

# 23:4 BERKELEY TECHNOLOGY LAW JOURNAL

Pages  
1141  
to  
1530

Fall  
2008

**Production:** Produced by members of the *Berkeley Technology Law Journal*. 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 © 2008 Regents of the University of California.**  
All Rights Reserved.

*Berkeley Technology Law Journal*  
Berkeley Law  
West Basement, Room 3  
University of California, Berkeley  
Berkeley, California 94720-7200

(510) 643-6454 (Phone)  
(510) 643-6816 (Fax)  
btlj@law.berkeley.edu  
www.btlj.org

# BERKELEY TECHNOLOGY LAW JOURNAL

---

VOLUME 23

NUMBER 4

FALL 2008

## TABLE OF CONTENTS

### ARTICLES

- THE FORMAL STRUCTURE OF PATENT LAW AND THE LIMITS OF ENABLEMENT ..... 1141  
By Jeffrey A. Lefstin
- EMOTION, DILUTION, AND THE TRADEMARK CONSUMER ..... 1227  
By Laura R. Bradford
- RECENT DEVELOPMENTS AFFECTING THE ENFORCEMENT, PROCUREMENT, AND  
LICENSING OF RESEARCH TOOL PATENTS ..... 1299  
By Joshua D. Sarnoff & Christopher M. Holman
- CONGRESS SHOULD AMEND THE COPYRIGHT ACT TO PROTECT TRANSACTIONAL  
WATERMARKS ..... 1367  
By Matt Williams
- EXPLORING EUROPEAN UNION COPYRIGHT POLICY THROUGH THE LENS OF THE  
DATABASE DIRECTIVE ..... 1411  
By Miriam Bitton
- PHYSIOLOGICAL STEPS DOCTRINE ..... 1471  
By Andrew W. Torrance

### NOTES & COMMENTS COMPETITION WINNER

- SCHOOL ADMINISTRATORS AS CYBER CENSORS: CYBER SPEECH AND FIRST  
AMENDMENT RIGHTS ..... 1507  
By Tova Wolking

## SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN 1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited by the students of the University of California School of Law, Berkeley (Boalt Hall), and published four times each year (March, June, September, January) by the Regents of the University of California, Berkeley, Journal Publications, School of Law, 311 Boalt Hall, University of California, Berkeley, CA 94720-7200. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, 311 Boalt Hall, School of Law, University of California, Berkeley, CA 94720-7200.

**Correspondence.** Address all correspondence regarding subscriptions, address changes, claims for non-receipt, single copies, advertising, and permission to reprint to Journal Publications, 311 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 non-receipt 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 *Journal* conforms to THE CHICAGO MANUAL OF STYLE (15th ed. 2003) 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 23 BERKELEY TECH. L.J. \_\_\_\_ (2008).

## BTLJ ONLINE

The full text and abstracts of many previously published *Berkeley Technology Law Journal* articles can be found at <http://www.btlj.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-submissions@law.berkeley.edu](mailto:btlj-submissions@law.berkeley.edu). Authors should submit double-spaced, single-sided manuscripts with generous margins. 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.

**Mailing Address.** BTLJ only accepts electronic submissions sent as Microsoft Word attachments to [btlj-submissions@law.berkeley.edu](mailto:btlj-submissions@law.berkeley.edu).

# DONORS

The *Berkeley Technology Law Journal* and the Berkeley Center for Law & Technology acknowledge the following generous donors to Boalt Hall's Law and Technology Program:

## Benefactors

COOLEY GODWARD KRONISH LLP	PERKINS COIE LLP AND AFFILIATES
COVINGTON & BURLING LLP	SKADDEN, ARPS, SLATE, MEAGHER & FLOM LLP AND AFFILIATES
FENWICK & WEST LLP	WEIL, GOTSHAL & MANGES LLP
KIRKLAND & ELLIS LLP	WHITE & CASE LLP
LATHAM & WATKINS LLP	WILSON, SONSINI, GOODRICH & ROSATI
MORRISON & FOERSTER LLP	WINSTON & STRAWN LLP
ORRICK, HERRINGTON & SUTCLIFFE LLP	

## **Members**

ALSTON + BIRD LLP	HICKMAN, PALERMO, TRUONG & BECKER, LLP
BAKER BOTTS LLP	JONES DAY LLP
BINGHAM MCCUTCHEN LLP	KEKER & VAN NEST LLP
DLA PIPER RUDNICK GRAY CARY	KNOBBE MARTENS OLSON & BEAR LLP
DAY CASEBEER MADRID & BATCHELDER LLP	MAYER BROWN LLP
DECHERT LLP	MCDERMOTT WILL & EMERY
DREIER STEIN KAHAN BROWNE WOODS GEORGE LLP	MORGAN, LEWIS & BOCKIUS LLP
FINNEGAN HENDERSON FARABOW GARRETT & DUNNER LLP	ROPES & GRAY LLP
FISH & RICHARDSON P.C.	SONNENSCHN NATH & ROSENTHAL LLP
GOODWIN PROCTER LLP	TOWNSEND AND TOWNSEND AND CREW LLP
GUNDERSON DETTMER STOUGH VILLENEUVE FRANKLIN & HACHIGIAN, LLP	WILMER CUTLER PICKERING HALE AND DORR LLP
HELLER EHRMAN LLP	

## **Patrons**

BAKER & MCKENZIE LLP	O'MELVENY & MYERS LLP
GREENBERG TRAURIG, 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 Center for Law and Technology*, Boalt Hall School of Law, 355 Boalt Hall, University of California, Berkeley, California 94720. Telephone: 510-642-8073. E-mail: [bclt@law.berkeley.edu](mailto:bclt@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  
Morrison Foerster LLP  
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

# 2008-2009

---

*Editor-in-Chief*

JUSTIN LEE

*Managing Editor*

CAMERON YAHR

*Senior Article Editors*

BRANDON BROWN  
ASHLEY DOTY

*Senior Executive Editor*

KIM NATIVIDAD

*Senior Annual Review Editors*

TAMAR GUBINS  
DANNY PRATI

*Executive Editors*

---

*Submissions Editors*

AARON EPSTEIN  
JADE HOFFMAN

*Production Editors*

ALICE AHN  
HILLIARY CREELY

*External Relations Editors*

REED MCBRIDE  
NATHAN SALHA

*Bluebook Editors*

ALYSSA KNUTSON  
JENNIFER SAIONZ

*Notes & Comments Editors*

TOLGA GULMEN  
JENNIFER RAAB

*Symposium Editors*

SUNIL BECTOR  
ANDREW DUFRESNE

*Member Relations Editor*

HERSH REDDY

*Annual Review Editors*

ADAM BLANKENHEIMER  
NATHAN KAMESAR

*Web Editor*

AARON BRAUER-RIEKE

*Assistant Managing Editor*

JENNIFER MIDDLETON

*Publishing Editor*

MARGARET ZIEGLER

*Article Editors*

---

ANGELIQUE ABED  
CESAR ALVAREZ  
VARTY DEFTERDERIAN  
JUSTIN HAAN

ELIZABETH IGLESIAS  
ANTHONY JACOBSON  
VIVIAN KIM  
PETER NAGLE  
MICHELE PATTON

MIKE SAWYER  
STEPHEN ULLMER  
JOEL WALLACE  
ROSAMOND XIANG

# BERKELEY CENTER FOR LAW & TECHNOLOGY

---

*Executive Director*

ROBERT BARR

*Faculty Directors*

AMY KAPCZYNSKI  
DEIRDRE MULLIGAN  
HOWARD SHELANSKI '92

PETER MENELL  
PAMELA SAMUELSON

ROBERT MERGES  
PAUL SCHWARTZ  
MOLLY VAN HOUWELING

*Assistant Director*

LOUISE LEE

*Assistant Director*

DAVID GRADY

*Affiliated Faculty and  
Scholars*

---

STEPHEN BARNETT  
ROBERT BERRING '74  
AARON EDLIN  
JOSEPH FARRELL  
RICHARD GILBERT  
BRONWYN HALL  
THOMAS JORDE

MICHAEL KATZ  
DAVID MOWERY  
ERIN MURPHY  
DAVID NIMMER  
DANIEL RUBINFELD  
ANNALEE SAXENIAN  
SUZANNE SCOTCHMER

CARL SHAPIRO  
MARJORIE SHULTZ  
LON SOBEL  
DAVID TEECE  
HAL R. VARIAN  
OLIVER WILLIAMSON  
BRIAN WRIGHT

# MEMBERSHIP

# 2008-2009

---

## *Associate Editors*

---

AMIR AMIRI	ELIZABETH ERAKER	LUCIAN ORBAI
ALEX BAXTER	CHRIS FRANKLIN	DANIEL PARK
KYLE BRADY	KENNETH GANTZ	EZEKIEL RAUSCHER
TAYLOR BURRAS	KRISTIN KEMNITZER	JACLYN SETILI
CHARLOTTE CHANG	PAN LEE	TOM SPAHN
ALLISON CROW	YITE JOHN LU	ALLEN WANG
EARTH DUARTE- TRATTNER	ARASH MOUSSAVIAN	ALEX WIDMANN
APRIL ELLIOTT	SHAMILA NATHAN	MENG XI
	LIZ OFFEN-BROWN	KATHY YU

---

## *Members*

---

JASON AMSEL	DANIEL KIM	PAUL PARK
SARA ANSARI	ASHLEY KUSTU	JAKE PHILLIPS
TOPE AWOYODE	JOANNE KWON	JAY PURCELL
DAVID BARNHILL	BRIAN LAHTI	SHAHRZAD RADBOD
CHRIS BEBENEK	BERT LAO	VIVEK RAO
JESSICA BROWN	MITCH LAUFER	JOE ROSE
ALEXIS BURGESS	DAVID LEE	PHILIPP SANDNER
MICHAEL CHENG	STEPHANIE LEE	VALENTINA SCHULTE- BRAUCKS
CHARLES CIACCIO	TERRY LI	GRACE SHEN
MATT DAL SANTO	AMANDA LOPEZ	SCOTT SIERA
ERIC FONG	DAVID LOWRY	SPEARLT
CANDACE FRAZIER	JESSICA LYON	DAVID STARK
SARA GIARDINA	ELANOR MANGIN	JESSE STEINBACH
MORGAN HAGUE	DAVID MARTY	PAUL STELLMAN
LAURA HALL	JEROME MATTHEWS	DAVE TROSSEN
PETER HALPERN	ROGER MICHALSKI	CHRISTINA TSAKONA
HEATHER HANEY	TODD MOBLEY	LINFONG TZENG
DEVIN HECKMAN	WILL MOSELEY	PRANAVA UPADRASHTA
NIMA HEFAZI	JANA MOSER	JULIA VAN DE WALLE
BLAKE HOLLAND	ANNA NATHORST- WESTFELT	SHANE WITNOV
ALEX JADIN	TAL NIV	ALAN XIANG
TANEEN JAFARKHANI	KURUVILLA OLASA	CECILIA ZINITI
RYAN KELLY	ANGELICA ORNELAS	
CHANG KIM	HENRY PAN	



# THE FORMAL STRUCTURE OF PATENT LAW AND THE LIMITS OF ENABLEMENT

By Jeffrey A. Lefstin, Ph.D.<sup>†</sup>

## TABLE OF CONTENTS

I. INTRODUCTION: THE NEW FORMALISM AND THE LAW OF ENABLEMENT .....	1142
A. THE NEW FORMALISM IN PATENT LAW .....	1142
B. PATENT DISCLOSURE .....	1151
II. PATENT LAW AS A FORMAL SYSTEM .....	1159
III. THE LIMITS OF ENABLEMENT .....	1167
A. THE PROBLEM OF INFINITE SCOPE.....	1168
1. <i>All the World's a Genus: All Claims are Infinite</i> .....	1168
2. <i>What is the "Full Scope" of Infinite Scope?</i> .....	1174
B. CAN ENABLEMENT LIMIT CLAIM SCOPE? .....	1181
1. <i>Claim Scope Without Written Description</i> .....	1182
2. <i>The Relation Between Enablement and Nonobviousness</i> .....	1186
C. RETHINKING ENABLEMENT .....	1189
1. <i>Converging Enablement and Nonobviousness</i> .....	1190
2. <i>"Enablement Plus"</i> .....	1193
IV. WRITTEN DESCRIPTION REVISITED .....	1197
A. WHAT WRITTEN DESCRIPTION IS NOT .....	1197
1. <i>Written Description as Possession</i> .....	1197
2. <i>Written Description as a Priority Doctrine Only</i> .....	1200
B. WHAT WRITTEN DESCRIPTION IS: A DOCTRINE OF DEFINITION.....	1204
C. LOSING THE PATH .....	1207
D. ANCHORING THE DEFINITIONAL HIERARCHY.....	1211
V. THE SIGNIFICANCE OF DEFINITIONAL INFORMATION .....	1215
A. CONSEQUENCES OF AN OBLIGATION FOR DEFINITIONAL INFORMATION .....	1215

---

© 2008 Jeffrey A. Lefstin, Ph.D.

<sup>†</sup> Associate Professor of Law, University of California, Hastings College of the Law. The author thanks Margreth Barrett, Robin Feldman, John Golden, Timothy Holbrook, Evan Lee, Michael Meurer, Craig Nard, and participants at the Michigan State University College of Law Intellectual Property and Telecommunications Workshop, and the Eighth Annual Intellectual Property Scholars' Conference for useful comments.

