the law of patents and copyrights.”).

77. This parallels other regimes that use a so-called private attorney general, for which scholars urge heightened judicial scrutiny. See STEVE SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW, 391 n.26 (2004) (observing that although the “the private incentive to bring suit is fundamentally misaligned with the socially optimal incentive to do so,” many of the forces that drive a wedge between private and social incentives to litigate could operate in either direction).

78. See, e.g., *Nw. Corp. v. Gabriel Mfg., Co.*, 48 U.S.P.Q.2d 1902, 1908 (N.D. Ill. 1998) (“The policy underpinning trademark protection is not a constitutional mandate, such as the one which creates patents.”); Stephen J. Davidson & Nicole A. English, *Applying the Trademark Misuse Doctrine to Domain Name Disputes*, COMPUTER LAWYER, Aug. 1996, available at <http://www.cla.org/trademark%20misuse.pdf>.

79. See, e.g., *Nw. Corp.*, 48 U.S.P.Q.2d at 1908. (“[I]t is inappropriate to predicate trademark misuse upon the same anti-competitive practices which comprise patent misuse.”).

80. See Kathryn Judge, Note, *Rethinking Copyright Misuse*, 57 STAN. L. REV. 901, 912-13 (2004). (“The structure of the constitutional grant of power embodied in the Copyright Clause, by specifying a purpose as well as a means, arguably creates a heightened duty for courts to ensure the statutory scheme is implemented in a way that furthers the constitutionally defined purpose.”).

81. The Lanham Act expressly provides that even infringement claims involving incontestable marks are subject to equitable defenses, 15 U.S.C. § 1115(b)(8), and the Act requires a balancing of equities when determining whether to enjoin an otherwise infringing act. 15 U.S.C. § 1116(a) (Lanham Act provides that court may grant injunctions “according to principles of equity”); see also 1A J. GILSON, TRADEMARK PROTECTION AND PRACTICE 8.12[13][a], at 8-295 to 8-296 (1990 ed.) (“[a] trademark plaintiff with unclean hands is one whose conduct relative to his mark has been so illegal or unconscionable that the court will refuse to hear him”).

ment jurisprudence, one mode of argument, which Eugene Volokh calls the “constitutional tension method,” identifies certain values that the Constitution explicitly protects and then suggests that the Constitution’s free speech guarantee must sometimes yield to these values.<sup>82</sup> Of course, this method is not determinative; the Constitution contains several “values,” such as democracy, private property, and equality, to which the First Amendment cannot automatically defer without eviscerating free speech. Rather, the constitutional origins of patent and copyright law simply weigh in favor of less exacting First Amendment scrutiny.<sup>83</sup>

### 3. *Trademark Law’s Vague, Fairness-Based Boundaries*

Due to its intangible nature, the boundaries of intellectual property vary more than those of real property. For example, claim construction is notoriously inconsistent in patent law; copyright requires the fact-finder to make a difficult subjective decision whether the defendant unlawfully appropriated the plaintiff’s expressive work.<sup>84</sup> Particular doctrines within intellectual property only exacerbate the problem, such as the doctrine of equivalents and the idea-expression dichotomy. But even these patent and copyright doctrines appreciate the benefit of extensive explication in the case-law and relatively stable, albeit vague, standards. Trademark’s trajectory is somewhat different: while its boundaries are similarly vague,<sup>85</sup> they continue to expand with only meager limiting principles.<sup>86</sup>

Congress expanded the ambit of trademark protection in 1996 with the passage of the Federal Trademark Dilution Act,<sup>87</sup> which prohibits the “commercial use in commerce” of a “famous” mark, if that use causes

---

82. Eugene Volokh, *Freedom of Speech and the Constitutional Tension Method*, 3 U. CHI. L. SCH. ROUNDTABLE 223, 224 (1996).

83. There is some evidence that this method affects First Amendment doctrine within patent and copyright law. In *Zacchini v. Scripps-Howard Broadcasting Co.*, the Supreme Court applied deferential free speech scrutiny because the state was encouraging investment similar to the policies that underlie patent and copyright protection. 433 U.S. 562, 569 (1977); see also *id.* at 577 (noting that Constitution doesn’t “prevent Ohio from making a similar choice here in deciding to protect the entertainer’s incentive”).

84. PAUL GOLDSTEIN, *COPYRIGHT, PATENT, TRADEMARK AND RELATED STATE DOCTRINES* 750 (5th ed. 2002).

85. Likelihood of confusion has developed as the cornerstone of the infringement inquiry and is determined by courts through a case-by-case application of a multi-factor balancing test.

86. See, e.g., Jennifer E. Rothman, *Initial Interest Confusion: Standing at the Crossroads of Trademark Law*, 27 CARDOZO L. REV. 105, 108 (2005); see also Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777, 784-86 (2004).

87. Lanham Act, § 43, codified at 15 U.S.C. §§ 1125, 1127.

“the lessening of the capacity” of the mark “to distinguish goods or services.”<sup>88</sup> Compounding this statutory expansion, recent rejuvenation of the court-created doctrine of “initial interest confusion” permits findings of trademark infringement solely on the basis that a consumer might be “interested,” “attracted,” or “distracted” by a competitor’s product or service.<sup>89</sup>

Some scholars worry that these recent additions unmoor trademark from its original purpose of protecting against consumer confusion.<sup>90</sup> They characterize this shift as the propertization of trademark since it empowers trademark owners to capture *all* value derived from the mark, notably contrary to Justice Holmes’s directive that “property, a creation of law, does not arise from value.”<sup>91</sup> While this account appears correct, it misconstrues one important respect: the recent additions do not facilitate property-like predictability. Instead, courts tend to couch trademark analysis in terms of fairness and a concern with “free-riding.”<sup>92</sup>

In response to these changes, several commentators suggest retreating in some form or another.<sup>93</sup> Misuse, by contrast, offers a fundamentally different approach: concede that trademark law entails vague, fairness-based

---

88. Also, in 1999 Congress further amended the Lanham Act to add a new Section 43(d), which provides for statutory damages when a cybersquatter registers another’s trademark as its domain name with the intent to traffic in the domain name.

89. See sources cited *supra* note 86.

90. See, e.g., Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687 (1999); Sara Stadler Nelson, *The Wages of Ubiquity in Trademark Law*, 88 IOWA L. REV. 731 (2003); Hannibal Travis, *The Battle for Mind-share: The Emerging Consensus that the First Amendment Protects Corporate Criticism and Parody on the Internet*, 10 VA. J.L. & TECH. 3 (2005) (contending that trademark rights should be restricted to policing commercial competition, rather than non-commercial internet speech).

91. *Int’l News Serv. v. Assoc. Press*, 248 U.S. 215, 246 (1918) (Holmes, J., concurring). Rochelle Dreyfuss similarly characterizes this instinct as “if value, then right.” Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 297, 405 (1990).

92. See generally Monica Hof Wallace, *Using the Past to Predict the Future: Refocusing the Analysis of a Federal Dilution Claim*, 73 U. CIN. L. REV. 945 (2005) (characterizing the recent changes as rendering trademark “unwieldy and unpredictable”). See also David J. Franklyn, *Debunking Dilution Doctrine: Toward a Coherent Theory of the Anti-Free-Rider Principle in American Trademark Law*, 56 HASTINGS L.J. 117, 118 (2004) (arguing that plaintiff success rate in dilution claims is “due largely to the fact that judges and juries seek to . . . punish free-riding.”). Note that some European countries explicitly created cause of action for “unfair advantage.” *Id.* at 120.

93. See sources cited *supra* note 90; see also Clarisa Long, *Dilution*, 106 COLUM. L. REV. 1029 (2006) (finding that dilution has not been as powerful a theory as commentators expected because courts have been unwilling to grant injunctions for dilution).

rights, but impose a countervailing doctrine to render the fairness interests symmetrical.<sup>94</sup> That is, if trademark protection now considers the fairness to the property-holders, it should also integrate fairness to potential infringers and society in general. This symmetry is pervasive in property law, which imposes all kinds of fairness-based obligations on property owners.<sup>95</sup> Part III describes a proposal as to how to broker this balance. The conclusion thus far is that trademark law's recent growth suggests a need for misuse as an equitable counterbalance, perhaps even more so than in the areas of patent or copyright law.

### III. WHAT CONDUCT SHOULD CONSTITUTE TRADEMARK MISUSE?

As mentioned at the outset, the doctrine of misuse comes in two varieties: an antitrust approach, which hones in on anticompetitive conduct, and a public policy approach, which scrutinizes behavior that is "violative of the public policy embodied in the grant" of the intellectual property right.<sup>96</sup>

For reasons discussed below, if limited to these varieties, misuse is either narrow and redundant with antitrust law or broad to the point that it lacks basic definition. Neither is an acceptable option for trademark misuse. In particular, discarding the antitrust approach should not lead courts into a doctrinal abyss, as it has in copyright law.<sup>97</sup> Instead, for a public policy approach to function effectively, it must have an identifiable scope and clear underlying justification.

---

94. This is especially true considering trademark offers little in the way of countervailing interests such as a broad fair use doctrine.

95. Michael A. Carrier, *Cabining Intellectual Property Through a Property Paradigm*, 54 DUKE L.J. 1, 75 (2004); Joseph William Singer, *The Reliance Interest in Property*, 40 STAN. L. REV. 611, 641 n. 108 (1988).

96. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 978 (4th Cir. 1990). The dichotomy is a bit puzzling; the former approach appears to be lesser-included component of the latter. After all, antitrust policy itself is no doubt a "public policy" embodied in the grant of an intellectual property right.

97. See Frischmann & Moylan, *supra* note 36 (noting that the case law and legal commentary regarding copyright misuse "seem in disarray"). The assortment of theories that may underlie the copyright misuse doctrine is reflected in recent lawsuits concerning this doctrine. See, e.g., Complaint at ¶ 122-34, *Carol Loeb Shloss v. Joyce Estate*, 06-3718 (N.D. Cal 2006), available at <http://cyberlaw.stanford.edu/Complaint%20Endorsed%20Filed%206-12-06.pdf>.

### A. Anticompetitive and Public Policy Approach

Under the traditional approach, misuse would target anticompetitive conduct by a trademark holder, even when that action falls short of violating antitrust laws. Critics of the misuse doctrine typically focus their ire on this approach. Antitrust law, after all, encompasses the law's collective wisdom concerning the analysis of anticompetitive conduct. Misuse ignores this wisdom if it permits courts to ignore antitrust law's essential requirement—the presence of market power.<sup>98</sup>

When compared to patents and, to some extent, copyrights, trademarks only rarely generate market power. As the court articulated in *Clorox Co. v. Sterling Winthrop, Inc.*,<sup>99</sup>

A trademark, unlike other intellectual property rights, does not confer a legal monopoly on any good or idea; it confers rights to a name only. Because a trademark “merely enables the owner to bar others from use of the mark, as distinguished from competitive manufacture and sale of identical goods bearing another mark,” “the opportunity for effective antitrust misuse of a trademark . . . is so limited that it poses a far less serious threat to the economic health of the nation [than does patent misuse].”<sup>100</sup>

Thus, in accordance with the general trend in misuse jurisprudence, this Article opposes the adoption of an antitrust-based doctrine of trademark misuse.

The alternative, equitable approach avoids redundancy with antitrust law, but in doing so tends to expand to the point of incoherence. In its

---

98. See HERBERT HOVENKAMP, MARK D. JANIS & MARK LEMLEY, *IP & ANTITRUST: AN ANALYSIS OF ANTITRUST PRINCIPLES APPLIED TO INTELLECTUAL PROPERTY LAW* 3.4b2(B) (2004) (asserting that it makes sense to test copyright misuse using antitrust principles because “[a]ntitrust has a long history of evaluating such conduct; copyright misuse does not.”). Granted, antitrust law imprecisely characterizes the presence of market power as an either-or binary inquiry, rather than a matter of degree. Cf. JAMES E. HARTLEY, *THE RULE OF REASON* 139 (1999) (discussing the view of the per se and rule of reason analysis as parts of a continuum rather than two distinct inquiries). Thus, misuse could treat conduct by firms with semi-market power and simply use less drastic remedies than antitrust. While theoretically sound, this approach still ignores what antitrust has learned: an oversimplified inquiry already runs the risk of penalizing and deterring behavior that benefits social welfare—a sliding-scale approach only stands to increase uncertainty. Also, without guidance from antitrust law, courts may be prone to resort to their personal instinct under the guise of competition policy.

99. 117 F.3d 50 (2d Cir. 1997).

100. *Id.* at 56 (quoting *Carl Zeiss Stiftung v. V.E.B., Jena*, 298 F. Supp. 1309, 1314 (S.D.N.Y. 1969)). But see Glynn S. Lunney, Jr., *Trademark Monopolies*, 48 EMORY L. J. 367 (1999).

most general terms, misuse under a public policy approach discerns whether the trademark is being used in a manner violative of the public policy embodied in the grant of the right.<sup>101</sup> In other areas of intellectual property law, this standard essentially invites judges to implement their own, often idiosyncratic, appraisal of the intellectual property system. As a result, discussions of equity-based misuse often constitute a jumbled mix of platitudes and reprimands.<sup>102</sup> This encourages critics to attack the doctrine for lacking coherence and predictability, which, they charge, ultimately undermines the value of intellectual property rights.<sup>103</sup> Within this historical backdrop, trademark law presents the opportunity for misuse to avoid its past pitfalls.

## B. The Proposed Approach

An equity-based approach to the doctrine of trademark misuse should provide clear guidance to courts. With that goal in mind, this Section first sets out the proposed test for a doctrine of trademark misuse; it then justifies the test and applies it to various factual situations.

Broadly speaking, the elements of trademark misuse should be a trademark holder's:

- 1) incorrect assertion of rights
- 2) with an improper purpose.

The first element calls on a court to assess the legal basis of a trademark holder's assertion of rights. The second element requires inquiry into the subjective intent of the trademark holder. If trademark holders do not seek the proper adjudication of the claim upon which the assertion of rights is based, their purpose is improper under the second element. Yet ascertaining the purpose of a trademark holder poses obvious evidentiary

---

101. Note, this does not mean that such violations cannot have an anticompetitive effect as well.

102. See, e.g., *Practice Mgmt. Info. Corp. v. Am. Med. Ass'n.*, 121 F.3d 516, 521 (9th Cir. 1997) (concluding that "[t]he adverse effects of the licensing agreement are apparent" because "[t]he terms under . . . gave [the defendant] a substantial and unfair advantage over its competitors."). Brett Frischmann and Dan Moylan's article constitutes an exception. Frischmann & Moylan, *supra* note 36 (identifying three functions of misuse: a corrective function, a coordination function, and a safeguarding function); see also Dan L. Burk, *Anticircumvention Misuse*, 50 UCLA L. REV. 1095, 1123 (2003) (remarking that an equitable doctrine should be applied to "legal claims that might be technically legitimate, but which would lead to socially perverse outcomes.").

103. See e.g., F. Scott Kieff & Troy A. Paredes, *The Basics Matter: At the Periphery of Intellectual Property*, 73 GEO. WASH. L. REV. 174 (2004).

problems; thus certain factual situations in which a trademark holder's purpose is unlikely to be related to consumer confusion should raise a presumption of an improper purpose. This Article contends that one situation meriting such a presumption is when a trademark holder attempts to enforce rights against a defendant's criticism or parody.<sup>104</sup>

### 1. *Incorrect Assertion of Rights*

One could object at the outset that victims of a trademark holder's incorrect assertion of rights have a ready solution: take the culprit to court. But misbalanced incentives render litigation an inadequate response, at least under the current legal regime. Under normal circumstances, a trademark holder will proceed in a suit so long as the potential damages,  $D$ , multiplied by the probability of infringement,  $P_i$ , exceed the enforcement costs,  $C$ .

Bring suit for infringement if:  $D * P_i > C$

Many potential infringers must research the law or hire an attorney in order to assess liability, thereby expending information costs,  $I$ . A rational trademark holder would account for these costs when deciding whether to bring suit.

Enforcement decision, accounting for information cost differential:

$$D * P_i > C + I$$

Because trademark holders possess, on average, better information about the scope and validity of their rights, they can sometimes successfully bluff a defendant, despite having a weak lawsuit.<sup>105</sup> More fundamen-

---

104. See, e.g., *In re Napster*, 191 F. Supp. 2d 1087, 1105 (N.D. Cal. 2002). The court stated:

Additional confusion arises because while courts have repeatedly stated that misuse is different from antitrust, they still rely on antitrust-like inquiries in determining what licensing agreements violate public policy. Of the cases reviewed by the court, all mimic the *per se* rules of antitrust in holding that the relevant licensing agreements constitute copy-right misuse because they are unduly restrictive *on their face*.

*Id.*

105. This information costs approach derives from the broader insight that the law should place burdens on the least cost avoider. See GUIDO CALABRESI, *THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* 26, 135-73 (1970) (arguing that efficiency in accident cost reduction is maximized by imposing liability on the party best able to choose between accident and safety costs); see also William Hubbard, Note,

tally, defendants must pay litigation costs without any chance of payoff.<sup>106</sup> Finally, trademark holders may have an interest in establishing a reputation in prosecuting even weak suits through to the end, so even losing a weak lawsuit may be a profitable investment in that reputation.<sup>107</sup>

These factors bias trademark holders' enforcement decisions when infringement is possible, but prove more influential on such decisions when the probability of infringement is negligible. The law does not normally concern itself with situations in which the probability of infringement is zero. Why would a trademark holder care if, by definition, there exists no likelihood of confusion? The reason is straightforward: the absence of confusion does not equate to an absence of costs to a trademark holder. This is one reason parody and critique, which impose costs without (usually) creating confusion, are the focus of this Article.

Accordingly, from the trademark holder's perspective, attempting to enforce via a cease-and-desist letter makes sense so long as the damage to the mark wrought by the defendant's critique,  $D_{\Delta}$ , reduced by the defendant's probability of complying with the letter,  $P_c$ , exceeds the cost of sending the letter,  $C$ .