B. MEANING AND THE PLACE OF DEFINITIONAL INFORMATION IN PATENT DOCTRINE.....	1220
VI. CONCLUSION: THE PAST AND FUTURE OF CLAIMING .....	1222

## I. INTRODUCTION: THE NEW FORMALISM AND THE LAW OF ENABLEMENT

### A. The New Formalism in Patent Law

Modern patent law has recapitulated the circumstances of its birth. American patent law was first consolidated in the period between the Civil War and the end of the 19th Century, when the highly formal system of thought, described as classical legal orthodoxy, captivated the American legal mind.<sup>1</sup> As an intellectual structure, classical legal orthodoxy was primarily characterized by:

A strong commitment to abstract legal categories, and the clear differentiation of one category from another.<sup>2</sup>

A desire to derive bottom-level legal rules analytically from a few basic top-level categories and higher principles, akin to Euclid's derivation of the whole of geometry from five fundamental axioms.<sup>3</sup> Inherent in this process was the condensation of legal rules, previously scattered among functional categories or forms of action, around a few key principles such as negligence and fault in tort, or offer, acceptance, and consideration in contract.<sup>4</sup>

A preference for objective rules over vague standards.<sup>5</sup> If not motivated by the ascendant business community's demand for legal predictability, this preference met the community's needs for clear legal rules upon which investment decisions could be predicated.<sup>6</sup>

Classical legal orthodoxy's emphasis on generalization, abstraction, and categorization extended to nearly all fields of American law,<sup>7</sup> and

---

1. See Thomas C. Grey, *Langdell's Orthodoxy*, 45 U. PITT. L. REV. 1, 2 (1983).

2. See *id.* at 12.

3. See *id.* at 11-12, 16.

4. See *id.* at 9-10.

5. See *id.* at 11.

6. See *id.* at 33-35.

7. See MORTON J. HORWITZ, *THE TRANSFORMATION OF AMERICAN LAW 1870—1960: THE CRISIS OF LEGAL ORTHODOXY* 10-31 (1992).

patent law was no exception.<sup>8</sup> Like many other bodies of law, patent law became committed to a formal and abstract system inspired by the logical principles of science. The most fundamental change in patent law during this period concerned the mode by which patentee's rights were defined. Before the mid-19th Century, patent infringement was judged simply by comparing the invention described in the patent's disclosure with the accused device or process.<sup>9</sup> However, over the latter half of the century, an inventor's rights came to be defined not by what the inventor actually made or disclosed, but by formal "claims" that specified the precise boundaries of the inventor's exclusive right. Precise claiming provided clear notice of the patent's boundaries to the public, competitors, and other inventors. For classical legal thinkers, however, a precise claiming system was more than a means to achieve predictability in patent rights. It was essential in a conceptually ordered legal system. The deep theoretical connection between the claim system and the ideal of conceptual order is clearly evident in *Merrill v. Yeomans*, a pivotal case that cemented the primacy of the patent claim:

The growth of the patent system in the last quarter of a century in this country has reached a stage in its progress where the variety and magnitude of the interests involved require accuracy, precision, and care in the preparation of all the papers on which the patent is founded. It is no longer a scarcely recognized principle, struggling for a foothold, but it is an organized system, with well-settled rules, supporting itself at once by its utility, and by the wealth which it creates and commands. The developed and improved condition of the patent law, and of the principles which govern the exclusive rights conferred by it, leave no excuse for ambiguous language or vague descriptions. The

---

8. See generally Samuel Oddi, *Regeneration in American Patent Law: Statutory Subject Matter*, 46 IDEA 491, 520-34 (2006).

9. See Karl B. Lutz, *Evolution of the Claims of U.S. Patents*, 20 J. PAT. OFF. SOC'Y 134, 147 (1938) (explaining that in the period from 1836 to 1870 "claims rarely, if ever, received consideration on the question of infringement"). This practice may seem puzzling today, but the notion that the patentee has exclusive rights to a set of things bearing the properties recited by the claim is a relatively recent development. In the early 19th Century, the patentee was required to show what was *novel* about the invention. A patentee could not describe as his invention, a clock for example, with a novel mainspring, because the other components of the clock would be "old" and the patentee would be accused of claiming the old and the new together. Claims defined the novel feature or principle of the invention, not necessarily an actual embodiment, and therefore ensured that the patent would not be held invalid for lack of novelty. See also text accompanying notes 259-265 *infra*.

public should not be deprived of rights supposed to belong to it, without being clearly told what it is that limits these rights. The genius of the inventor, constantly making improvements in existing patents—a process which gives to the patent system its greatest value—should not be restrained by vague and indefinite descriptions of claims in existing patents from the salutary and necessary right of improving on that which has already been invented. It seems to us that nothing can be more just and fair, both to the patentee and to the public, than that the former should understand, and correctly describe, just what he has invented, and for what he claims a patent.<sup>10</sup>

For the *Merrill* Court, adequate notice to the public and other inventors was not the only consideration in favor of limiting the patentee's rights to the claim. In parallel with these practical concerns was an intellectual linkage between the system of claiming and the ideal of patent law as “an organized system, with well-settled rules.”<sup>11</sup> An ordered and logical system of patent law demanded nothing less than an ordered and logical system for defining patent rights. The system of peripheral claiming, in which the claims set forth the boundaries of the patent, served both ends.

If the late 19th Century was the first great period of formalism in American patent law, then the past several decades have been the second. It is critical to understand that by “formalism,” I do not mean simply that the substantive doctrines of patent law have gravitated towards bright-line rules in place of looser standards, although scholars have identified such trends in the patent jurisprudence of the Supreme Court and the Court of Appeals for the Federal Circuit.<sup>12</sup> Instead, I refer primarily to the classical ideal of a conceptually ordered system founded upon a small number of abstract categories and concepts. Thus by formalism I mean that the courts have become committed to a highly formal conception of the patent itself; it is the legal structure of the patent system that has become formal, not just its particular doctrines.<sup>13</sup>

---

10. *Merrill v. Yeomans*, 94 U.S. 568, 573-574 (1876).

11. *Id.* at 573.

12. See Timothy R. Holbrook, *The Supreme Court's Complicity in Federal Circuit Formalism*, 20 SANTA CLARA COMP. & HIGH TECH. L.J. 1 (2003); John R. Thomas, *Formalism at the Federal Circuit*, 52 AM. U. L. REV. 771 (2003) (citing Holbrook). The Federal Circuit since 1982 has been vested with nearly complete and exclusive appellate jurisdiction over questions of patent law.

13. See Grey, *supra* note 1, at 1-12 (distinguishing between a preference for rules over standards, and a commitment to analytically deriving those rules from top-level

The cornerstone of that formal system is the claim, which today is the defining characteristic of patent law. Claims delineate exactly the subject matter over which the inventor is entitled to exclude others from manufacture, use, or sale. The claim is the sole measure of the invention; all questions of patent infringement, validity, and inventorship, are resolved by reference to the subject matter defined by the claim.<sup>14</sup> In modern parlance, the claim, “the invention,” and “the patent” are essentially synonymous.<sup>15</sup>

Modern claims are themselves highly formal entities. They recite a set of characteristics, or properties, that define the subject matter encompassed by the patent. The more properties or characteristics the claim recites, the smaller the scope of the subject matter it defines. This structure corresponds to the concepts of intension and extension prevalent in classical logic and derived ultimately from Aristotle: as the number of properties or characteristics defining a class grows larger, the number of objects to which it applies becomes smaller.<sup>16</sup> Most patents contain ordered, hierarchical pyramids of claims in which more and more properties are recited to define successively smaller slices of subject matter. Typically, patentees begin with a broad claim reciting as few properties as possible to yield the broadest claim permissible in light of the prior art and the patentee’s disclosure. To these broad claims they generally attach narrower “dependent” claims. Dependent claims refer back to a broader independent claim, but recite additional properties and therefore define a smaller sub-category of subject matter within the independent claim.<sup>17</sup>

---

principles). According to Grey, late 19th-Century classical legal orthodoxy was committed to formalism at both levels. *See id.* at 10-11.

14. An exception to this principle is that one who is co-inventor of a claim receives co-ownership of *all* the claims in the patent. *See Ethicon, Inc. v. U.S. Surgical Corp.*, 135 F.3d 1456, 1460 (Fed. Cir. 1998).

15. Patents typically carry more than one claim, each of which technically defines an “invention.” Though the essential synonymy of invention and claim was well-established by the time of the 1952 Patent Act, the equivalence of the invention and claim is never explicitly demanded by the Act itself. The Act’s substantive validity and infringement provisions speak of “the invention” rather than “a claim”; section 112 of the Act merely requires the patentee to conclude the specification with one or more claims distinctly pointing out what he regards as his invention. 35 U.S.C. § 112 (2000).

16. *See WILLIAM T. PARRY & EDWARD A. HACKER, ARISTOTELIAN LOGIC* 65-67 (1991).

17. The practical motivation to construct hierarchies of successively narrower claims is to ensure that some claims remain valid. If broad claims are invalidated because it is later discovered that they read upon the prior art, or are too broad in light of the

Though claims have been the primary measure of the inventor's rights since the mid-19th Century, the Federal Circuit's jurisprudence seems to have driven towards an ideal in which the patentee has an absolute entitlement to all things within the boundaries defined by the claims, but, with the possible exception of developments unforeseeable at the time of patenting,<sup>18</sup> no rights over any things outside the literal boundaries of the claim. In the law of infringement, the court has worked towards a regime in which *any* use of subject matter falling within the claims is an act of infringement, regardless of its extent or purpose.<sup>19</sup> Conversely, with its hostility to the doctrine of equivalents,<sup>20</sup> the court has tried to prevent patentees from asserting infringement against subject matter lying outside the literal scope of the claims.<sup>21</sup> With respect to patent validity, questions

---

patentee's disclosure, the patentee may be able to fall back on narrower dependent claims that are still valid. Because narrower claims encompass less subject matter, they are less likely to encompass prior art or subject matter that the patentee did not enable or describe.

18. *See Johnson & Johnston Assocs. Inc. v. R.E. Serv. Co.*, 285 F.3d 1046, 1056 (Fed. Cir. 2002) (Rader, J., concurring) (arguing that doctrine of equivalents should not extend to "subject matter that the patent drafter reasonably could have foreseen during the application process and included in the claims"). Judge Rader's argument that foreseeability ought to be the sole principle underlying the doctrine of equivalents recalls the classical program of systematizing unruly legal regimes around central organizing principles.

19. The court has essentially denied the existence of a common-law experimental-use exemption or an exception for "de minimis" infringement, *see Madey v. Duke Univ.*, 307 F.3d 1351 (Fed. Cir. 2002), and (before being reversed by the Supreme Court) accorded narrow scope to the statutory exemption for activities directed to approval of generic drugs mandated by the Hatch-Waxman Act, *see Merck KGaA v. Integra Lifesciences I, Ltd.*, 331 F.3d 860 (Fed. Cir. 2003), *rev'd*, 545 U.S. 193 (2005). The Federal Circuit's rule that permanent injunctions would issue upon proof of infringement absent exceptional circumstances was another example of the absolutist strain in the law of infringement. The Supreme Court limited that rule in *Ebay v. MercExchange*, holding that injunctions should issue only upon satisfaction of the traditional tests for equitable relief. *See eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006).

20. Created by the courts, the doctrine of equivalents ("DOE") permits a patentee to reach beyond the literal scope of the claims. Under the doctrine of equivalents, accused subject matter may infringe if its elements or properties are substantially similar to those defined by the patent's claims. *See Warner-Jenkinson Co. v. Hilton Davis Chemical Co.*, 520 U.S. 17, 21 (1997) ("[A] product or process that does not literally infringe upon the express terms of a patent claim may nonetheless be found to infringe if there is 'equivalence' between the elements of the accused product or process and the claimed elements of the patented invention.").

21. In *Johnson & Johnston Associates Inc. v. R.E. Service Co.*, the court has held that any subject matter disclosed, but not explicitly claimed, in the patent's specification is abandoned to the public and cannot be reached under the doctrine of equivalents,

of novelty have been reduced to exact rules of inclusion or exclusion within the boundaries of the claim, in derogation of the more nuanced approach of earlier times.<sup>22</sup>

In addition to reducing the substantive doctrines of patent law to formal questions of inclusion or exclusion with respect to claim scope, the Federal Circuit, since its founding, has maintained a rigid conceptual separation between the substantive doctrines themselves. The questions of patent infringement and patent validity are both binary determinations; the patent is infringed or not, and it is valid or invalid. The court has determined patent validity independently of the infringement inquiry and without reference to the allegedly infringing device. In particular, the question of patent scope—whether the patentee is entitled to assert

---

though the court later held that the alleged equivalent must be disclosed specifically in the written description to trigger the dedication rule. *See generally Johnson & Johnston*, 285 F.3d 1046; *see also* *PSC Computer Prods. Inc. v. Foxconn Int'l, Inc.*, 355 F.3d 1353, 1360 (Fed. Cir. 2004). The court also attempted to impose a strict regime of prosecution history estoppel in which any subject once within the claims during prosecution, but not within the final claims, was surrendered and beyond the reach of the doctrine of equivalents. *See Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 234 F.3d 558 (Fed. Cir. 2000). The Supreme Court tempered the Federal Circuit's absolute rule by specifying when the patentee could rebut a presumption of surrender. *See Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002). While the *Festo* and *Johnson & Johnston* cases drew public attention to the court's hostility to the DOE, Professor Nard had earlier identified this trend commencing in 1991. *See* Craig Allen Nard, *A Theory of Claim Interpretation*, 14 HARV. J. L. & TECH. 1, 68-69 (2000) (noting skepticism towards DOE in *London v. Carson Pirie Scott & Co.*, 946 F.2d 1534, 1538 (Fed. Cir. 1991)).