Enforcement against critique or parody is rational if:  $D_{\Delta} * P_c > C$

Assessing the probability that a defendant will comply with a cease-and-desist letter requires us to examine the situation from the defendant's perspective. For non-infringing use, defendants will nevertheless comply so long as the information costs, i.e., the costs of assessing liability,  $I$ , exceed the benefits they gain from using the mark,  $G$ .

---

*Communicating Entitlements: Property and the Internet*, 22 YALE L. & POL'Y REV. 401 (2004) (connecting this insight to notice in exchange between property holder and potential trespasser/infringer); Lior Strahilevitz, *Information Asymmetries and the Rights to Exclude*, 104 MICH L. REV. 1835 (2006). Henry Hansmann and Reinier Kraakman discuss the efficiency of altering verification rules to reduce aggregate user and non-user costs. See Henry Hansmann & Reinier Kraakman, *Property, Contract, and Verification: The Numerus Clausus Problem and the Divisibility of Rights*, 31 J. LEGAL STUD. 373 (2002). Applying this logic to trademark law suggests placing the burden on trademark holders to accurately represent their rights.

106. Note that Section 35(a) of the Lanham Act provides that a "court in exceptional cases may award reasonable attorney fees to the prevailing party." 15 U.S.C. §1117(a).

107. Patrick Bolton et al., *Predatory Pricing: Strategic Theory and Legal Policy*, 88 GEO. L.J. 2239, 2999-2301 (2000) (noting the reputation effect of predation); see also David M. Kreps & Robert Wilson, *Reputation and Imperfect Information*, 27 J. ECON. THEORY 253 (1982).

Defendant complies if:  $I > G$

This basic assessment of incentives reveals two characteristics of a trademark holder's enforcement against a defendant's non-infringing use. First, unlike in a full-scale lawsuit, the marginal cost of enforcement—in essence, one Google search and a cease-and-desist letter—is nearly negligible. Second, for creators of websites that parody or critique a trademark, the benefit from using the mark tends to be more psychic than financial; hence, these creators are understandably unwilling to hire a lawyer to determine the merits of their case. Consequently, a defendant will often comply, and a trademark holder needs only limited success for enforcement to be worthwhile.

In the absence of a trademark misuse doctrine, trademark holders have every incentive to abuse this lawful tool: they can send cease-and-desist letters to anyone who criticizes their product, and exaggerate or even lie about their rights or the potential repercussions.<sup>108</sup> Imposing misuse damages,  $D_M$ , to the equation above counteracts this bias by imposing a cost, in addition to the cost of enforcement, that would balance the costs and benefits of enforcement when there is no infringement.

Imposing misuse damages:  $D_\Delta * P_C > C + D_M$

For trademark misuse to cancel out the perverse incentives arising from non-infringing conduct, the damages should equal the amount the trademark holder stands to gain from enforcing, less the cost of enforcement, i.e.,  $D_\Delta * P_C - C$ .

In practice, this amount would be inadequate, however. For the same reason defendants often comply with baseless cease-and-desist letters—namely, expensive lawyers—they will also fail to enforce the misuse remedy. To neutralize a trademark holder's incentives, then, the damage must be multiplied by the probability that the defendant enforces her misuse remedy,  $P_E$ .

Imposing misuse damages, revised:  $D_\Delta * P_C > C + (D_M * P_E)$

Thus, misuse damages neutralize a trademark holder's incentives if:

---

108. Note, the Lanham Act already provides an exception to incontestability that includes intentionally false and misleading designations of origin. 15 U.S.C. §1115(b).

$$D_M = \frac{D_\Delta P_C - C}{P_E}$$

This focus on the trademark holder is open to critique because both parties share responsibility for information asymmetry regarding the scope of trademark rights. To some extent, then, one could argue that the alleged infringer acquiesced in the trademark holder's assertion of extra-legal rights because otherwise acquiring sufficient information simply costs too much. Punishing the trademark holder for an alleged infringer's rational ignorance may seem unfair, as well as potentially inefficient.<sup>109</sup> But rational ignorance may not seem so troublesome if one accepts that its 'rationality' derives solely from flawed law, which creates unclear property lines in conjunction with asymmetric consequences: drastic damages for infringers and little-to-none for trademark owners. As mentioned in the Introduction, revitalizing misuse would bring symmetry to trademark law by regulating potential defendants *and* trademark holders.

Directing courts to determine whether the trademark holder incorrectly asserted her rights also narrows the inquiry much more than copyright law's mandate that copyright holders not attempt to extend the scope of their right.<sup>110</sup> Applying copyright law's standard would, if construed broadly, appear to prohibit most standard licensing arrangements. Misuse should avoid undermining freedom of contract in this way, especially for contractual relations between a trademark holder and its licensees.<sup>111</sup>

Ironically, copyright law's focus on 'scope expansion' also induces courts to focus too narrowly on explicit licensing terms rather than implicit tactics. To courts, the terms of a licensing contract provide the best evidence for the improper acquisition of rights. For example, in *Lasercomb*, the licensor restricted the licensee's ability to develop similar software.<sup>112</sup> But in trademark law, most problematic conduct occurs when the

---

109. The concept of rational ignorance is that any person will spend only a certain amount of time or money to obtain a piece of information. If obtaining that information costs more than the information is worth, they will (or should) rationally choose to remain ignorant of it. See RALPH T. BYRNS & GERALD W. STONE, *ECONOMICS* 433 (4th ed. 1989) (discussing rational ignorance).

110. See *Practice Mgmt. Info. Corp. v. Am. Med. Ass'n.*, 121 F.3d 516, 521 (9th Cir. 1997); *A&M Records, Inc. v. Napster*, 239 F.3d 1004, 1026 (9th Cir. 2001) ("The misuse defense prevents copyright holders from leveraging their limited monopoly to allow them control of areas outside the monopoly.").

111. See Robert P. Merges, *Intellectual Property and the Costs of Commercial Exchange: A Review Essay*, 93 MICH. L. REV. 1570, 1573-74 (1995) (describing how various intellectual property rights facilitate distinct organizational structures).

112. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 973 (4th Cir. 1990).

“expansion” is only implicit.<sup>113</sup> As such, instructing courts to look for expansion of the trademark right is bound to fail: such evidence is rare. Screening for the incorrect assertion of rights prevents over-claiming, without the distraction of culling evidence of expansion. Courts, moreover, are eminently capable of assessing the legal basis of an infringement allegation.

## 2. *Improper Purpose*

As Judge Posner asked in *USM Corp. v. SPS Technologies, Inc.*,<sup>114</sup> “[i]f misuse cases are not tested by conventional antitrust principles, by what principles shall they be tested?”<sup>115</sup> Judge Posner’s skepticism is appropriate: if we cannot identify a policy reason for intervention other than the efficiency reasons already accounted for in antitrust law, misuse wastes our time at best and otherwise deters welfare-enhancing interactions. At a minimum, an equitable misuse doctrine requires us to define what kind of equity we seek. This Article selects free speech as an objective that a trademark misuse doctrine should advance.

Potential policy interests related to a trademark misuse doctrine fall into two general categories: those specific to trademark law and those independent of it. The former category is less controversial; indeed, some courts already partially recognize that certain trademark policies are important enough for misuse’s protection. For example, even the *Northwestern* court, which characterized the defense as a “phantom,” begrudgingly acknowledged that misuse may apply in situations in which the trademark owner “somehow [does] violence to the public policy which establishes [the] trademark right.”<sup>116</sup> To the *Northwestern* court, this only occurs when trademark holders misrepresent their product.<sup>117</sup> But as a matter of logic, it is hard to see why misuse should stop at misrepresentation when other behaviors probably contravene trademark public policy as well.<sup>118</sup>

---

113. From an ex ante perspective, infringement of an intellectual property right is inherently probabilistic, so conduct that simply raises the probabilities functionally “expands” scope as well. See, e.g., Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSPECTIVES 75 (2005) (making the point with respect to patent rights).

114. 694 F.2d 505 (7th Cir. 1982).

115. *Id.* at 512.

116. *Nw. Corp. v. Gabriel Mfg., Co.*, 48 U.S.P.Q.2d 1902, 1908-09 (N.D. Ill. 1998).

117. *Id.* at 1909.

118. For example, a second option is trademark law’s underlying policy: protecting against consumer confusion. Cf. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 978-79 (4th Cir. 1990) (concluding that a broad covenant not to develop other software conflicts with the copyright goal of stimulating original expression).

Employing misuse to advance policy objectives that are at best tangential to trademark law invites more controversy, largely due to concern that unmooring misuse from its intellectual property origins will lead to the overexpansion mentioned above, undermining the incentive structure embodied in the grant of the property right. But this Article contends that a trademark misuse doctrine could advance free speech without undermining trademark law's underlying policies.