22. In assessing whether an invention has been anticipated by a disclosure in the prior art, the court has held that *any* prior use or sale of subject matter falling within the scope of the claims invalidates the claim, regardless of whether the use or even existence of the subject matter was known at the time. *See Schering Corp. v. Geneva Pharms.*, 339 F.3d 1373, 1377-80 (Fed. Cir. 2003). Earlier case law, which the Federal Circuit in *Schering* characterized as dicta, had suggested that prior accidental or unknown existence of subject matter within the scope of the claims might not render the patent invalid. *See id.* at 1378-79 (dismissing *In re Seaborg*, 328 F.2d 996 (C.C.P.A. 1964)). Other developments towards a regime of absolute novelty include expanding the range of prior art under section 102(g) of the Patent Act that can destroy patentability despite the absence of publication or public knowledge. *See, e.g., Thomson, S.A. v. Quixote Corp.*, 166 F.3d 1172, 1175 (Fed. Cir. 1999) (explaining that activity not otherwise "prior art" may anticipate claim unless abandoned, suppressed, or concealed). Of course, the benchmark for bright-line rules of public use was laid down during the heyday of classical legal orthodoxy, when the Supreme Court held that corset springs were in "public use" when their inventor gave his "intimate friend" a single pair to wear within her corset. *See Egbert v. Lippmann*, 104 U.S. 333, 335 (1881).

exclusive rights over a broad range of subject matter—is resolved independently from the question of whether the patentee is entitled to assert exclusive rights over the particular subject matter practiced by the accused infringer.<sup>23</sup> Formally, it is no defense that the accused infringer is practicing something which was in the public domain before the patent,<sup>24</sup> nor is it a defense that the accused technology is beyond what the patentee’s disclosure enabled. Once the validity is determined, there is no formal relationship between what the patentee is asserting rights over (the accused subject matter) and either the patentee’s disclosure, or the prior art.<sup>25</sup>

---

23. Historically, patent law *could* take into account the relationship between these inquiries via the so-called “reverse doctrine of equivalents,” which exonerated accused subject matter of infringement even though it fell within the literal boundaries of the patent’s claims. This doctrine permitted a court to assess infringement in light not only of the degree to which the claimed invention represented an advance over the prior art, but the marginal advance and functional similarity of the accused subject matter as well. *See, e.g.,* *Boyden Power-Brake Co. v. Westinghouse*, 170 U.S. 537, 569 (1898) (“Even if the patent for a machine be a pioneer, the alleged infringer must have done something more than reach the same result.”). However, the Federal Circuit has suggested that the 1952 Patent Act essentially destroyed the doctrine, and has never affirmed a finding of noninfringement under the reverse doctrine of equivalents. *See* *Tate Access Floors, Inc. v. Interface Arch. Res., Inc.*, 279 F.3d 1357, 1368 (Fed. Cir. 2002). But the disappearance of the reverse doctrine of equivalents has little to do with the passage of the 1952 Act, which largely codified common-law patent doctrine. The reverse doctrine of equivalents is untenable in modern patent law because it is premised on the existence of “the invention” and “the claims” as separate entities, or at least on the significance of that distinction. *See* *Boyden Power-Brake*, 170 U.S. at 568 (explaining that the doctrine exonerates the defendant who “has so far changed the principle of the device that *the claims of the patent, literally construed, have ceased to represent his actual invention*” and invoking a distinction between violation of the letter of a statute as opposed to the spirit or intent thereof) (emphasis added). Such a distinction is incompatible with the modern synonymy of claim and invention.

24. *See* *Tate Access Floors*, 279 F.3d at 1365-66 (refusing to accept “practicing the prior art” as a defense to literal infringement). However, infringement under the doctrine of equivalents is explicitly limited by the scope of the prior art. *See* *Wilson Sporting Goods Co. v. David Geoffrey & Assocs.*, 904 F.2d 677, 683 (Fed. Cir. 1990) (“Even if this test is met . . . there can be no infringement if the asserted scope of equivalency of what is literally claimed would encompass the prior art.”).

25. The questions of validity, claim scope, and infringement are still connected *in practice* by the question of claim interpretation. The parties in suit tailor their claim interpretations to suit their arguments on infringement and validity; a broader claim is more likely to be infringed but less likely to be valid, and vice versa. Moreover, one of the maxims of claim interpretation is that claims should be construed, if possible, to preserve their validity. Thus, as a matter of interpretation, claim scope is determined with an eye towards the arguments raised in connection with infringement and invalidity. If,

The program of condensing patent law into a rigorously defined system centered on the claim has not proceeded without interruption. The Supreme Court in particular has resisted the Federal Circuit's tendencies toward an absolutist or minimalist model of patent infringement.<sup>26</sup> However, there can be little doubt that classical legal theorists would have received favorably the vision of patent law as a set of conceptually differentiated, binary determinations founded on the abstract concept of the claim,<sup>27</sup> as well as the Federal Circuit's desire to provide certain and

---

however, the claims are textually clear and unambiguous, there is no room for interpretive discretion, and the validity and infringement determinations proceed as entirely disconnected inquiries. We might regard clarity in claim scope as the sine qua none of the formalist program: if claim scope is not certain, then that uncertainty limits the precision of patent determinations no matter how closely the substantive doctrines adhere to the boundaries of the claims. Two trends in the Federal Circuit's claim construction jurisprudence—the emphasis on dictionaries as a source of meaning, and the suggestion that claim construction should proceed by an ordered algorithm—have attempted to formalize claim interpretation as well. However, the Federal Circuit put an end to these trends in its en banc decision in *Phillips*, de-emphasizing the role of dictionaries and denying the existence of any rigid structure to the claim construction process. The formalist program might well be futile if the boundaries of claims cannot be determined with precision. However, as an *intellectual* structure, the claim system is still conceptually ordered if a single principle governs the outcome of claim construction. Thus, the principle of *Phillips v. AWH Corp.*—that claim terms mean what an ordinary artisan in the field of the invention would think them to mean after having read the patent specification and prosecution history—provides a conceptually ordered unity to claim interpretation regardless of whether the outcome in individual cases is determinate or not. *Phillips* clearly reflects the aspirations of classical legal thinkers: a single unifying principle from which the bottom-level rules of the regime may be derived. See *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (en banc) (emphasizing role of context in claim interpretation).

26. The Supreme Court has restored some of the ground lost by the doctrine of equivalents, see *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002); broadened the scope of the statutory exemption for generic drug approval under the Hatch-Waxman Act, see *Merck KGaA v. Integra Lifesciences I, Ltd.*, 545 U.S. 193 (2005); and emphasized that the grant of injunctive relief is subject to the traditional principles of equitable discretion, see *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006). One might also view the Supreme Court's decision in *KSR Int'l Co. v. Teleflex Inc.*—which declared that the Federal Circuit's "teaching, suggestion, or motivation" test for combining prior art references was but one of several permissible ways to demonstrate obviousness—as rejecting the notion that obviousness can be condensed around a single unifying principle. See *KSR Int'l Co. v. Teleflex Inc.*, 127 S. Ct. 1727 (2007).

27. Horwitz describes how classical legal thinking was dominated not only by the tendency to draw bright-line classifications of legal phenomena but also by the tendency to structure legal inquiries as binary questions of inclusion or exclusion from those abstract categories. See HORWITZ, *supra* note 7, at 17-18. Twentieth-Century legal

predictable patent law upon which investment decisions may be based.<sup>28</sup> So too would classical legal theorists endorse the efforts by some Federal Circuit judges to condense complex bodies of patent law around single unifying principles, such as the notion that the subject matter requirement of section 101 may be reduced to the question of whether an invention yields a “useful, concrete, and tangible result.”<sup>29</sup> And just as classical legal theorists sought to order law as a system of rules formally derivable from a minimum set of higher-order principles, the vision behind the formalist program seems to be a patent law in which the rules of infringement and validity can essentially be derived from a few simple axioms. We might in fact represent the formalist ideals of patent eligibility, patent validity, and patent infringement by three axioms:

All things useful may be claimed;

---

thinking, according to Horwitz, was more receptive to balancing tests. *Id.* at 131. One could, for example, conceive of a patent system in which the question of infringement might depend on the degree to which a patent was novel and nonobvious, rather than being an entirely separate question once the statutory thresholds of novelty and nonobviousness have been met.

28. *See, e.g., Festo Corp. v. Shoketsu Kinzoku Kabushiki Co.*, 234 F.3d 558, 577-78 (Fed. Cir. 2000) (justifying limitations on doctrine of equivalents as stimulating investment in improvements by competitors); *Aerojet-Gen. Corp. v. Mach. Tool Works, Oerlikon-Buehrle Ltd.*, 895 F.2d 736, 744 (Fed. Cir. 1990) (en banc) (“The availability of a clear, stable, uniform basis for evaluating matters of patent validity/invalidity and infringement/noninfringement renders more predictable the outcome of contemplated litigation, facilitates effective business planning, and adds confidence to investment in innovative new products and technology.”), *overruled by*, *Holmes Group, Inc. v. Vornado Air Circulation Sys., Inc.*, 535 U.S. 826 (2002).

29. *State St. Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368, 1375 (Fed. Cir. 1998); Thomas, *supra* note 12, at 788 (“It is difficult to imagine a more simple rule governing patent-eligible subject matter.”). The Federal Circuit *en banc* overruled the *State Street* “useful, concrete, and tangible result” standard in *In re Bilski*, 545 F.3d 943, 959-60 (Fed. Cir. 2008) (overruling *State Street*). Judge Rader, however, believed that the court’s complex exegesis of Supreme Court precedent and its “machine or transformation” test could have been condensed into the single principle that abstract ideas are not patentable. *See id.* at 1011 (Rader, J., dissenting) (“This court labors for page after page, paragraph after paragraph, explanation after explanation to say what could have been said in a single sentence: ‘Because Bilski claims merely an abstract idea, this court affirms the Board’s rejection.’”). The same desire to condense law around a single principle—so attractive to classical legal theorists—seems to explain Judge Rader’s view that the myriad restrictions on the doctrine of equivalents may be condensed into the question of whether the alleged equivalent was foreseeable. *See Johnson & Johnston Assocs. Inc. v. R.E. Serv. Co.*, 285 F.3d 1046, 1056-59 (Fed. Cir. 2002) (Rader, J., concurring); *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 344 F.3d 1359, 1374-77 (Fed. Cir. 2003) (en banc) (Rader, J., concurring).

No claim may have within its boundaries any thing existing in, or obvious from, the prior art; and

All things within the boundaries of the claim infringe, and nothing outside those boundaries infringes.

## **B. Patent Disclosure**

The three axioms given above are insufficient to derive a conceptually ordered system of patent law, because we have not defined an axiom of permissible claim scope. While the doctrines of novelty and nonobviousness define the limits of the inventor's claims imposed by the prior art, an axiom of claim scope must define the extent of the inventor's entitlement as a function of what the inventor has created or described in his patent application. The question of permissible claim scope lies at the heart of patent law. The extent of the inventor's entitlement determines not only the incentive offered to the patentee for creating his invention and disclosing it to the public, but also the balance of incentives between the patentee and future inventors whose improvements may fall within the scope of the patent's claims.

Patent claims delimit the inventor's entitlement. In turn, how broadly the inventor is entitled to claim depends on the nature and extent of what the inventor discloses about the invention in his specification. The modern expression of these principles is in section 112 of the 1952 Patent Act:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.<sup>30</sup>

The second paragraph of section 112 requires that the inventor delimit the scope of his invention with claims. In the jurisprudence of the Federal Circuit and one of its predecessor courts, the Court of Customs and Patent Appeals, this obligation is known as the "definiteness" requirement.

---

30. 35 U.S.C. § 112 (2000). The third through sixth paragraphs of section 112 control the drafting of dependent claims and claims defining the invention by function rather than by structure or, in the case of processes, by acts.

Modern definiteness doctrine is concerned not with how broad the claims are, but whether the claim language clearly communicates to the public the boundaries of the patent right.<sup>31</sup> The requirement of linguistic definiteness is modest: only if claims are “not amenable to construction” or “insolubly ambiguous” will they be invalid as indefinite.<sup>32</sup>

In contrast, the first paragraph of section 112 requires that the inventor disclose his invention in the patent specification. One of the central functions of section 112, ¶ 1 is to implement the quid pro quo of the patent system: in exchange for disclosing the invention to the public, the inventor receives for a limited time the exclusive rights provided by the patent grant. As set forth by the Court of Customs and Patent Appeals and the Federal Circuit, the doctrinal elements of section 112 are threefold. One element is the “best mode” requirement, which obligates the inventor to disclose (if he has one) his preferred mode of carrying out the claimed invention. The best mode requirement is not a scope doctrine in the sense that I use the term. Although the scope of the claims determines whether a particular preference held by the inventor is subject to the best mode requirement because it is a preferred mode of carrying out “the claimed invention,”<sup>33</sup> a greater or more detailed disclosure of the preferred mode does not entitle the inventor to broader claims. Instead, the other two doctrines rooted in section 112, ¶ 1, “enablement” and “written description,” control the inventor’s entitlement to a broader scope of protection.

Under the modern doctrine, enablement requires that one of ordinary skill in the art, relying on the disclosure and the information known to those of skill in the art, must be able to make and use the claimed invention without “undue experimentation.”<sup>34</sup> The requirement that the

---

31. *See, e.g.,* *Datamize, LLC v. Plumtree Software, Inc.*, 417 F.3d 1342, 1347 (Fed. Cir. 2005) (“Because the claims perform the fundamental function of delineating the scope of the invention, the purpose of the definiteness requirement is to ensure that the claims delineate the scope of the invention using language that adequately notifies the public of the patentee’s right to exclude.”) (citations omitted).

32. *Id.* (stating that test for definiteness is whether claim terms “can be given any reasonable meaning”).

33. *See Bayer AG v. Schein Pharms., Inc.*, 301 F.3d 1306, 1315 (Fed. Cir. 2002) (“[T]he best mode disclosure requirement only refers to the invention defined by the claims.”).

34. *See, e.g., Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384 (Fed. Cir. 1986) (“Enablement is a legal determination of whether a patent enables one skilled in the art to make and use the claimed invention, [and] is not precluded even if

specification's disclosure be commensurate with the scope of the patent claims was articulated early in the history of patent law,<sup>35</sup> and the existence and nature of the enablement doctrine are uncontroversial—though its application is by no means certain.<sup>36</sup>

In contrast, the nature of the separate written description requirement in the first paragraph of section 112 of the 1952 Act has been enormously controversial. The distinction between the enablement requirement and the written description requirement is, crudely speaking, the distinction between disclosing *how* to create the claimed subject matter, and disclosing *what* the claimed subject matter is.<sup>37</sup> A patent disclosure may provide sufficient technical information for one of ordinary skill in the art to make and use things within the boundaries defined by the claims, but fail to satisfy the written description requirement because it does not disclose the identity or characteristics of the subject matter within the claim.<sup>38</sup> The Court of Customs and Patent Appeals first articulated a separate doctrine of written description in *In re Ruschig*, invoking the doctrine to reject claims filed after the original patent application that were directed to an invention not disclosed in the original patent specification.<sup>39</sup> This priority-policing function of the written description is generally accepted, though it has been argued that section 132 of the Act, which

---

some experimentation is necessary, although the amount of experimentation needed must not be unduly extensive. . . .”) (citations omitted).

35. See *O'Reilly v. Morse*, 56 U.S. 62, 112-20 (1853) (holding invalid claim to all means of communication by electromagnetism when inventor had disclosed telegraph apparatus only).

36. See *infra* Part III (describing difficulties with enablement doctrine).

37. To illustrate, one may be able to teach an ordinarily skilled mariner how to sail from Spain to North America, but such instruction may not constitute a description of North America. One could also clearly define the boundaries of the United States but such delineation may not suffice as a description of the United States.

38. For example, suppose a patentee discloses a new method of chemical synthesis and a particular compound A that can be synthesized by the method. The new method might enable one of skill in the art to make other compounds B and C, but if the patentee has disclosed only A he may not have described the more general family of compounds embracing A, B, and C. See *In re DiLeone*, 436 F.2d 1404, 1405 (C.C.P.A. 1971). Likewise, one who describes a generic class may not have described individual species within that class. I might satisfactorily describe the United States but fail to describe a smaller portion of it, such as California.

39. See *In re Ruschig*, 379 F.2d 990, 995-96 (C.C.P.A. 1967). Without such a requirement, patent applicants could continue to claim inventions not disclosed in their applications while still relying on their original filing date to circumvent the requirements of novelty and nonobviousness.

prohibits claim amendments from introducing “new matter” into the disclosure, suffices to fulfill this function.<sup>40</sup>

What has been far less accepted is that the written description doctrine also applies to claims that were filed in the original patent application. This aspect of written description doctrine arose, at least in modern jurisprudence, in *Regents of the University of California v. Eli Lilly*.<sup>41</sup> In *Lilly*, the patentee had identified and disclosed the sequence of a DNA molecule encoding the insulin polypeptide from rats.<sup>42</sup> Despite not disclosing additional insulin-encoding DNA molecules, the patentee claimed DNA molecules encoding human insulin, as well as the broader genera of DNA molecules encoding mammalian or vertebrate insulins.<sup>43</sup> The accused infringer defended not on the grounds that the patent’s disclosure failed to enable any insulin-encoding DNA molecules other than rats’, but on the grounds that the patent did not provide a written description of human insulin DNA or the broader genera of mammalian or vertebrate insulin-encoding DNA molecules.<sup>44</sup> The Federal Circuit, affirming the district court, held that neither the claim to human insulin-encoding DNA, nor the claims to mammalian or vertebrate molecules, were valid under the written description requirement of section 112.<sup>45</sup>

In the time since *Lilly*, criticism of the written description requirement has been intense. A minority of judges on the Federal Circuit have vigorously denied the existence or utility of a separate written description requirement in section 112, arguing that enablement alone should define the scope of the patentee’s claims.<sup>46</sup> A majority of the Federal Circuit

---

40. See 35 U.S.C. § 132 (2000 & Supp. II 2002); *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 978 (Fed. Cir. 2002) (Rader, J., dissenting from the denial of rehearing en banc) (arguing that priority policing function of written description requirement is redundant with section 132).

41. *Regents of Univ. of Cal. v. Eli Lilly & Co.*, 119 F.3d 1559 (Fed. Cir. 1997).

42. *Id.* at 1562-63.

43. *Id.* at 1563. The claims were specifically directed to “cDNAs,” which are synthetic DNA molecules generated from reverse-transcription of protein-encoding RNA molecules.

44. *Id.* at 1563-64.

45. *Id.* at 1568-69.

46. To the extent the minority of judges concede that the need for a written description requirement to prevent patentees from adding unsupported claims to existing applications, they believe that such a function is more properly found in section 132 of the Patent Act, which forbids patentees from adding new matter to the specification by amendment. This Article refers to arguments for and against written description as arguments for the position that written description and enablement limit the scope of originally filed claims, or for the position that enablement alone limits the scope of

remains opposed to revisiting *Lilly*, though the issue has yielded no fewer than three spirited disagreements on the court's refusal to take the question en banc.<sup>47</sup> A wealth of scholarly commentary has sympathized with the minority position by arguing that the written description doctrine is a dangerous and unnecessary graft onto the traditional law of claim scope.<sup>48</sup> The arguments against an independent written description doctrine take both the doctrine and its application to task: the doctrine is obsolete under modern claiming practice;<sup>49</sup> it is unnecessary because original claims by statute constitute their own description;<sup>50</sup> there has been no coherent differentiation between the requirements imposed by enablement and the requirements imposed by written description;<sup>51</sup> it is incompatible with current claim interpretation methodology;<sup>52</sup> it deviates markedly from earlier Federal Circuit and Court of Customs and Patent Appeals precedent;<sup>53</sup> it is a biotechnology-specific doctrine;<sup>54</sup> and it has been

---

original claims. One's position on best mode is peripheral to this question. The inventor need only disclose *the* best mode of carrying out the claimed invention, if he has a preferred mode. Best mode is therefore not a doctrine of claim scope in the sense that enablement and written description are.

47. See *Lizardtech, Inc. v. Earth Res. Mapping, Inc.*, 433 F.3d 1373 (Fed. Cir. 2006) (denying rehearing en banc); *Univ. of Rochester v. G.D. Searle & Co.*, 375 F.3d 1303 (Fed. Cir. 2004) (denying rehearing en banc); *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956 (Fed. Cir. 2002) (denying rehearing en banc).

48. For review of the critical literature, see *Rochester*, 375 F.3d at 1309 (Rader, J., dissenting from denial of rehearing en banc) (citing thirty-one articles criticizing *Lilly*, seven articles defending it, and sixteen neutrally commenting upon it); Christopher M. Holman, *Is Lilly Written Description a Paper Tiger?*, 17 ALB. L. J. SCI. & TECH. 1, 17-25 (2007) (collecting criticisms).

49. See, e.g., Mark D. Janis, *On Courts Herding Cats: Contending with the "Written Description" Requirement (And Other Unruly Patent Disclosure Doctrines)*, 2 WASH. U. J.L. & POL'Y 55, 63 (2000).

50. See, e.g., *Enzo*, 323 F.3d at 988 (Linn, J., dissenting from denial of rehearing en banc) ("[T]he claims themselves—having been filed as part of the original application—provide their own written description.").

51. See, e.g., Holman, *supra* note 48, at 80 (arguing that courts have failed to articulate a standard for compliance with written description distinct from enablement).

52. See *LizardTech*, 433 F.3d at 1376-77 (Rader, J., dissenting from denial of rehearing en banc) (arguing that doctrine is incompatible with recent claim construction jurisprudence).

53. See, e.g., Janice M. Mueller, *The Evolving Application of the Written Description Requirement to Biotechnological Inventions*, 13 BERKELEY TECH. L. J. 615, 633-36 (1998).

54. See, e.g., *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 285 F.3d 1013, 1025 (Fed. Cir. 2002) (Dyk, J., dissenting) (arguing that *Lilly* applies a "unique" doctrine to biotechnology).

applied inconsistently between cases in different technologies<sup>55</sup> as well as between cases in the same technologies.<sup>56</sup>

Most, though not all, of the critics advocate abolishing the doctrine in favor of enablement as the unitary standard for patent disclosure.<sup>57</sup> Many of these criticisms stand on their own merits. However, if we place them within the context of the modern program of systematizing patent law into a conceptually ordered formal system, then it becomes clear that the elimination of the written description requirement is an essential element of that program. Obviously, any formal description of patent law requires at least one axiom of permissible claim scope. Why should the formalist conception of patent law entail that enablement be the only doctrine of claim scope? One answer is simply guilt by association. The judges of the Federal Circuit who have consistently opposed the *Lilly* doctrine also have authored some of the most notable articulations of the formalist program. Nonetheless, there is a deeper intellectual connection between the attack on written description and the desire for a classically ordered system of patent law. Classical legal theorists sought to condense scattered rules and doctrines around core principles such as “fault” or “will.”<sup>58</sup> Likewise, the modern classicists of patent law condense the disclosure doctrine around enablement alone, relegating both written description and best mode to peripheral roles.<sup>59</sup> Classical legal theorists sought unity in generalized categories such as “contract,” rather than separate bodies of law devoted to particular industries or relationships.<sup>60</sup> So too, modern critics who view the written description as a doctrine peculiar to chemistry and biotechnology cases assail the doctrine on the grounds that patent law ought not to be technology-specific. Classical legal theorists sought

---

55. See, e.g., Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L. J. 1155, 1183-85 (2002) (arguing that application of written description doctrine is inconsistent between industries).

56. See, e.g., *Univ. of Rochester v. G.D. Searle & Co.*, 375 F.3d 1303, 1308 (Fed. Cir. 2004) (Rader, J., dissenting from denial of rehearing *en banc*).

57. For example, Dan Burk and Mark Lemley seem to approve of the existence of the doctrine, but not its current application by the Federal Circuit. See *infra* note 235, at 1682-83.

58. See HORWITZ, *supra* note 7, at 13.

59. Judge Rader has argued that a separate best mode requirement is largely unnecessary. See *Bayer AG v. Schein Pharms., Inc.*, 301 F.3d 1306, 1325 (Fed. Cir. 2002) (Rader, J., concurring) (contending that best mode is “little more than . . . a trap for the unwary” and that it was wise for the Federal to have limited the requirement with “the statutory ‘scope of the claimed invention’ rule”).

60. See WILLIAM M. WIECEK, *THE LOST WORLD OF CLASSICAL LEGAL THOUGHT* 102-03 (1998).

certainty in the law in part to provide stability for business expectations;<sup>61</sup> modern theorists criticize the *Lilly* written description inquiry as one that yields no certain results in scope inquiries.<sup>62</sup>

However, these criticisms are secondary to the main intellectual thrust of the attack on written description. Hostility to the written description requirement derives fundamentally from the belief that written description is incompatible with the modern claim. The claim is the abstract legal creature at the heart of modern patent law, and the cornerstone of the conceptually ordered system pursued by formalist thinkers. For those who seek a conceptually ordered system of patent law, written description plays no role in a rational modern system because it is an obsolete relic of an earlier patent law: a patent law without claims. When the earliest United States patent statutes demanded that the patent applicant “deliver a written description of his invention,”<sup>63</sup> patents did not include claims defining the scope of inventors’ rights. Early judicial interpretations of the patent statute therefore required that the specification not only enable practice of the invention so as to satisfy the quid pro quo of the patent system, but also define the invention so as to put the public on notice of infringement, and to permit courts to assess the novelty of the invention.<sup>64</sup> The description of the invention in the specification therefore served the function of modern claims: to define the extent of the patentee’s exclusive rights. Claims evolved gradually over the course of the 19th Century, first as formal statements of the invention’s novelty rather than definitions of the scope of the inventor’s rights.<sup>65</sup> Later, the claim evolved to represent the subject matter against which the patentee could claim infringement—initially “central” claims, which defined an embodiment around which judges determined the actual scope of patent rights, then later as modern “peripheral” claims, which themselves define the boundaries of the patent right.