Using misuse to advance free speech<sup>119</sup> may seem redundant, at least at first glance, because the First Amendment already provides a free speech defense in trademark infringement and dilution cases.<sup>120</sup> The Lanham Act itself also appears to integrate such First Amendment values. By requiring a "likelihood of confusion," for example, traditional infringement law obliquely follows Supreme Court deference to government restrictions on "forms of communication more likely to deceive the public

---

119. Some commentators similarly recommend using misuse to advance free speech within copyright. See JuNelle Harris, *Beyond Fair Use: Expanding Copyright Misuse to Protect Digital Free Speech*, 13 TEX. INTELL. PROP. L.J. 83 (2004); Judge, *supra* note 80. In general, the overlap of the First Amendment and copyright has been examined extensively. See, e.g., Neil Weinstock Netanel, *Locating Copyright Within the First Amendment Skein*, 54 STAN. L. REV. 1 (2001); Jed Rubinfeld, *The Freedom of Imagination: Copyright's Constitutionality*, 112 YALE L.J. (2002); Alfred C. Yen, *Eldred, the First Amendment, and Aggressive Copyright Claims*, 40 HOUS. L. REV. 673 (2003). An interesting study relevant to free speech and trademark found that American novels use an average of 160 generic names per 10,000 words. Monroe Friedman, *The Changing Language of a Consumer Society: Brand Name Usage in Popular American Novels in the Postwar Era*, 11 J CONSUMER RES. 927, 932-33 (1985).

120. In 1980, the Supreme Court set forth a four-part analysis for commercial speech cases:

[First], we must determine whether the expression is protected by the First Amendment. For commercial speech to come within that provision, it at least must concern lawful activity and not be misleading. [Second], we ask whether the asserted governmental interest is substantial. [Third, if] both inquiries yield positive answers, we must determine whether the regulation directly advances the governmental interest. [And fourth], if the governmental interest could be served as well by a more limited restriction on commercial speech, the excessive restrictions cannot survive.

Cent. Hudson Gas & Elec. Corp. v. Public Serv. Comm'n of N.Y., 447 U.S. 557, 566 (1980); see also Richard B. Biagi, *The Intersection of First Amendment Commercial Speech Analysis and the Federal Trademark Dilution Act: A Jurisprudential Roadmap*, 91 TRADEMARK REP. 867 (2001). If dilution law restricted non-commercial speech, it wouldn't matter that it only restricts certain words, but leaves speakers free to express facts and ideas. As *Cohen v. California* held, "forbid[ding] particular words" also "run[s] a substantial risk of suppressing ideas in the process." 403 U.S. 15, 26 (1971).

than to inform it.”<sup>121</sup> Likewise, the Federal Trademark Dilution Act ostensibly integrates First Amendment policy by exempting “non-commercial use” of a mark from actions for dilution.<sup>122</sup>

In practice, however, both of these provisions fall short of standard First Amendment protection.<sup>123</sup> Most notably, in the dilution context, only Judge Kozinski in *Mattel, Inc. v. MCA Records, Inc.* identified what is obvious in other parts of First Amendment jurisprudence: speech is not made commercial by the fact that someone sells it; it must actually propose a commercial transaction.<sup>124</sup> The majority of courts, on the other hand, categorize trademark use as unprotected “commercial” speech on the basis of any commercial intent or impact on a trademark holder’s own commercial interests.<sup>125</sup> For example, in *Planned Parenthood Federation of America, Inc. v. Bucci*,<sup>126</sup> the court enjoined Bucci, a pro-life activist, from espousing his view on his website *plannedparenthood.com*. As a basis for categorizing Bucci’s speech as commercial, the court mentioned that the site harmed Planned Parenthood commercially and that Bucci solicited funds on his radio show while mentioning the website.<sup>127</sup> But extending this standard outside of trademark law would lead to absurd results: “commer-

---

121. *Cent. Hudson*, 447 U.S. at 563.

122. 15 U.S.C. § 1125(c)(4)(B). Congress included the non-commercial use exemption in the FTDA in light of First Amendment concerns that the statute would otherwise prohibit protected forms of expression. *L.L. Bean, Inc., v. Drake Publishers, Inc.*, 811 F.2d 26 (1987) (reversing lower court’s holding that non-commercial use of a trademark in a parody is subject to Maine’s anti-dilution statute). Other doctrines indirectly promote free speech as well. For example, the trademark fair use doctrine for comparative advertising first asks whether the defendant can readily identify the plaintiff’s product without using a trademark, and then whether the defendant used more of the plaintiff’s mark than necessary. *New Kids on the Block v. New Am. Publ’g*, 971 F.2d 302, 308 (9th Cir. 1992).

123. For example, the likelihood of confusion requirement fails to integrate Supreme Court instruction that while inherently misleading advertising may be prohibited outright, the government “may not place an absolute prohibition on certain types of potentially misleading information . . . if the information also may be presented in a way that is not deceptive.” *Peel v. Attorney Registration & Disciplinary Comm’n*, 496 U.S. 91, 100 (1990).

124. 296 F.3d 894, 906 (9th Cir. 2002); see also EUGENE VOLOKH, *THE FIRST AMENDMENT: PROBLEMS, CASES AND POLICY ARGUMENTS* 235 (2001); Alex Kozinski & Stuart Banner, *The Anti-History and Pre-History of Commercial Speech*, 71 TEX. L. REV. 747 (1993); Alex Kozinski & Stuart Banner, *Who’s Afraid of Commercial Speech?*, 76 VA. L. REV. 627 (1990).

125. See generally Patrick D. Curran, Comment, *Diluting the Commercial Speech Doctrine: “Noncommercial Use” and the Federal Trademark Dilution Act*, 71 U. CHI. L. REV. 1077 (2004).

126. 42 U.S.P.Q.2d 1430 (S.D.N.Y. 1997).

127. *Id.* at 1435-36.

cial” speech would include all speech that is sold, including books, newspapers, and magazines, and the government could ban any speakers with economic intent, such as workers who rally for higher pay.<sup>128</sup>

The fact that courts underutilize First Amendment defenses suggests adopting Judge Kozinski’s more adequate conception, but does not necessarily support incorporating free speech values into a misuse doctrine. Indeed, at first glance, Judge Kozinski’s robust construal obviates the need for misuse to preserve free speech—a conclusion arrived at in most recent trademark law scholarship sympathetic to the freedom of speech.<sup>129</sup>

While this recommendation advances free speech in theory, it falls markedly short of upholding the level of free speech its proponents desire in practice. Driving the wedge between theory and practice are the tactics deployed by trademark holders described in this Article.<sup>130</sup> That is, certain kinds of coercive conduct undermine free speech independent of formal First Amendment defenses—the paradigmatic example being pre-litigation threats sent to alleged infringers or internet intermediaries based on dubious claims.<sup>131</sup> Misuse restores this constricted free speech space by providing a more aggressive weapon than a First Amendment shield.

Before proceeding, it is worth mentioning that this step from shield to sword does not derive from the text of the First Amendment itself, but from so-called free speech *values*. This latter viewpoint is not as outlandish as one would suspect; indeed, an entire school of thought, and certain swaths of First Amendment jurisprudence,<sup>132</sup> conceptualize the First Amendment more broadly than its traditional negative role, obliging the government to protect speech actively.<sup>133</sup> Under such a conception, a

---

128. One could argue that trademark law deserves explicit exemption from First Amendment analysis, such as the Court has done for copyright. However, courts instead disingenuously pretend that trademark law actually comports with commercial speech law, and then apply their watered-down version of First Amendment protection.

129. See, e.g., K.J. Green, *Abusive Trademark Litigation and the Incredible Shrinking Doctrine—Trademark Abuse in the Context of Entertainment Media and Cyberspace*, 27 HARV. J.L. PUB. POL’Y 609 (2004); Hannibal Travis, *The Battle for Mindshare: The Emerging Consensus that the First Amendment Protects Corporate Criticism and Parody on the Internet*, 10 VA. J.L. & TECH. 3 (2005) (contending that trademark rights should be restricted to policing commercial competition, rather than non-commercial internet speech); see also Lemley, *supra* note 90; Sara Stadler Nelson, *The Wages of Ubiquity in Trademark Law*, 88 IOWA L. REV. 731 (2003).

130. This Article also argues that misuse as a free speech weapon should apply, albeit more deferentially, to non-coercive licensing transactions.

131. See *infra* Section III.C.

132. An example of such active protection is the public forum doctrine. See VOLOKH, THE FIRST AMENDMENT, *supra* note 124, at 419.

133. See *infra* note 134.

merely libertarian First Amendment that grants property owners free reign within their property right to determine what speech to permit and what speech to restrict fails to realize certain collective values that are necessary for a well-functioning system of free speech.<sup>134</sup>

Indeed, grafting positive speech protection onto trademark law, rather than simply using the First Amendment itself, evades the most powerful critique of this notion of free speech—that the First Amendment’s broad scope renders positive rights too complex to apply in light of the context-specific judgments they require.<sup>135</sup> Trademark misuse offers a narrow, though vital, section of law to instill such rights. In other words, while we lack the capacity to subsidize dissident and other disfavored or unpopular speech wholesale, misuse can provide affirmative protection in trademark-related disputes.

Scholars who frame this positive protection as a “subsidy” nonetheless misconstrue its effect; the benefits to the “subsidized” individual are extracted from another party, not society in general.<sup>136</sup> For example, what looks like “speech protection” to the petitioning protester is “compelled speech” to the land owner whose private shopping mall cannot exclude him. Likewise, in preventing trademark owners from privately restricting speech, misuse penalizes trademark holders for the statements they make in their cease-and-desist letters, thereby promoting one form of speech, critique and parody, by restricting another. But the fairness of this transfer

---

134. CASS SUNSTEIN, *DEMOCRACY AND THE PROBLEM OF FREE SPEECH* 18 (1993). See also LAURENCE H. TRIBE, *AMERICAN CONSTITUTIONAL LAW* 964 (2d ed. 1988). (“[I]t is not enough for the government to refrain from invading certain areas of liberty. The State may, even at some cost to the public, be required to provide at least a minimally adequate opportunity for the exercise of certain freedoms.”). Richard Posner likewise acknowledges such a role for the government and free speech. See Richard A. Posner, *Free Speech in an Economic Perspective*, 20 *SUFFOLK U.L. REV.* 1, 52 (1986).

Until fairly recently it was assumed that the purpose of the first amendment was the negative one of preventing undue interference with private markets in ideas rather than the positive one of promoting the effective functioning of such markets. The “public forum” doctrine, however, requires the government in some cases to make public facilities available for persons wanting to express themselves.

*Id.* But see Richard A. Epstein, *Cybertrespass*, 70 *U. CHI. L. REV.* 73 (2003). See generally Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 *N.Y.U. L. REV.* 1, 7 (2004); Dawn C. Nunziato, *The Death of the Public Forum in Cyberspace*, 20 *BERKELEY TECH. L.J.* 1115 (2005).

135. See, e.g., David G. Post, *What Larry Doesn’t Get: Code, Law, and Liberty in Cyberspace*, 52 *STAN. L. REV.* 1439, 1457 (2002).

136. See Wesley Newcomb Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 *YALE L.J.* 16 (1913).

is contingent on an already-synthetic legal baseline. As argued above, the law already intervenes to grant trademark holders special speech rights—misuse intends to counteract where these rights go too far.

### C. Applying the Test

The threshold question under this Article's proposed test is whether a trademark holder has incorrectly asserted her rights. Next, the doctrine of misuse inquires whether the trademark holder asserted rights with an improper purpose, i.e., a purpose collateral to the object of the process upon which the assertion of rights is based. Due to evidentiary concerns, this Article recommends presuming an improper purpose in certain situations, namely, when the defendant critiques or parodies the trademark holder. As such, this next Section demonstrates how one would apply the test in two situations: when an improper purpose is presumed and when it is not.

#### 1. *Presumption of an Improper Purpose*

The cease-and-desist letter is the tool that most often abets the troublesome conduct that falls within this category. But, as is the problem with many tools, it serves both lawful and (potentially) unlawful means.

When enough money is on the line for the alleged infringer, the relevance of trademark law's misbalanced incentives recedes. Lawsuits over commercial products, even questionable ones such as Mattel's suit over the "Barbie Girl" song<sup>137</sup> or Fox's suit against Al Franken over the title of his book, *Lies and the Lying Liars Who Tell Them: A Fair and Balanced Look at the Right*,<sup>138</sup> surely waste resources, but usually fail to coerce the alleged infringer. While these lawsuits naturally draw the most attention,<sup>139</sup> scholarly and otherwise, they are not the focus of this Article.

Instead, this Article is concerned with enforcement of questionable claims when the financial stakes are not so high. For example, when former employees of Radio Shack created the website Radioslack.com to serve as a forum to discuss recent changes in pay structure (from commission to hourly wage), one harsh cease-and-desist letter<sup>140</sup> alleging trade-

---

137. *Mattel, Inc. v. MCA Records, Inc.*, 296 F.3d 894 (9th Cir. 2002).

138. Complaint at 2, *Fox News Network, LLC v. Penguin Group, Inc.*, Index No. 602514/2003 (N.Y. Sup. Ct. 2003), available at <http://fl1.findlaw.com/news.findlaw.com/hdocs/docs/ip/foxpenguin80703cmp.pdf>.

139. See, e.g., DAVID BOLLIER, *BRAND NAME BULLIES* (2005).

140. See Chilling Effects Clearinghouse, *Radioshack not amused by radioslack.com*, <http://www.chillingeffects.org/protest/notice.cgi?NoticeID=324> (last visited Dec. 17, 2006).

mark infringement—even claiming they have no right to use any “portion[s] of the mark”<sup>141</sup>—was all it took for the site to disappear.

A recent study by the Brennan Center for Justice of chillingeffects.org—a website that collects and publishes a database of cease-and-desist notices sent to internet users—confirms that the bias outlined above is more than theoretical.<sup>142</sup> Of the letters that were possible to evaluate, more than a fifth stated extremely weak claims or had strong First Amendment defenses, and many had reasonable free speech defenses.<sup>143</sup> And yet within the first category of weak claims, almost half of the letters successfully caused removal.<sup>144</sup> These results are particularly troubling in view of the sample’s obvious selection bias: individuals or entities aware of the chillingeffects.org website likely know their legal rights better than the average internet user.

State legislatures across the country increasingly recognize the consequences of these tactics and the disturbing implications for free speech rights. To discourage lawsuits brought primarily to chill the valid exercise of free speech rights, several legislatures have responded by enacting so-called anti-SLAPP (“Strategic Lawsuits Against Public Participation”) statutes that grant individuals procedural mechanisms to dismiss these non-meritorious lawsuits at an early stage.<sup>145</sup> Under California’s anti-SLAPP<sup>146</sup> statute, for example, lawsuits that arise from any act in furtherance of a person’s right to petition or to free speech<sup>147</sup> are subject to a spe-

---

141. *Id.*

142. BRENNAN CENTER FOR JUSTICE, WILL FAIR USE SURVIVE? FREE EXPRESSION IN THE AGE OF COPYRIGHT CONTROL (2005) [hereinafter WILL FAIR USE SURVIVE?]. Despite the title of the report, a sizable portion of the cease-and-desist letters stated claims of trademark infringement, particularly among the weak claims.

143. *Id.* at 36.

144. *Id.*

145. *See, e.g.*, FLA. STAT. § 768.295 (2000); IND. CODE 34-7-7 (2006); N.M. STAT. ANN. §§ 38-2-9.1 to 9.2 (West 2006); OR. REV. STAT. §§ 31.150–31.155 (2005); 27 PA. CONS. STAT. ANN. §§ 7707, 8301–8305 (West 2001); UTAH CODE ANN. §§ 78-58-101 to -105 (West 2006).

146. CAL. CIV. PROC. CODE § 425.16(b)(1) (West 2005); *see also* *Metabolife Int’l, Inc. v. Wornick*, 213 F. Supp. 2d 1220, 1221 (S.D. Cal. 2002) (“The California Legislature passed the statute recognizing ‘the public interest to encourage continued participation in matters of public significance . . . and [finding] that this participation should not be chilled through abuse of the judicial process.’”).

147. *See* CAL. CIV. PROC. CODE § 425.16(e) (defining “act in furtherance of a person’s right of . . . free speech.”).

cial motion to strike unless the plaintiff can demonstrate a probability of prevailing.<sup>148</sup>

Trademark holders need not focus on the alleged infringer exclusively to quash criticism; with increasing frequency, they achieve similar results by sending cease-and-desist letters to internet intermediaries.<sup>149</sup> By informing the intermediary that a certain site infringes their trademark, often under the dubious basis that the site merely uses the company name, trademark holders are likewise empowered to suppress criticism, and thus ignore otherwise narrow free speech exceptions.<sup>150</sup>

Under the original limitations of contributory liability for trademark infringement, such letters posed little in the way of legal consequence unless the intermediary somehow “intentionally induce[d]” infringement or “continue[d] to supply its product to one whom it knows or had reason to know is engaging in trademark infringement.”<sup>151</sup> But such letters require attention and even compliance now that several courts have expanded contributory infringement’s striking range.<sup>152</sup>

Intermediaries face strong incentives to comply with complaints of infringement in their search results or hosted content. These incentives do not stem from information asymmetry, as before, but result from a principal-agent problem. Liability for intermediaries, who act as agents for al-

---

148. *See* *Batzel v. Smith*, 333 F.3d 1018, 1024 (9th Cir. 2003) (“The plaintiff must demonstrate that the complaint is legally sufficient and supported by a prima facie showing of facts to sustain a favorable judgment if the evidence submitted by plaintiff is credited.”). As with a motion for summary judgment, the evidence the plaintiff submits in support of its claims must be admissible evidence. If the court denies an anti-SLAPP motion to strike, the parties continue with discovery and the anti-SLAPP statute no longer applies. *Id.*

149. *See, e.g.,* Chilling Effects Clearinghouse, *Keyword: Google and the DMCA*, <http://www.chillingeffects.org/dmca512/keyword.cgi-?KeywordID=2> (last visited Jan. 6, 2007).

150. Section 230 of the Communications Decency Act exempts service providers from liability for defamation committed by their users, rendering all such actions not Google’s problem. 47 U.S.C. § 230.

151. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 854 (1982). The Court outlined the basis for contributory trademark infringement:

If a manufacturer or distributor intentionally induces another to infringe a trademark, or if it continues to supply its product to one whom it knows or has reason to know is engaging in trademark infringement, the manufacturer or distributor is contributorily responsible for any harm done as a result of the deceit.