In modern practice the claim is “the invention.” The set of properties recited by the claim define both the subject matter over which the patentee

---

61. See Robert W. Gordon, *Legal Thought and Legal Practice in the Age of American Enterprise, 1870-1920*, in PROFESSIONS AND PROFESSIONAL IDEOLOGIES IN AMERICA 92 (Gerald L. Geison ed., 1983). Modern historians have distanced themselves from the position that architects of classical legal thought intended it to fortify the emerging capitalist class. See *id.*

62. See, e.g., Timothy R. Holbrook, *Possession in Patent Law*, 59 SMU L. REV. 123, 161-62 (2006).

63. Patent Act of 1793, ch. 11, § 3, 1 Stat. 318, 321 (repealed 1836).

64. See *Evans v. Eaton*, 20 U.S. 356, 380-81, 391-92 (1822).

65. See *supra* note 9.

may assert infringement and the subject matter which may invalidate the patent if known or obvious from the prior art. For those who doubt a modern role for the written description doctrine, claims have entirely supplanted the notice function once performed by the specification's description of the invention.<sup>66</sup> On this view, written description as articulated by *Lilly* is an atavism of the time before claims and should have no role in limiting the permissible scope of modern peripheral claims.<sup>67</sup> The function of the description in the specification is to fulfill the quid pro quo of the patent system: to disclose the invention to the public, so that the public may practice the invention upon expiration of the inventor's exclusive right. That is the function of the enablement doctrine alone. To complete the process of systematizing patent law into a conceptually ordered system grounded upon the peripheral claim, it is necessary to discard remnants of the pre-claim system such as the written description doctrine. Thus, for those who would deny the doctrine of written description, the conceptual ordering of patent law can be largely achieved by adding only one more formal axiom to the set described above:

The maximum permissible boundaries of the claim are what the patentee has enabled in the specification.

In this axiom, "enabled" means what one in the ordinary skill in the art could make and use without undue experimentation.

The object of this Article is to evaluate the formalist conception of patent law and its implications for patent law's substantive doctrines. Identifying the modern patent program as "formalist" or "classical," does not suggest that it represents an obsolete or futile endeavor. The decay of the classical system of legal thought may have come more from its political assumptions than from its epistemological ones, and it is beyond the scope of this Article to evaluate whether the epistemological criticisms leveled at classical legal thought negate the possibility of a conceptually ordered patent system as well. Nor does this Article address whether a conceptually ordered system serves the ends of patent law; both the 19th

---

66. See *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 977 (Fed. Cir. 2002) (Rader, J., dissenting from the denial of rehearing en banc) ("In later enactments, this function was assigned to claims, leaving enablement as the only purpose of the 'written description' language.").

67. Again, because the Court of Customs and Patent Appeals grounded the "new matter" rejection of new claims in a continuation application in part in section 112, those who feel bound to respect the CCPA's precedent concede that the "written description" language of section 112 serves that important role. See *supra* note 46.

Century and the modern formalist agendas have become too central to patent law to realistically recommend their abolition. Rather, the focus of this Article is on whether the formalist program can succeed on its own terms in formulating a conceptually ordered patent system and whether the written description doctrine is necessary to such a system. This Article asks whether patent law can be reduced to a set of uniform principles centered on a formal and hierarchical model of claim scope. In particular, this Article assesses whether the enablement requirement can satisfactorily limit claim scope in our modern peripheral claiming system or whether additional constraints—such as the written description requirement—are necessary to limit claim scope based on the inventor's disclosure. This Article's approach is in part formal, as it gauges whether a theoretically coherent doctrine of claim scope is possible within the confines of the peripheral claiming system. However, because the question of permissible claim scope is so foundational to patent law, many of the theoretical difficulties this Article highlights will entail balancing the incentives and costs the patent system allocates between patentees, competitors, later inventors, and the public at large.

Part II will consider how the various doctrines of patent law may be expressed as formal statements of inclusion or exclusion from claim scope. It will show that enablement, in contrast to other doctrines, cannot be expressed as a truth function defined in terms of claim scope. Part III will detail the characteristics of enablement that make it incapable, in its present form, of adequately constraining claim scope, and evaluate proposals for modifying enablement to account for these problems. Part IV will revisit the doctrine of written description and argue for its hitherto unrecognized role as a doctrine of *definition*. Once the written description doctrine is recognized as a doctrine of definition, it will become clear that written description is necessary, though undoubtedly not sufficient, to define the scope of patentable inventions in a peripheral claiming system. Finally, Part V will consider the implications of treating the written description doctrine as a doctrine of definition.

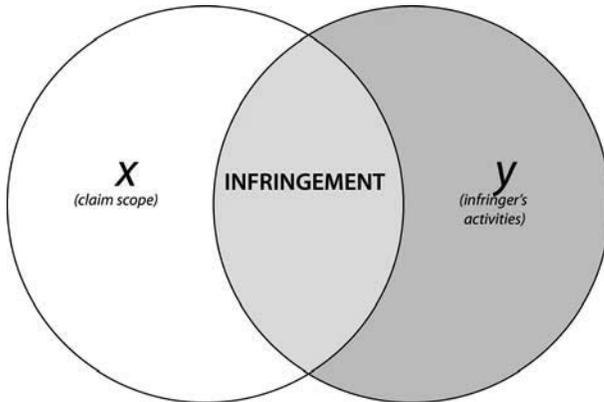
## II. PATENT LAW AS A FORMAL SYSTEM

Nearly all the doctrines of patent law can be described as a precise relationship between the legal inquiry and the subject matter within the claim's boundaries. These doctrines may be posed almost as mathematical set-functions whose truth value is described in terms of the claimed subject matter. Take a claim reciting particular properties, and call the set of all possible things or events characterized by those properties as  $x$ . In general, a patent is infringed by the manufacture, use or sale, of anything

possessing all the properties recited by the patent claims. We may easily represent the question of patent infringement in terms of the members of  $x$ :

Let  $y$  be the set of all things the accused infringer has made, used, sold, or offered to sell within the United States. The claim is infringed if and only if  $x$  and  $y$  intersect.

We might diagram this rendition of the infringement inquiry as follows:



The basic test for patent infringement, therefore, may be reduced to a test of intersection between the accused subject matter and the set of things encompassed by the claim. In practice, this test may be uncertain, for we must construe the language of the claim to define the set of things it encompasses. Claim construction can be a difficult and unpredictable exercise. But the uncertainty associated with claim construction is inherent to nearly all inquiries in patent law, because nearly all patent law doctrines depend on the claim. Moreover, the difficulties of linguistic interpretation are hardly peculiar to patent law. If we concede the baseline uncertainty in interpreting the words of the claim, then at least in theory it is straightforward to define the set  $x$ , the set of things possessing the properties recited by the claim. Once that set is defined, then the infringement inquiry asks merely whether the accused subject matter falls within that set.<sup>68</sup>

Can the statutory requirements of patent validity be expressed in similar terms? Consider first the novelty provisions of section 102, which generally deny patentability if the invention was known or used by others

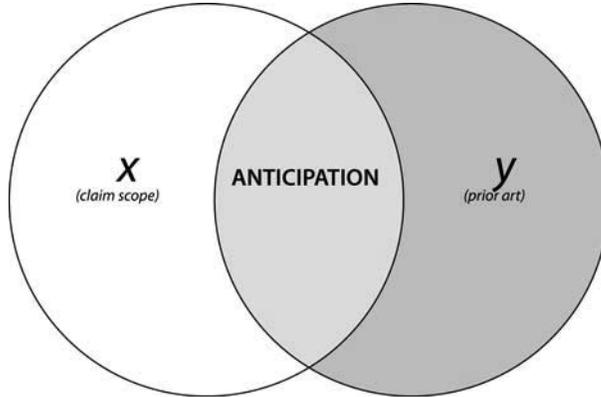
---

68. Of course, there may be factual disputes over whether the accused subject matter actually has the properties recited by the claim, but few areas of law are free from that kind of factual uncertainty.

before the date of invention.<sup>69</sup> If the invention is synonymous with the claims, then we can express the requirement of novelty as a simple intersection between the subject matter of the claim and the prior art:

Let  $y$  be the set of all things known and used, or patented or described in a printed publication, prior to the date of invention (the prior art). The claim is novel if and only if there is no intersection between  $x$  and  $y$ .

Thus, the claim is anticipated if there is any overlap between the claim scope and the prior art:



The law governing inventorship, also expressed in section 102, functions similarly: one who contributes to the conception of any element of a claim becomes an inventor not only of the entire claim, but the entire patent, as well.<sup>70</sup> Likewise, in priority contests between two inventors, reduction to practice of a single embodiment within the scope of the claim generally suffices to establish priority.<sup>71</sup>

---

69. The statutory bar provisions of section 102 deny patentability to inventions if patent applications are not filed within one year of disclosure or commercialization of the inventor; these provisions function like novelty with respect to claim scope. See 35 U.S.C. § 102(b) (2000 & Supp. II 2002).

70. See *Ethicon, Inc. v. U.S. Surgical Corp.*, 135 F.3d 1456 at 1460-64 (Fed. Cir. 1998) (granting inventorship to worker who contributed to conception of claimed features or embodiment falling within means-plus-function claim).

71. Notwithstanding that an inventor establishes priority over a rival by being the first to reduce to practice a single species of the generic invention, the inventor may still be denied a patent if the single species fails to satisfy the disclosure requirements of section 112 with respect to the generic invention. See *In re Zletz*, 893 F.2d 319, 32-23 (Fed. Cir. 1989); *Fried v. Murray*, 268 F.2d 223, 225 (C.C.P.A. 1959) (stating that lack of support for full scope of interference count could be a question only of patentability, not priority). The law with respect to conception is not so clear. See *In re Jolley*, 308 F.3d

The test of novelty may not be entirely certain, because the scope of the claim—the set of things  $x$  having the properties recited by the claim—may be uncertain. In practice, claims may be precise or vague, and the more vague the claim the more uncertain the test of novelty. But practically every inquiry in patent law shares this uncertainty. This Article’s aim is not to show that particular doctrines in patent law are certain or uncertain in practice. It is to illustrate where uncertainty lies in the various doctrines of patent law, and, more importantly, to distinguish between the fundamental *kinds* of uncertainty inherent in the doctrines. Uncertainty in claim scope means that the set  $x$  may be difficult to define. However, once we have defined  $x$  to whatever degree we think satisfactory or practical, the underlying formal structure of the novelty inquiry is precise. Likewise, the set of prior art  $y$  may be uncertain because the standard of whether a thing is “known or used” is not precise.<sup>72</sup> However, once a satisfactory definition of set  $y$  is achieved, the expression of the novelty inquiry in terms of sets  $x$  and  $y$  is clear.

The requirement of section 103 that the claimed subject matter be nonobvious may also be framed as a relationship between the set of prior art and the set of things encompassed by the claim:<sup>73</sup>

Let  $y$  again be the set of all things known and used prior to the date of the invention. The claim is obvious if, for any  $y$  or set of

---

1317, 1322 n.2 (Fed. Cir. 2002) (noting that conception of species may, but not necessarily, constitute conception of a genus).

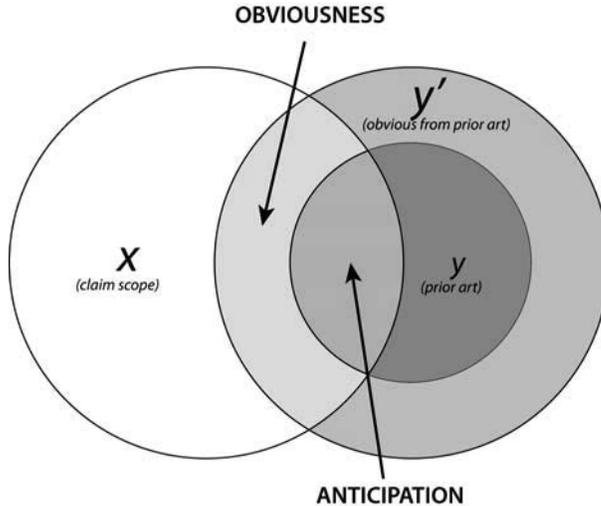
72. For example, the question of whether subject matter was “known or used” if the property defining the subject matter was not perceived at the time has divided the Federal Circuit. *See* *Elan Pharms., Inc. v. Mayo Found. for Med. Educ. & Research*, 304 F.3d 1221, 1229-351 (Fed. Cir. 2002) (holding that inherent feature of transgenic mouse was not disclosed by reference suggesting method of making mouse); *id.* at 1241-45 (Dyk, J., dissenting) (arguing that reference inherently disclosed feature that would be present if method performed); *see also* *Elan Pharms., Inc. v. Mayo Found. for Med. Educ. & Research*, 346 F.3d 1051 (Fed. Cir. 2003) (deciding case on enablement rather than inherency grounds); *Elan Pharms., Inc. v. Mayo Found. for Med. Educ. & Research*, 314 F.3d 1299 (Fed. Cir. 2002) (order granting rehearing en banc).

73. The Supreme Court made this point in its *KSR* decision, explicitly defining the nonobviousness inquiry in terms of the set of things encompassed by the claim. *See* *KSR Int’l Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1742 (2007) (“What matters is the objective reach of the claim. If the claim extends to what is obvious, it is invalid under § 103.”). The Court many years earlier in *Graham* made definition of the set of prior art the first factual inquiry under section 103. *See* *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1, 17 (1966) (“Under § 103, the scope and content of the prior art are to be determined. . .”).

$y$ , the difference between any  $y$  and any  $x$  would be obvious to one of ordinary skill in the art.

or:

Let  $y'$  be the set of all things for which the difference between any  $y'$  and any  $y$  or set of  $y$  would be obvious to one of ordinary skill in the art. The claim is obvious if and only if  $x$  and  $y'$  intersect.