*Id.*

152. For a critique of this expansion see Benjamin Aitkin, *Keyword-Linked Advertising, Trademark Infringement, and Google’s Contributory Liability*, 2005 DUKE L. & TECH. REV. 21; Dogan & Lemley, *supra* note 86.

legedly infringing principals, exposes them to all of the downside of infringement although they appreciate little of its upside.<sup>153</sup> Accordingly, intermediaries tend to comply so long as litigation costs and potential damages exceed the indirect benefit of infringement.<sup>154</sup> And such compliance—i.e., removal of the hosted content or delisting from search results—often functionally eliminates the site from the internet itself, especially for powerful intermediaries such as Google. In other words, cease-and-desist letters based on weak claims stand to succeed with effects similar to those sent to defendants themselves.

One infamous example of the private censorship enabled by these misaligned incentives is when Dow Chemical was able, merely by alleging trademark violations to an internet service provider (ISP), to shut down a website critical of Dow's alleged responsibility for a gas leak at a Union Carbide plant in Bhopal, India, that killed thousands of people.<sup>155</sup> Because the ISP in the case had no private stake in communicating the message, it deferred to the trademark holder's judgment rather than face any litigation costs despite the weakness of Dow Chemical's claim.

A final type of conduct that may fall within this category is the practice of so-called "reverse domain-name hijacking,"<sup>156</sup> in which trademark owners use the Uniform Domain Name Dispute Resolution Policy to strip legitimate holders of their domain names by asserting expansive trade-

---

153. 17 U.S.C. § 1201(a) (1998). For an excellent article outlining the misaligned incentives of ISPs, see Assaf Hamdani, *Who's Liable for Cyberwrongs?*, 87 CORNELL L. REV. 901, 905-06 (2002) (discussing why intermediaries that do not reap the benefits of the wrong should not be held strictly liable, otherwise they "would assign disproportionate weight to the risk of legal liability and disregard the loss caused to [the direct users] as a result of the newly blocked or removed user information.").

154. Note that intermediaries may also fight infringement in an effort to change the law. This is arguably Google's strategy in *Google Inc. v. American Blind & Wallpaper Factory Inc.*, No. C03-5340-JF(RS), 2006 WL 2318803, (N.D. Cal. Aug. 10, 2006), and *Government Employees Insurance Co. v. Google, Inc.*, 77 U.S.P.Q.2d 1841 (E.D. Va. 2005).

155. See Chilling Effects Clearinghouse, *Dow slap Yes Men and upstream ISP for parody/remix*, <http://www.chillingeffects.org/protest/notice.cgi?NoticeID=500> (last visited Dec. 17, 2006); see also C. Carr, *Dow v. Thing: A Free-Speech Infringement That's Worse Than Censorship*, VILLAGE VOICE, Jan. 17, 2003, at 49; Matthew Mirapaul, *Cyberspace Artists Paint Themselves Into a Corner*, N.Y. TIMES, Dec. 23, 2002, at E2.

156. According to Uniform Domain Name Dispute Resolution Policy Rule 1, the term "Reverse Domain Name Hijacking" means "using [the UDRP] in bad faith to attempt to deprive a registered domain-name holder of a domain name." ICANN, UNIFORM DOMAIN NAME DISPUTE RESOLUTION POLICY (Oct. 24, 1999), available at <http://www.icann.org/dndr/udrp/uniform-rules.htm>.

mark rights.<sup>157</sup> This strategy often focuses on gripe sites that combine the trademarked company name with a pejorative, e.g., “sucks” or “anti.” After a spate of gripe-site litigation, both through the courts and ICANN’s domain dispute process, the law appeared to settle on the rule that such sites were fair use, as most forcefully articulated in the case over the “ballysucks” website.<sup>158</sup> Nevertheless, companies still attempt, and succeed, in acquiring the domain names for these gripe sites. For example, recently, after a disgruntled customer created a non-commercial consumer criticism site, nextelsucks.org, he received a cease-and-desist letter claiming that his site infringes Nextel’s trademark rights and demanded that he immediately transfer ownership of the domain name.<sup>159</sup>

Victims of reverse domain-name hijacking may file an action in court to declare that their registration or use of the domain name is not unlawful under the Anticybersquatting Consumer Protection Act (ACPA).<sup>160</sup> While this provision intends to balance the rights of domain name registrants against the rights of trademark holders,<sup>161</sup> it only provides for injunctive relief. At best, domain holders get their name back,<sup>162</sup> and trademark holders have little reason to curb their abusive practices.<sup>163</sup> Analogous to the ACPA’s approach to cease-and-desist letters, misuse should rebalance the scales in domain name disputes.<sup>164</sup>

---

157. The standard for a finding of reverse domain name hijacking was defined in *Goldline International, Inc. v. Gold Line*: “To prevail on such a claim [of reverse domain name hijacking], Respondent must show that Complainant knew of Respondent’s unassailable right or legitimate interest in the disputed domain name or the clear lack of bad faith registration and use, and nevertheless brought the Complaint in bad faith.” WIPO Domain Name Decision: D2000-1151 (Jan. 2, 2001), available at <http://arbitrator.wipo.int/domains/decisions/html/2000/d2000-1151.html>.

158. *Bally Total Fitness Holding Corp. v. Faber*, 29 F. Supp. 2d 1161, 1167 (C.D. Cal. 1998).

159. The cease-and-desist letter is available at <http://www.nextelsucks.org/legal/letter.htm>.

160. See 15 U.S.C. § 1114(2)(D)(v) (2005).

161. See, e.g., *Barcelona.com. v. Excelentísimo Ayuntamiento de Barcelona*, 330 F.3d 617, 624 (4th Cir. 2003).

162. See 15 U.S.C. § 1114(2)(D)(v).

163. Note, the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) also allows recovery for abuse of authority and abuse of enforcement procedures. Injured parties may recover any monetary damages resulting from such abusive practices, including attorneys’ fees. Agreement on Trade-Related Aspects of Intellectual Property Rights, Including Trade in Counterfeit Goods, § 2, art. 48, para. 1, Dec. 15, 1993, 33 I.L.M. 81, available at [http://www.wto.int/english/docs\\_e/legal\\_e/27-trips.pdf](http://www.wto.int/english/docs_e/legal_e/27-trips.pdf).

164. The lone article advocating for a doctrine of trademark misuse urges its application in precisely this situation. See Stephen J. Davidson & Nicole A. English, *Applying the Trademark Misuse Doctrine to Domain Name Disputes*, *Computer Lawyer*, Aug.

In sum, courts should presume an improper purpose when trademark holders privately censor critique or parody through pre-litigation threats such as cease-and-desist letters against alleged infringers or internet intermediaries or by reverse domain-name hijacking.

## 2. *No Presumption of Improper Purpose*

In contrast, courts should not necessarily presume an improper purpose when trademark holders respond to speech that does not constitute parody or critique. This distinction derives from the model of incentives presented in Section III.B.1, above. Recall that the model revealed that misuse is necessary to prevent a trademark holder from enforcing rights for purposes unrelated to consumer confusion. Censoring critique or parody is a paradigmatic example, but not an exclusive one. Other improper purposes include a trademark holder's threat to file suit to force a person to settle or drop a claim in an unrelated matter, to impose a financial hardship on a competitor, or to accumulate precedent of compliance for other cases. This latter purpose likely accounts for the notorious enforcement practices of New York City's *Village Voice*.<sup>165</sup> Each time a defendant complies with a baseless cease-and-desist letter, *Village Voice* cites it as an example for the next letter. When *The Cape Cod Voice*, a regional journal, received a cease-and-desist letter from the *Village Voice*, asserting both trademark infringement and dilution, the letter noted: "We hope to resolve this matter amicably . . . Our previous experience with the *Bloomington Voice*, *Dayton Voice* and *Tacoma Voice* newspapers . . . indicates that we can accomplish this goal."<sup>166</sup> To *Village Voice*, each cease-and-desist reinforces the

---

1996, available at <http://www.cla.org/trademark%20misuse.pdf>. Moreover, other scholars recommend using damages to deter reverse domain name hijacking. See, e.g., Stacey Knapp, *Balancing the Crucible: The Revolving Conflict Between Fair Use and Corporate Use in the Battle to Control Domain Names*, 1 OKLA J.L. & TECH. 10 (2004).

165. Alexander Zaitchik, *Village Voice Media Inc. is Bullying Tiny Papers all Across the Country*, N.Y. PRESS, Apr. 22, 2003, available at [http://www.nypress.com/print.cfm?content\\_id=8137](http://www.nypress.com/print.cfm?content_id=8137).

166. This category is the most common one on the [chillingeffects.org](http://chillingeffects.org) website. As mentioned previously, companies comply with these letters to a surprising degree. For example, the Muzak company, which creates playlists for stores, successfully forced a website that compiled webpages for independent musicians to change its URL from [muzak.smoe.org](http://muzak.smoe.org) to [www.smoe.org](http://www.smoe.org). On the site, the owner explained: "if you're wondering – yes, there is a reason why we changed our series' name. It has to do with a certain concern which supplies elevator music to the masses, and a cease-and-desist letter we received therefrom. Ain't legalities grand?" See Chilling Effects Clearinghouse, *Don't Play that Funky Muzak*, <http://www.chillingeffects.org/acpa/notice.cgi?NoticeID=1437> (last visited Dec. 17, 2006); [smoe.org](http://www.smoe.org), [http://www.smoe.org/cgi-bin/mj\\_wwwusr?func=info&list=hom-announce](http://www.smoe.org/cgi-bin/mj_wwwusr?func=info&list=hom-announce) (last visited Dec. 17, 2006); see also WILL FAIR USE SURVIVE?, *supra* note 142.

next, a shrewd but inappropriate method of broadening the effective scope of its trademark rights. Evidence of this tactic merits a finding of misuse but should not raise a presumption because the relevant facts are more context specific than in the case of a trademark holder's censoring of critique or parody.

Trademark holders also enforce claims unrelated to consumer confusion by contesting domain names, such as in *Juno Online Services L.P. v. Juno Lighting, Inc.*,<sup>167</sup> previously discussed in Part I. In the dispute, Juno Lighting attempted to cancel Juno Online's domain name, [www.juno.com](http://www.juno.com), compelling Network Solutions, Inc. to place Juno Online's domain name on "hold" and inflicting sizeable damages, despite the fact that Juno Online rightfully held the name. Juno Lighting's purpose in this dispute would have been improper if, for example, it sought to capitalize on Juno Online's successful website.

Without a presumption of improper purpose, trademark holders have an opportunity to offer innocent explanations for their misuse. For example, *Village Voice* or Juno Lighting could avoid misuse by proffering legitimate, trademark-related reasons for their enforcement. One such explanation occurred after the owner of the Godzilla trademark sent a cease-and-desist letter to the blogging site [Davezilla.com](http://Davezilla.com). Although the internet community rushed to judgment against this seemingly weak claim for infringement, the trademark holder apparently sent the letter believing that the blog referred to its logo as Godzilla,<sup>168</sup> which rendered the infringement claim much more tenable. In this circumstance, a reasonable explanation for this belief should excuse what would otherwise be a strong candidate for misuse.<sup>169</sup>

In distinguishing between this category and the previous one that raises a presumption, misuse tends to favor critical and parodic treatments over other uses of protected material. Some scholars worry that this preference within copyright law's fair use doctrine violates the First Amend-

---

167. 979 F. Supp. 684 (N.D. Ill. 1997). Other examples include Gap's dispute over [gap.com](http://gap.com), first registered by "Genesis Access Point"; Hasbro's dispute over [clue.com](http://clue.com), registered to Clue Computer; or Fry's Electronics, Inc.'s dispute over [frys.com](http://frys.com), registered by Frenchy Frys.

168. Chilling Effects Clearinghouse, *Godzilla Tries to Stomp Davezilla.com*, <http://www.chillingeffects.org/trademark/notice.cgi?NoticeID=372> (last visited Nov. 22, 2006).

169. The fact that the owner of the Godzilla mark could assert an explanation, however, does not mean it was reasonable.

ment for not maintaining evaluative neutrality.<sup>170</sup> But the proposed misuse doctrine elides this controversy because it favors parody and critique not by assessing the speech's value; rather, a harsh response to critique tends to indicate an improper purpose because such speech imposes costs on trademark holders unrelated to consumer confusion.

The vast majority of trademark holders' conduct will fall outside of misuse's purview. Most commercial licenses and even most enforcement proceedings directly advance trademark law's central purpose of reducing consumer search costs. Moreover, maintaining a restricted scope acknowledges the fact that new doctrines are not costless; they impose direct administrative costs as well as indirect costs on private actors by creating uncertainty and complexity.<sup>171</sup> Hence, courts should vigilantly refrain from applying the doctrine of trademark misuse to all trademark holder behavior. The narrow scope of the proposed test is the best way to ensure such restraint.

#### IV. WHAT IS THE APPROPRIATE REMEDY?

##### A. Court-Based Equitable Remedies

At first glance, the most glaring barrier to realizing a doctrine of trademark misuse is the lack of statutory support, especially considering the Supreme Court's conclusion in *Park-N-Fly, Inc. v. Dollar Park & Fly, Inc.* that only defenses enumerated by the Lanham Act are permissible for incontestable marks.<sup>172</sup> The prior misuse cases usually rely on Section 1115(b)(7) of the Lanham Act, which provides that use of a mark to violate the antitrust laws is a defense to incontestability. Subsequently, courts have extended this clause to provide a complete defense to an infringement suit.<sup>173</sup>

Instead of the antitrust defense, an equitable misuse doctrine may derive from Section 1115(b)(9), which makes "equitable principles, including laches, estoppel, and acquiescence" an applicable defense to an incon-

---

170. See Jeb Rubinfeld, *The Freedom of Imagination: Copyright's Constitutionality*, 112 YALE L.J. 1, 17 (2002); see also LARRY ALEXANDER, IS THERE A RIGHT OF FREEDOM OF EXPRESSION? 17 (2005).

171. Not to mention the costs of error. See Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 VA. L. REV. 2099 (2004).

172. 469 U.S. 189, 202-03 (1985).

173. See *Phi Delta Theta Fraternity v. J.A. Buchroeder & Co.*, 251 F. Supp. 968, 975-80 (W.D. Mo. 1966) (concluding that this provision may provide a complete defense to a trademark infringement action).

testably registered mark.<sup>174</sup> The Act also requires a balancing of equities when determining whether to enjoin an otherwise infringing act.<sup>175</sup> In practice, courts have narrowly applied these defenses, largely due to concerns about confusing the public, as previously discussed.<sup>176</sup>

Yet, questionable statutory support has not stopped misuse's spread within other areas of intellectual property. In *Lasercomb*, for example, the court reasoned that a misuse of copyright defense is inherent in the law of copyright just as a misuse of patent defense is inherent in patent law.<sup>177</sup> And for trademark law, courts would have a deeper tradition to draw on: the Supreme Court declined to enforce marks on the basis of equity throughout the late nineteenth and early twentieth centuries.<sup>178</sup> As such, lack of statutory support should not by itself preclude a doctrine of trademark misuse.

Traditionally, after a finding of misuse, courts withhold any remedy for infringement until the misuse is purged by abandoning the abusive practice. Several commentators rightfully criticize this remedial scheme because the severity of the penalty bears no relation to the level of harm<sup>179</sup> or the chance of detection.<sup>180</sup> The consequences of withholding a trademark right would instead relate to the utility of the mark. In other words, Coca-Cola stands to suffer more damages than less-known trademarks for legally irrelevant reasons. While these large companies with valuable trademarks may not win over everyone's sympathy, it surely makes no economic sense to punish them for being successful.

Misuse's remedy of unenforceability also directly undermines trademark law's central goal of preventing consumer confusion. If the law permits all generic brands to mimic the Diet Coke label, the diverted sales injure the public as much as Coca-Cola. Clearly, for an unenforceability

---

174. 15 U.S.C. § 1115(b)(9).

175. 15 U.S.C. § 1116(a) (providing that courts should grant injunctions "according to principles of equity").

176. See discussion *supra* Section I.B.2.

177. *Lasercomb Am., Inc. v. Reynolds*, 911 F.2d 970, 973 (4th Cir. 1990); see also *Morton Salt Co. v. G.S. Suppiger*, 314 U.S. 488 (1942) (using principles of equity to recognize the doctrine of misuse in patent law).

178. See *supra* Section I.B.1.

179. See, e.g., Lemley, *supra* note 16; Edward F. Sherry & David J. Teece, *The Misuse Doctrine: An Economic Reassessment*, in ABA SECTION OF ANTITRUST LAW, INTELLECTUAL PROPERTY MISUSES: LICENSING AND LITIGATION 125 (2000).

180. Cf. A. Mitchell Polinski & Steven Shavell, *The Economic Theory of Public Enforcement of Law*, 38 J. ECON. LIT. 45, 47 (2000); A. Mitchell Polinski & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869-962 (1998).

remedy, trademark misuse must demand a nexus between the alleged misuse and the defendant.<sup>181</sup>

But in requiring a nexus, misuse is prone to irrelevance: unenforceability does not punish trademark holders for frivolous claims because, presumably, courts will not enforce these anyway, while legitimate claims avoid misuse altogether. To be sure, an unenforceability remedy would deter claims for which the chance of infringement is low, but nonzero. Then, evidence of conduct that requires high scrutiny transforms the chance of infringement to zero, deterring trademark holders and protecting non-infringing users of the mark.

Unenforceability might deter misuse on the margin, but it ultimately reveals the inadequacy of misuse's traditional remedial measures. As discussed in Part III, an equitable doctrine of misuse should focus on zero probability claims, for which unenforceability poses no real consequence. A second option for these claims is to cost-shift: evidence of misuse would not only render the right unenforceable, but would force the trademark holder to pay for the defendant's attorney fees.<sup>182</sup> As compared to unenforceability, this remedy better aligns with the relevant policy concerns—information asymmetry and costs of litigation—while minimizing threats of widespread public confusion. Yet, even cost-shifting depends upon the claim proceeding to court. The misuse discussed in this Article consists mostly of trademark holders attempting to expand rights extra-judicially, without any intention of actually proceeding to court.