Again, the reader may object that this is not a precise relationship at all. The transformation between  $y$ , the prior art, and  $y'$ , the set of all things obvious in light of the prior art, is vague and indeterminate. True, there has never been a certain test of whether the difference between a  $y$  and a  $y'$  would be obvious to one of ordinary skill in the art, and the test has become even less certain after the Supreme Court's opinion in *KSR Int'l Co. v. Teleflex Inc.*<sup>74</sup> This kind of uncertainty, however, has nothing to do with the scope of the claim in question. The uncertainty lies in how far the penumbra of obvious objects  $y'$  extends from the boundaries of the prior art objects  $y$ . Once we develop or posit a determinate method of defining the extent of this penumbra, then the structure of the nonobviousness inquiry is identical to the novelty inquiry: we simply ask whether claim

74. See *KSR Int'l Co.*, 127 S. Ct. at 1727. Before *KSR*, the Federal Circuit required as an element of obviousness a teaching, suggestion, or motivation for the artisan to combine or to modify the prior art and arrive at the claimed subject matter. See *id.* at 1734. Under this so-called "TSM test," the set of obvious subject matter  $y'$  might have been defined more precisely: something is a member of  $y'$  if there exists a teaching, suggestion, or motivation connecting the putative member with a member of  $y$ . See also Holbrook, *supra* note 62, at 172.

scope  $x$  and penumbra  $y'$  intersect, rather than whether the claim scope  $x$  and prior art  $y$  intersect.<sup>75</sup>

Similarly, for the doctrine of utility—the requirement of section 101 that the invention be “useful”—we may define a straightforward relationship between the validity inquiry and the scope of the claim.<sup>76</sup> As a general matter, if an applicant or patentee establishes the utility of a species encompassed by the claim, then the utility of the claim is established.<sup>77</sup> Therefore, if we can agree on a satisfactory standard of whether a particular embodiment is useful or not, whether a claim satisfies the utility requirement generally reduces to whether any member of the set  $x$  possesses the quality of utility.<sup>78</sup>

---

75. Professor Durham’s treatment has also suggested the restatement of obviousness as intersection. See Alan L. Durham, *Patent Symmetry*, 87 B. U. L. REV. 969, 995, 997-98 (2007). Durham’s discussion of the difficulty in assessing obviousness with respect to “the claim,” versus assessing whether some *subject matter* within the claim is obvious, in some way mirrors the difficulty identified herein with enablement doctrine. See *id.* at 995-96. This is not an issue for obviousness doctrine, but arises in connection with Durham’s suggestion to recast infringement by equivalents in terms of obviousness.

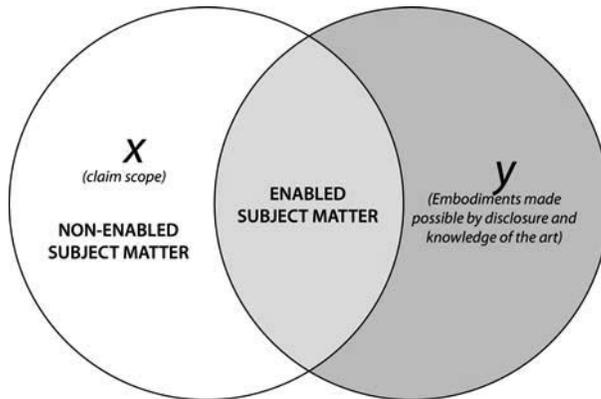
76. As far as the requirement that the invention be within the class of statutory subject matter defined by section 101, the Court of Customs and Patent Appeals assumed, without deciding, that a claim reading on both statutory and nonstatutory subject matter would be invalid under section 101. See *In re Mahony*, 421 F.3d 742, 745-46 (C.C.P.A. 1970). This interpretation may be untenable. Any open claim may be construed to include some form of nonstatutory subject matter, because adding additional elements to subject matter meeting the limitations of the claim does not remove that subject matter from the scope of the claim. The Federal Circuit had the opportunity to resolve this question in *In re Bilski*, as one of the questions posed for en banc review was whether claims that contain both mental and physical steps are eligible subject matter under section 101. See *Bilski*, 264 F. App’x. 897 (Fed. Cir. 2008) (order granting en banc review). However, the court did not directly address this question in its en banc decision. See *In re Bilski*, 545 F.3d 943 (Fed. Cir. 2008).

77. See UNITED STATES PATENT AND TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE § 2107.02. (8th ed. rev. 6, 2007) (“Where an applicant has established utility for a species that falls within an identified genus of compounds, and presents a generic claim covering the genus, as a general matter, that claim should be treated as being sufficient under 35 U.S.C. § 101.”).

78. The question becomes more complicated if some members of  $x$  have utility, and some do not. Unlike novelty and nonobviousness—for which the claim is invalid if any species encompassed by the claim lacks those qualities—the utility requirement of section 101 still may be met even if some members of the set  $x$  are not useful. At least in recent case law, this question of “inoperative embodiments” has been treated not as a matter of utility *per se*, but as a matter of compliance with the enablement requirement of section 112: so long as one of ordinary skill in the art can distinguish between the operative and inoperative embodiments without “undue experimentation,” then one of skill in the art can “make and use” the invention in accordance with the first paragraph of

Unlike sections 101, 102, and 103, the disclosure requirements of section 112 *cannot* be reduced to an inquiry defined strictly in terms of the claim scope. Consider the enablement aspect of section 112, which according to the reductionist position, is the only disclosure doctrine necessary to define the proper scope of allowable claims. Section 112 frames the enablement inquiry as whether one of ordinary skill in the art can “make and use” the invention, a standard which the Federal Circuit has explained requires that the ordinary artisan be able to make and use the invention without “undue experimentation.” Let us define as  $y$  the set of all things which the skilled artisan, equipped with the teachings of the patent and the knowledge of the art, could make and use without undue experimentation. We cannot express enablement as a simple intersection as we could for sections 101, 102, and 103. Simply because an inventor has enabled *something* within the scope of the claims, he is not necessarily entitled to *everything* within the scope of the claims. The proposition:

The claim is enabled if  $x$  and  $y$  intersect

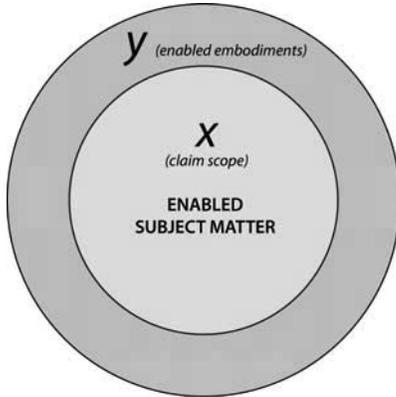


section 112. *See Atlas Powder Co. v. E.I. du Pont de Nemours & Co.*, 750 F.2d 1569, 1576-77 (Fed. Cir. 1984) (holding that claim encompassing inoperative embodiments may be enabled if one of ordinary skill can distinguish inoperative embodiments without undue experimentation). We might express this doctrine as follows:

*Let  $x'$  be the members of  $x$  which are operative, and let  $x''$  be the members of  $x$  which are not operative. The claim is valid if and only if one of ordinary skill in the art can identify members of  $x'$  without undue experimentation.*

is false, because enablement of some members of  $x$  does not necessarily imply that the full claim scope is enabled. An inventor may very well disclose how to make and use some things within the scope of the claim, but still fail to enable the full scope of the claim.<sup>79</sup> Nor must the inventor enable *all* things falling within the scope of the patent claim. The proposition

The claim is enabled only if all members of  $x$  are also members of  $y$



is false, because an inventor need not—and in most instances can not—enable all things falling within the scope of his claim.<sup>80</sup>

The point to emphasize is that the uncertainty inherent in the disclosure requirement of section 112 is qualitatively different from the uncertainties inherent in other doctrines of patent law. Once the predicate facts have been established,<sup>81</sup> questions of infringement and novelty are completely determinate provided that the scope of the claims is precise. Nonobviousness is also determinate if, in addition to precise claim scope, we can determine whether a specific thing  $y'$  falling within the scope of the claims would or would not have been obvious to one of ordinary skill in the art. Not so for enablement. Even if we have a perfect technique for construing claims, and a perfect test of whether one of ordinary skill in the art could make and use a particular thing  $y$ , we cannot necessarily determine whether the claim meets the enablement requirement of section 112. Validity has no certain relationship to claim scope even if the

79. See *infra* Section III.A.2.

80. See *infra* Section III.A.1.

81. That is, whether the accused subject matter or the prior art actually has the properties recited by the claim.

theory of claim construction or the underlying substantive doctrine of enablement is further refined.

Disclosure is therefore different from infringement, anticipation, or nonobviousness. Of what significance are these differences? The analysis of patent law as a formal system suggests two things about the disclosure doctrines necessary for a coherent system of patent law. First, if we aspire to a conceptually ordered system, the disclosure doctrines are going to be more troublesome than other aspects of patent law. Consequently, we might a priori be less confident that we can condense the problem of disclosure around a single unifying principle. Second, this analysis tells us something about the nature of the doctrines necessary to complete a coherent system of patent law. Doctrines like anticipation and enablement depend on the application of a legal standard to particular embodiments within the claim rather than the claim itself: anticipation asks whether a particular thing was known or used in the prior art, and enablement asks whether one of ordinary skill in the art could make and use a particular thing without undue experimentation. Because the problem of disclosure cannot be reduced to a simple inquiry in terms of embodiments falling within the claim, our disclosure doctrines may need to be founded in part upon a different intellectual framework—one that tests the scope of the claim itself against a legal standard, rather than one that tests a particular embodiment or collection of embodiments against a legal standard. This Article contends not only that the written description doctrine can fulfill this role, but also that the doctrine is absolutely necessary to complete the set of constraints on a coherent system of patent law. The next section therefore explores in depth why the enablement doctrine alone is insufficient to provide a coherent doctrine of permissible claim scope.

### III. THE LIMITS OF ENABLEMENT

The inability to formulate a doctrine of enablement as a simple function of exclusion or inclusion is no mere quirk in the law of enablement. Rather, it is an inherent and necessary property of the particular claim system that has been at the heart of the United States patent system for a century, and more recently, abroad. It arises because we have consolidated three formerly separate concepts in patent law—the invention, the claim, and the scope of the inventor's exclusive rights—into a unitary conception founded upon the peripheral claim. The nature of the peripheral claim itself is what makes patent law not reducible to a simple set-theoretic system. In turn, the inability to reduce patent law to a simple set-theoretic system means that the attempt to squeeze out the last remnant

of the pre-claim conception of patent law—the written description doctrine—cannot succeed if patent law is to be a coherent system.

### A. The Problem of Infinite Scope

#### 1. *All the World's a Genus: All Claims are Infinite*

Why is it that enablement alone cannot provide adequate limits of claim scope within the structure of modern patent law? It cannot because modern patent law must reconcile two apparently contradictory principles:

The claim completely and exclusively<sup>82</sup> defines the class of things over which the inventor may exercise his rights.

All patent claims are of infinite scope.

The first principle is familiar; the second may be less so. It is an essential characteristic of all patent claims that they cover a set of entities rather than a single entity. Otherwise claims could not be infringed, save perhaps by the use of the one physical entity that the inventor constructed. Yet, the set of entities covered by a claim, despite being bounded by the language of the claim and the various doctrines of patent law, must be infinite in scope. This conclusion follows not from legal doctrine, but from the ontological nature of patent claims themselves.

The distinction between “genus” claims—claims covering a class of entities characterized by a common property—and “species” claims—claims covering only a single entity—is familiar in chemical and biotechnological practice. An inventor might synthesize a novel molecule with antibiotic efficacy, and file a claim defining the specific structure of the molecule that she synthesized. However, because molecules with minor modifications to the chemical backbone may share the antibiotic efficacy, inventors typically also draft a claim to a genus of related molecules sharing the same backbone but varying in the atoms or groups attached to the backbone. Likewise, because multiple DNA molecules can encode the same polypeptide,<sup>83</sup> an inventor discovering a novel protein will typically claim the genus of all DNA molecules encoding that protein. Chemical patent law has long recognized the problems of adequate

---

82. The doctrine of equivalents of course renders this statement not literally true if we take “define” to mean only the literal claim scope. The argument is the same whether we consider only the literal claim scope or literal claim scope plus equivalents; if the point is that all claims are infinite, then the extension of those claims by the doctrine of equivalents is a relatively trivial matter.

83. This is due to the degeneracy of the genetic code.

disclosure that arise from a patentee's attempt to claim a genus that encompasses a variety of different but related molecules.<sup>84</sup>

Less evident is that essentially all patent claims—not just those defining chemical and biotechnological inventions—are genus claims. Modern patent claims define the scope of the inventor's rights by reciting *properties*; all things having those properties fall within the scope of a patent's claims.<sup>85</sup> A claim to a molecule recites the atomic structure of the molecule, or the physical properties that characterize a composition of the molecule. A claim to an apparatus recites the structure of the components of the apparatus, or functional language describing its operation. A claim to a method or process recites the steps of the method, or the physical conditions under which the process take place. Regardless of form, however, most patent claims define an infinite number of existing and possible objects or acts.<sup>86</sup> Consider a simple claim to a chair having four legs:

1. An object for supporting a human body, comprising a substantially flat surface sized to accommodate a human posterior, and four legs supporting said surface.

This claim is unremarkable and, supposing the inventor to be the first to conceive of the idea of a chair with four legs, we would not think this claim poses any issue of adequate disclosure. Yet this claim, even more so than the typical chemistry or biotechnology claim, covers an infinite variety of embodiments. Like nearly all patent claims, this claim is written

---

84. See, e.g., RIDSDALE ELLIS, PATENT CLAIMS §§ 214-42 (1949) (discussing problems of class definitions in chemical cases).

85. In metaphysical terms, the patent claim is thereby synonymous with the extension of the properties, or class: "A *class* is often thought of as the extension of a property (or concept), the collection of all those things . . . which have that property or fall under that concept." A COMPANION TO METAPHYSICS 86 (Jaegwon Kim & Ernest Sosa eds., 1996).

86. It might be possible in theory to draft claims limited to a particular instantiation of those properties. One might, for example, claim a chair with the property that "said chair being resident in Room 380 of 200 McAllister Street, San Francisco, on November 21, 2007." Obviously such claims have little commercial value, and are thus lacking in practice. This particular example raises the question of whether subject matter defined by temporal or spatial limitations would infringe if, having met those conditions at some point, ceased to meet them at a later point. The example is trivial but the general question is not. See *Novartis Corp. v. Ben Venue Labs., Inc.*, 271 F.3d 1043, 1049 n.5 (Fed. Cir. 2001) (noting theory of transitory infringement by chemical intermediate).

in the so-called “open” format, employing the word “comprising.”<sup>87</sup> Such claims are construed to cover all things that possess the recited properties. Subject matter with *additional* properties or elements still falls within the scope of the claim, so long as it retains those properties recited by the claim. Thus chairs made of all sorts of materials, chairs of all sizes, chairs including contoured backrests, and chairs with roller wheels, etc. are all within the claim so long as they possess the recited flat surface and four legs.<sup>88</sup> In reality, then there is no such thing as a “species” claim, for claims are never restricted to a single physical entity. Insofar as both genus and species are abstractions, the difference between the two is less in kind and more in degree.

Supposing the inventor to have disclosed the basic structure of the chair, we would have little difficulty concluding that claim 1 satisfies the enablement requirement of section 112. The inventor is entitled to assert exclusive rights over all chairs which include a flat surface and four legs. If the inventor has enabled those of skill in the art to make and use the genus of chairs defined by claim 1, then by definition, claims dependent on claim 1—claims reciting additional properties and thereby defining subsets of claim 1—are also enabled.<sup>89</sup> Nonetheless the inventor has not disclosed information sufficient to make and use all subsets of claim 1. Consider the claims:

2. The object of claim 1, wherein the legs and surface are composed of neutronium.<sup>90</sup>
3. The object of claim 1, wherein the object further comprises a portable fusion reactor.

---

87. There are more narrow patent claims drafted with the phrase “consisting of” instead of “comprising”; such “closed” claims extend only to subject matter possessing the recited elements and no others. Closed claims are quite rare and are generally employed only when the invention lies in the elimination of an element or step necessary in the prior art. Somewhat more common are “hybrid” claims employing the language “consisting essentially of,” which are open to the addition of elements that do not materially change the properties of the claimed subject matter.

88. Indeed, chairs with five or more legs (but not three) would also fall within the scope of the claim, because they have the recited four legs in addition to their others. *See Gillette Co. v. Energizer Holdings*, 405 F.3d 1367, 1374 (Fed. Cir. 2005) (holding that four-bladed razor infringed claims reciting “a razor comprising . . . a group of first, second, and third blades.”).

89. *See infra* note 93.

90. A material of unimaginable density found only in neutron stars, where gravitation has forced protons and electrons to combine.

Claim 2 defines a set of chairs composed of a material that cannot now (nor possibly ever) be made on this planet.<sup>91</sup> Claim 3 defines a set of chairs including a portable fusion power source, a technology that might be possible in the future but certainly is not available today. Clearly, the inventor's disclosure did not enable one of ordinary skill in the art to make chairs of neutronium or including fusion reactors. Yet, because of the hierarchical structure of patent claims, the sets of chairs defined by claims 2 and 3 are subsets of the set of chairs defined by claim 1.

These claims are, by statute, proper dependent claims because they incorporate the limitations of and add further limitations to define narrower subcategories of claim 1.<sup>92</sup> According to the hierarchical model of claim scope, if the inventor has satisfied the enablement requirement of section 112 with respect to claim 1, then he has done so for claims 2 and 3 as well. As the Board of Patent Appeals and Interferences reasoned in *Ex parte Forstova*:

We first express our concern about the anomalous situation confronting us where *dependent* claims 2-5 are rejected as being non-enabled while claim 1, the independent claim from which these claims directly or indirectly depend, is not rejected. It has long been held that a claim must be enabled throughout its scope. As a matter of logic, assuming claims 2-5 are proper dependent claims and we see no reason why they are not, the examiner's decision that claims 2-5 are non-enabled *necessarily means* that claim 1 is non-enabled.<sup>93</sup>

Thus, as a matter of enablement law, if one of ordinary skill in the art could make and use the set of chairs defined by claim 1, he can also make and use the smaller sets of chairs defined by claims 2 and 3. Even if we accept the notion that the inventor need not enable all embodiments within

---

91. Claim 2 presumably would be a proper dependent claim even if claim 1 was written in closed format. Claim 3 would not, because the addition of a fusion reactor would be an additional element. Note that additional elements narrow claims. If we are truly committed to the hierarchical claim structure, it is not entirely clear that the distinction between open and closed claims can be sustained. Consider a closed claim defining a chair "consisting of a seat and four legs." From a purely ontological viewpoint, there is no distinction between narrowing the set by adding the property "composed of wood," and narrowing the set by adding the property "having a backrest." Yet the chair composed of wood would infringe the closed claim, and the chair including a backrest would not.

92. See 35 U.S.C. § 112, ¶ 4 (2000).

93. *Ex parte Forstova*, No. 1998-0667, 2002 WL 3234992 3 (B.P.A.I. Apr. 11, 2002).

the scope of a claim for the claim to be enabled, claims 2 and 3 are curious. It seems that either the inventor is entitled to claims 2 and 3, or there must be some limitation on permissible claim scope beyond the enablement doctrine as currently conceived.<sup>94</sup>

This paradox may be more significant than is first supposed. Today, the chair claim clearly lacks novelty over known chairs. However, claims 2 and 3 are certainly novel and nonobvious because no prior art discloses or makes obvious the limitations added by claims 2 and 3; claims 2 and 3 may therefore be patentable where claim 1 is not. A real-world manifestation of this pattern appeared in *Amgen Inc. v. Hoechst Marion Roussel, Inc.*<sup>95</sup> Several of the patents in *Amgen* claimed a “non-naturally occurring erythropoietin [EPO] glycoprotein.” Because naturally occurring EPO was known in the prior art, addition of the “non-naturally occurring” limitation made the claims novel and potentially nonobvious over the prior art. The patentee’s disclosure of one method of making non-naturally occurring EPO was held to enable the claim. This claim was construed to cover *all* non-naturally occurring EPO, whether made by the patentee’s synthetic process or not. By adding a novelty-imparting limitation to a broad genus, the patentee was able to lay claim to all subsequent synthetic EPO molecules without having to enable the sub-genera of molecules made by different synthetic processes. Although the broader claim is more useful commercially, as with claims 2 and 3, it is difficult to understand in terms of enablement alone why *Amgen* could not have explicitly claimed synthetic methods of producing EPO that were not yet possible when it filed its application, given that the broader category of synthetic EPO was enabled.<sup>96</sup>

Likewise, in *Forstova*, the Board of Patent Appeals and Interferences reversed enablement rejections of claims on technology primarily useful

---

94. One might object that claims 2 and 3 lack utility as demanded by 35 U.S.C. § 101. If utility means the requirement that inventions confer some tangible benefit upon society, then the lack of utility can easily be remedied by changing the hypothetical to less outlandish objects that might become more useful by being composed of neutronium or including fusion reactors; such objects, if possible to create, would confer benefit upon society. Likewise, if concern exists about the prohibition against inventions that violate known laws of physics, examples can be chosen that are beyond current technology yet are more plausible than the ones given.

95. 314 F.3d 1313 (Fed. Cir. 2003).

96. Claims defining subject matter very far afield from the embodiments the inventor created may allow the inventor to circumvent certain limitations on the licensing of patent claims. See Robin Jacob, *Objectionable Narrowness of Claim*, in *PRINCIPLES OF PATENT LAW* 1097 (Donald S. Chisum et al. eds., 2d ed. 2001).

for gene therapy, notwithstanding that gene therapy is not yet clinically viable. In *Forstova*, dependent claims to a method of transferring DNA into a host cell with a “therapeutic effect” on an organism were allowed, despite the examiner’s rejection that applications of the method to clinical gene therapy in humans were not enabled. Because the claims were not directed to clinical gene therapy per se, the Board held that alleged difficulties in clinical gene therapy did not preclude enablement of the claim.<sup>97</sup> However, given the Board’s reasoning that enablement of a narrower claim is logically predicated on enablement of the broader claim, a dependent claim explicitly directed to clinical gene therapy ought to have been enabled as well. If so, the attachment of “non-enabled limitations” to broader enabled claims provides a means to circumvent the rule that an inventor cannot claim an improvement or additional feature on a base technology if the base technology itself is not enabled.<sup>98</sup> If the non-enabled base technology is attached as a limitation to a broader enabled claim, then the problem of enablement is circumvented.

The problem is not simply one of future technology, for the claim is infinite regardless of *when* its scope is assessed. As in *Amgen*, a concern exists with the problem of claim scope in the context of after-arising technology: a later inventor develops a marvelous new back-supporting chair, and we question whether the original inventor ought to be entitled to assert patent rights over chairs that did not exist or could not exist at the time the inventor filed for a patent. The inventor’s entitlement to future developments is important in allocating the proper incentives for innovation between earlier and later inventors.<sup>99</sup> However, the scope questions that arise in the context of after-arising technology are merely subsets of the more general problem of infinite claim scope. There are an infinite number of variations on the simple chair that can be constructed with contemporary technology, such as variations in material, proportions, and decoration, and are within the scope of the claim. It seems self-evident that without a coherent conception of claim scope with respect to *present-day* embodiments of the invention, we cannot hope to achieve a coherent

---

97. See *Forstova*, 2002 WL 32349992, at 4-5.

98. See *Gould v. Hellwarth*, 472 F.2d 1383, 1386 (C.C.P.A. 1973) (holding that improvement on laser could not be patented absent disclosure enabling construction of laser).

99. See, e.g., Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 295-298 (2003) (discussing problems raised by broad upstream patents in biomedical field); Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 885-91 (1990) (criticizing broad upstream rights).

conception of claim scope as applied to *future* embodiments of the invention.<sup>100</sup>

2. *What is the “Full Scope” of Infinite Scope?*

Notwithstanding the issues raised by cases like *Amgen* or *Forstova*, the puzzle posed by claims 2 and 3 might be considered as merely a quirk in current enablement doctrine. Abandoning, or at least modifying, the hierarchical model of claim scope with respect to enablement could eliminate the paradox. By discarding the Aristotelian requirement that all characteristics possessed by the genus must also inhere in the sub-genus, we could permit an enabled independent claim to include non-enabled dependent claims. After all, “enablement” is a legal property of the claim, not a physical property of the entities encompassed within the claim,<sup>101</sup>

---

100. Professor Merges has argued that we do not ask whether the inventor has enabled the subject matter recited by claim “generally,” but merely whether the inventor has enabled the embodiments known to be within the claim as of the filing date. See Robert P. Merges, *Rent Control in the Patent District: Observations on the Grady-Alexander Thesis*, 78 VA. L. REV. 359, 379 n.73 (1992). However, the proposition that the inventor need not enable subject matter infeasible with current technology seems to reduce to the tautological proposition that the inventor need not enable technology that is not enabled. The idea that claim scope is fixed by enablement at the time of filing provides a tool to resolve cases in which the denotation of a word used in the claim expands over time. See Christopher A. Cotropia, *After-Arising Technologies and Tailoring Patent Scope*, 61 N.Y.U. ANN. SURV. AM. L. 151, 165-68 (2005); Kevin E. Collins, *The Reach of Literal Claim Scope Into After-Arising Technology: On the Construction of Things and Meanings*, CONN. L. REV. (forthcoming 2008). However, the principle seems of little use in cases that do not involve a change in meaning over time. See *infra* text accompanying notes 108-124. Nor (*ipso facto*) can it constrain the scope of claims only involving current technology. See *infra* text accompanying notes 134-141. Still further, such a rule leaves unanswered the question of the scope of the patentee’s entitlement when future developments are known but not yet technologically possible. For example, in the *Amgen* case, the defendant’s technology (homologous recombination of transcription control sequences into human cells) was “known” in some sense at the time of the invention, because similar techniques existed for microorganisms and the application of such techniques to mammalian cells was an eagerly desired advance. The Federal Circuit has suggested that “nascent” technology must be specifically enabled by the disclosure but technology farther in the future need not be. See *Chiron Corp. v. Genentech, Inc.*, 363 F.3d 1247, 1254 (Fed. Cir. 2004) (leading to the peculiar result that, all other things being equal, a given disclosure “enables” more technology later in time than earlier); see also Collins *supra* (discussing *Chiron*, 363 F.3d 1247).