## **B. Statutory-Based Affirmative Claims**

An affirmative cause of action stands to deter conduct that would otherwise pass under the court's radar.<sup>183</sup> Conveniently, Section 512(f) of the Digital Millennium Copyright Act (DMCA) creates a counter-notification

---

181. Some courts have required this nexus in copyright misuse cases. *See* *Microsoft Corp. v. Jesse's Computers & Repair, Inc.*, 211 F.R.D. 681, 685 (M.D. Fla. 2002); *Microsoft Corp. v. Compusource Distribs., Inc.*, 115 F. Supp. 2d 800, 811 (E.D. Mich. 2000).

182. Section 35(a) of the Lanham Act authorizes an award of fees to the prevailing party only in "exceptional cases." 15 U.S.C. §1117(a); *see also* *Universal City Studios, Inc. v. Nintendo Co., Ltd.*, 797 F.2d 70, 77-78 (2d Cir. 1986) (upholding punitive damages for coercive assertion of inapplicable trademark rights, finding the claims amounted to "abuse of judicial processes" and "harm to the public as a whole").

183. Note, a few courts have already ruled against using trademark misuse as an independent cause of action. *See, e.g.*, *Ford Motor Co. v. Obsolete Ford Parts, Inc.*, 318 F. Supp. 2d 516, 521 (E.D. Mich. 2004) ("The court declines to announce or create an independent cause of action for trademark misuse."); *Dunn Computer Corp. v. Loudcloud, Inc.*, 133 F. Supp. 2d 823, 830 (E.D. Va. 2001) (holding that plaintiff's declaratory misuse claim was unripe).

procedure that serves as an ideal model.<sup>184</sup> Some substitutions and additions generate the following sample provision:

Any person who materially misrepresents that material or activity is infringing with an improper purpose shall be liable for any damages, including costs and attorneys' fees, incurred by the alleged infringer. If such misrepresentation was done willfully or knowingly, the court may award treble damages.

This provision would provide much-needed recourse for alleged infringers against overzealous trademark holders, and the threat of treble damages would help deter much of the conduct described in Part III. In short, an affirmative cause of action would best advance the misuse doctrine proposed in this Article.

Going forward, the primary impediment to an affirmative claim is the *Noerr-Pennington* doctrine, which, on the basis of the First Amendment's Petition Clause, immunizes from liability individuals who petition the judiciary by, for example, bringing a lawsuit.<sup>185</sup> Whether prelitigation threats of suit enjoy the same immunity as litigation itself under the *Noerr-Pennington* doctrine is currently unresolved.<sup>186</sup> If immune, no liability will arise from the trademark holder's cease-and-desist letter unless it qualifies as "sham" litigation under the Supreme Court's two-part test from *Professional Real Estate Investors, Inc. v. Columbia Pictures, Inc.*,<sup>187</sup> which requires the lawsuit to be "objectively baseless" and to intend to "interfere directly with the business relationships of a competitor."<sup>188</sup> Although many of the trademark holders' cease-and-desist letters cited above would

---

184. Section 512(f) reads:

Any person who knowingly materially misrepresents under this section (1) that material or activity is infringing [...] shall be liable for any damages, including costs and attorneys' fees, incurred by the alleged infringer, by any copyright owner or copyright owner's authorized licensee, or by a service provider, who is injured by such misrepresentation, as the result of the service provider relying upon such misrepresentation in removing or disabling access to the material or activity claimed to be infringing....

17 U.S.C. § 512(f).

185. *McGuire Oil Co. v. Mapco, Inc.*, 958 F.2d 1552, 1558 (11th Cir. 1992); see *United Mine Workers v. Pennington*, 381 U.S. 657 (1965); *E. R.R. Presidents Conference v. Noerr Motor Freight, Inc.*, 365 U.S. 127 (1961).

186. Compare *McGuire Oil Co. v. Mapco, Inc.*, 958 F.2d 1552, 1558-60 (11th Cir. 1992); *CVD, Inc. v. Raytheon Co.*, 769 F.2d 842, 850-51 (1st Cir. 1985); *Coastal States Mktg., Inc. v. Hunt*, 694 F.2d 1358, 1367 (5th Cir. 1983), with *Cardtoons, L.C. v. Major League Baseball Players Ass'n*, 208 F.3d 885 (10th Cir. 2000).

187. 508 U.S. 49 (1993).

188. *Id.* at 60-61 (emphasis in original).

satisfy this standard, such immunity would inhibit a trademark-misuse doctrine, at least one that adopts the affirmative cause of action described above.<sup>189</sup>

Most circuit courts have extended the *Noerr-Pennington* doctrine to apply to pre-litigation letters. The Fifth Circuit, for instance, reasons that “it would be absurd to hold that [petitioning immunity] does not protect those acts reasonably and normally attendant upon effective litigation. The litigator should not be protected only when he strikes without warning.”<sup>190</sup> Several commentators, including Herbert Hovenkamp, support this extension<sup>191</sup> on the theory that withholding immunity from prelitigation communication would curb practices that “facilitate the resolution of controversies.”<sup>192</sup> By contrast, an en banc court in the Tenth Circuit cabined the *Noerr-Pennington* doctrine, holding that the right to petition the government guaranteed by the First Amendment does not immunize prelitigation communications between private parties because such communications are not addressed to the government.<sup>193</sup>

The abusive behavior of trademark holders described in this Article highlights the imprudence of extending *Noerr-Pennington*'s immunity to prelitigation communication. No doubt, as Hovenkamp emphasizes, cease-and-desist letters have a legitimate role in preventing ongoing infringement. However, their utility is contingent on such communications (1) be-

---

189. See, e.g., *Thermos Co. v. Igloo Prods. Corp.*, No. 93 C 5826, 1995 WL 745832, \*6 (N.D. Ill. Dec. 13, 1995) (holding that “attempts to protect a valid and incontestable trademark” are privileged under the *Noerr-Pennington* doctrine).

190. *Coastal States*, 694 F.2d at 1367; see also *McGuire Oil Co. v. Mapco, Inc.*, 958 F.2d 1552 (11th Cir. 1992) (relying upon *Coastal States* in granting immunity to threats of litigation in an antitrust case); see, e.g., *Sosa v. Direct TV, Inc.*, 437 F.3d 923 (9th Cir. 2006); *IGEN Int'l, Inc. v. Roche Diagnostics GMBH*, 335 F.3d 303, 310 (4th Cir. 2003); *Cheminor Drugs, Ltd. v. Ethyl Corp.*, 168 F.3d 119, 128 (3d Cir. 1999); *South Dakota v. Kan. City S. Indus., Inc.*, 880 F.2d 40, 50-51 (8th Cir. 1989).

191. See, e.g., PHILLIP E. AREEDA & HERBERT HOVENKAMP, *ANTITRUST LAW* § 205e (rev. ed. 1997); HERBERT HOVENKAMP, *FEDERAL ANTITRUST POLICY* § 18.3d at 644 (1994) (immunizing prelitigation threats as vital to “[o]ur entire dispute resolution process[, which] is designed to encourage people to resolve their differences if possible before litigating”); see also Aaron R. Gary, *First Amendment Petition Clause Immunity from Tort Suits: In Search of a Consistent Doctrinal Framework*, 33 *IDAHO L. REV.* 67, 95 (1996) (“Innumerable federal and state courts have concluded that the *Noerr-Pennington* doctrine is rooted in the First Amendment right to petition and therefore must be applied to all claims implicating that right, not just to antitrust claims.”). But see Robert A. Zauzmer, Note, *The Misapplication of the Noerr-Pennington Doctrine in Non-Antitrust Right to Petition Cases*, 36 *STAN. L. REV.* 1243, 1251-53 (1984).

192. AREEDA & HOVENKAMP, *supra* note 191, at 237.

193. *Cardtoons, L.C. v. Major League Baseball Players Ass'n*, 208 F.3d 885 (10th Cir. 2000).

ing sent only to parties reasonably engaged in infringement and (2) accurately representing the property holder's rights and the potential ramifications for infringement. Yet the immunity urged by Hovenkamp provides no incentive to follow these conditions; in fact, it enables trademark holders to abuse this lawful tool with impunity.

## V. CONCLUSION

The persistent debate over the role of intellectual property on the internet usually gives the impression that only two solutions exist: wholesale strengthening or weakening of intellectual property rights. Or, as Paul Goldstein puts it, people view intellectual property as a glass half full or half empty.<sup>194</sup> But such competing approaches would try to fix far more than the problems they identify. For most of trademark law, the glass is about right; instead the problems cluster at the boundary of trademark law—a place in need of more precise solutions.

At this boundary, trademark law principally regulates potential infringers. This regime functioned well enough in the past because the law's ambit only extended to a clearly identifiable group: sellers of goods and services. But, as mentioned at the outset, the internet, by sweeping most of its users into trademark law's purview, confers to trademark holders power that is all too tempting to misuse. To respond, trademark law requires a revitalized doctrine of misuse to impose liability on trademark holders if they incorrectly assert their rights with a purpose unrelated to consumer confusion, especially if they do so in response to critique or parody. Only by shifting the focus of trademark law to the conduct of the trademark holder may it, in turn, refocus on the public welfare it intends to serve.

---

194. PAUL GOLDSTEIN, *COPYRIGHT'S HIGHWAY: THE LAW AND LORE OF COPYRIGHT FROM GUTENBERG TO THE CELESTIAL JUKEBOX* 15 (1994).