101. According to Bertrand Russell, the failure of Aristotle (or of Aristotle’s expounders) to recognize the ontological distinction between individuals and classes led to disastrous consequences in philosophy and in number theory. BERTRAND RUSSELL, A HISTORY OF WESTERN PHILOSOPHY, AND ITS CONNECTION WITH POLITICAL AND SOCIAL CIRCUMSTANCES FROM THE EARLIEST TIMES TO THE PRESENT DAY 198 (1945).

and there is no reason to demand that legal properties follow the rules applicable to physical entities.

Unfortunately, while simply declaring claims 2 and 3 non-enabled might solve the fanciful paradox presented here, it would not solve the very real problems that arise when assessing claims that include, as all claims do, non-enabled subject matter. The difficulties with current enablement doctrine, even when exotic technologies are not at issue, are evident from the Federal Circuit's recent enablement jurisprudence. The court has held claims on fairly conventional technologies invalid for lack of enablement under section 112, holding that the disclosure must enable the "full scope" of the patent claims.<sup>102</sup> The court has not defined "full scope," other than to indicate that section 112 requires "reasonable enablement,"<sup>103</sup> and to suggest that failure to enable "a significant portion of the subject matter encompassed"<sup>104</sup> by the claims renders the claims invalid under section 112. Due to the infinite scope of patent claims, a patentee certainly need not, and in most instances cannot, enable every embodiment falling within the "full scope" of the claims. For example, a patentee who discloses an operable industrial process may claim that process broadly, even though the patentee has *not* enabled commercially refined variants of the process within the scope of the claim.<sup>105</sup> Though the Patent Office has recognized that not every embodiment within the scope of the claims must be enabled,<sup>106</sup> it is not clear if the Federal

---

102. *See* *Sitrick v. Dreamworks, LLC.*, 516 F.3d 993 (Fed. Cir. 2008); *Pharm. Res., Inc. v. Roxane Labs., Inc.*, 253 F. App'x 26 (Fed. Cir. 2007); *Auto. Techs. Int'l, Inc. v. BMW of N. Am., Inc.*, 501 F.3d 1274 (Fed. Cir. 2007); *Liebel-Flarsheim Co. v. Medrad, Inc.*, 481 F.3d 1371 (Fed. Cir. 2007); *AK Steel Corp. v. Sollac*, 344 F.3d 1234, 1241 (Fed. Cir. 2003).

103. *AK Steel*, 344 F.3d at 1244.

104. *Id.* at 1245.

105. *See* *CFMT, Inc. v. Yieldup Int'l Corp.*, 349 F.3d 1333, 1338-40 (Fed. Cir. 2003) (rejecting argument of invalidity based on extensive experimentation necessary to achieve commercial embodiment). *But see* *Ormco Corp. v. Align Tech., Inc.*, 498 F.3d 1307, 1319 (Fed. Cir. 2007) ("If an inventor attempts but fails to enable his invention in a commercial product that purports to be an embodiment of the patented invention, that is strong evidence that the patent specification lacks enablement."). *CFMT* seems to highlight the inadequacy of temporal arguments alone to resolve the enablement paradox. *See supra* note 100. Were the experiments needed to optimize the drying step of the cleaning process in *CFMT* an "after-arising" or "nascent" technology? Only, it seems, if *anything* not known at the time of filing is considered after-arising.

106. *See, e.g., Ex parte Breakefield*, No. 2001-1686, 2002 WL 32346083, at 3-4 (B.P.A.I. Feb. 7, 2002) (holding claim enabled where skilled artisan could distinguish between enabled and non-enabled embodiments).

Circuit's recent "full scope" jurisprudence recognizes this basic principle.<sup>107</sup>

What *is* clear is that reconciliation of enablement doctrine with a formal conception of patent law is difficult, perhaps impossible, without resort to disclosure doctrines beyond enablement. As an illustration, consider the Federal Circuit's decision in *AK Steel Corp. v. Sollac*,<sup>108</sup> the case that inaugurated the current "full scope of enablement" line of authority. The patents at issue in *AK Steel* were compositions of matter: aluminum-coated stainless steel strips made by an improved process of hot-dipping the steel strips in a coating solution.<sup>109</sup> Standard industry coating solutions, known as "Type 1" coatings, included about 10% silicon. The inventors discovered that the inclusion of silicon inhibited the coating process. The patent specifications therefore stated that pure coating solutions with little or no silicon (known as "Type 2" coatings) were preferred for their invention. However, the patent issued with an

---

107. With respect to the "full scope" requirement, the court in *AK Steel* stated: That is not to say that the specification itself must necessarily *describe* how to make and use every possible variant of the claimed invention, for the artisan's knowledge of the prior art and routine experimentation can often fill gaps, interpolate between embodiments, and perhaps even extrapolate beyond the disclosed embodiments, depending on the predictability of the art.

*AK Steel*, 344 F.3d at 1244 (emphasis added). This language can be read to suggest that while the specification need not *describe* every embodiment within the scope of the claims, it must enable one of skill in the art to practice every embodiment within the scope of the claims without undue experimentation. Historically, case law has been clear that the specification need not disclose every embodiment within the scope of the claims, but has usually done so in the context of whether one of skill in the art would have to experiment unduly to identify operable species within the claimed parameters. The question of whether each embodiment within the claims must be embodied within the claims has not been addressed directly, but a rigid requirement would run counter to the sentiments expressed in the historical case law. *See, e.g., In re Angstadt*, 537 F.2d 498 (C.C.P.A. 1976). Interestingly, such precedent essentially rejects the synonymy of the claimed invention and the scope of the inventor's legal rights. *See id.* at 504.

By calling the claimed "invention" the "scope of protection sought" the dissent obscures the problem and frustrates the intended operation of the patent system. Depriving inventors of claims which adequately protect them and limiting them to claims which practically invite appropriation of the invention while avoiding infringement inevitably has the effect of suppressing disclosure.

*See also infra* Part IV (discussing conflation of invention, claim, and legal right).

108. 344 F.3d 1234.

109. *See id.* at 1236-37.

independent claim that did not limit the amount of silicon the coating solution:

1. A ferrous base ferritic strip continuously hot dip coated with a coating metal, comprising:
2. . . . the coating metal including aluminum or aluminum alloys . . .<sup>110</sup>

as well as a dependent claim explicitly reciting coating solutions with substantial amounts of silicon:

3. The strip of claim 1 wherein the aluminum coating metal contains up to about 10% by weight silicon.<sup>111</sup>

Thus, the patent claimed, in an independent claim, a broad genus not defined by any particular silicon content, and in a dependent claim, a narrower genus that encompassed the “Type 1” coating explicitly advised against by the disclosure.<sup>112</sup>

The patentee asserted the patent against a defendant whose stainless steel strips were coated with a solution containing about 8% silicon.<sup>113</sup> In light of evidence that one of ordinary skill in the art could not practice the invention using a coating solution with about 10% silicon, the Federal Circuit held that *both* independent claim 1 and dependent claim 3 were invalid for lack of enablement.<sup>114</sup>

What is remarkable about *AK Steel* is that its outcome was dictated by a formalist conception of patent law and of patent claim structure in particular. Consider the alternative courses the court could have taken. The Federal Circuit’s own case law had long suggested that a claim encompassing “inoperative embodiments” would not be invalid for lack of enablement, so long as one of ordinary skill in the art could identify the

---

110. Hot Dip Aluminum Coated Chromium Alloy Steel, U.S. Patent No. 5,066,549 (filed Nov. 22, 1988).

111. *Id.*

112. In the infringement suit, both the district court and the Federal Circuit construed the dependent claim to include Type 1 coatings with substantial silicon. *AK Steel*, 344 F.3d at 1240-43. The Federal Circuit held that, despite the maxim that claims are to be construed in order to preserve their validity, the clear literal word of the claim, and the prosecution history of the ’549 patent, demanded that the claim be construed to cover coating solutions with about 10% silicon.

113. *Id.* at 1238.

114. *Id.* at 1245. Other claims in the ’549 patent were essentially parallel to claims 1 and 3, and met the same fate.

inoperative embodiments without undue experimentation.<sup>115</sup> One of skill in the art would certainly not have to experiment unduly to exclude the high-silicon embodiments that fell within the claims, since the disclosure quite simply instructs him to avoid them.<sup>116</sup> The court's opinion does not refer to the "inoperative embodiments" doctrine, perhaps with good reason.<sup>117</sup> If enablement were the only disclosure requirement of section 112, then the logical conclusion of the "inoperative embodiments" doctrine would be that the patentee may draft a claim of the form:

1. Everything.

and have no difficulties with section 112, so long as the specification directs one of skill in the art to confine himself to one or two embodiments enabled by the disclosure.<sup>118</sup>

The second alternative would have been to construe at least claim 1 to exclude high-silicon coatings, given that the specification explicitly disclaimed such embodiments. Such a construction would not only comport with the maxim that claims are interpreted in light of the specification, but would also avoid the invalidation of claim 1 for lack of enablement. Why did the court not choose this course? In part, a formalist "plain meaning" principle of interpretation drove the court's decision. The claim, which recited a coating containing "aluminum or aluminum alloys," had no language limiting the coating's silicon content. Notwithstanding the principle that claims ought to be construed to preserve their validity,

---

115. *Atlas Powder Co. v. E.I. du Pont de Nemours & Co.*, 750 F.2d 1569, 1576-77 (Fed. Cir. 1984); *In re Cook*, 439 F.2d 730, 735 (C.C.P.A. 1971).

116. *See In re Vaeck*, 947 F.2d 488, 496 (Fed. Cir. 1991) ("[T]he disclosure must adequately guide the art worker to determine, without undue experimentation, which species among all those encompassed by the claimed genus possess the disclosed utility.").

117. It also may be that the litigants did not argue the "inoperative embodiments" doctrine, as there is no mention of the doctrine in the district court's lengthy opinion. *See AK Steel Corp. v. Sollac*, 234 F. Supp. 2d 711 (S.D. Ohio 2002).

118. Of course, prior art also limits claim scope; *see infra* Part III.B. To be fair, the *Atlas Powder* doctrine could be read more narrowly. *Atlas Powder* states that "Of course, if the number of inoperative combinations becomes significant, and *in effect* forces one of ordinary skill in the art to experiment unduly in order to practice the claimed invention, the claims might indeed be invalid." *Atlas Powder*, 750 F.2d at 1576-77 (emphasis added). One could interpret this passage to mean that a large number of inoperative embodiments is equivalent to undue experimentation, even if one of skill in the art could easily identify and exclude the inoperative embodiments.

the court would not apply that principle absent any lexical ambiguity in the claim language.<sup>119</sup>

However, the “plain meaning” argument was secondary to the decision in *AK Steel*. More important than the “plain language” principle was the Federal Circuit’s focus on the hierarchical structure of patent claims. According to the court, claim 1 *must* encompass coatings with up to 10% silicon, because claim 3, which depended from claim 1, explicitly recited “up to about 10% silicon.” In the court’s view, because claim 3 depended from claim 1, claim 3 must define a sub-genus *entirely contained* within the scope of claim 1. Given that claim 3 clearly encompassed high-silicon coatings, the supra-genus defined by claim 1 must encompass them as well.<sup>120</sup> This line of reasoning is predicated on a formal, hierarchical view of claim structure: every dependent claim, because it merely adds limitations to another claim, must constitute a sub-genus of its parent claim. The subject matter encompassed by every independent claim must therefore be a superset of the subject matter encompassed by its dependent claims.

Such a view of claim structure has an impeccable pedigree. It simply recapitulates Aristotle’s scheme of categorization, in which all things that exist may be classified in a hierarchical structure of genus and sub-genus.<sup>121</sup> It is not, however, required by the patent statutes. 35 U.S.C. § 112 simply requires that dependent claims add a further limitation to subject matter already claimed and are construed to include all the limitations of the independent claim.<sup>122</sup> It does not require that independent claims encompass all the subject matter defined by the dependent claim, nor does it require that any claim making reference to another claim be construed as a dependent claim. A priori, one could conceive of dependent claims that included all the limitations of an independent claim but whose subject matter was not entirely included within the independent claim, or one could conceive of claims that incorporate the limitations of other claims without being dependent on those claims. In fact, decisions of the PTO’s Board of Patent Appeals and Interferences have considered the possibility that a claim referring to another claim might be treated as an independent claim, and that the

---

119. See *AK Steel*, 344 F.3d at 1243. See also *MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1332 (Fed. Cir. 2007) (“[V]alidity construction should be used as a last resort, not a first principle. . .”).

120. *AK Steel*, 344 F.3d at 1242.

121. See *infra* Part IV.C.

122. 35 U.S.C. § 112, ¶ 4 (2000).

reference to another claim is only a shorthand form of drafting.<sup>123</sup> Nonetheless, the Federal Circuit's intrinsic commitment to the hierarchical conception of claim structure foreclosed this route in *AK Steel*, leaving the court with no choice but to conclude that the independent claims encompassed the subject matter defined by the dependent claims.

After defining the patent's scope according to a formal conception of claim structure, the *AK Steel* court proceeded to determine enablement of the claims, and it is at this point the uncomfortable interface between the formal structure of patent law and the enablement doctrine is apparent.<sup>124</sup> The evidence showed that the patent's specification did not enable one of ordinary skill in the art to use high-silicon "Type 1" coating solutions in the manufacture of the claimed steel strips. Having construed the independent claims to encompass Type 1 coating solutions, the court held those claims invalid because "the specification [did] not enable a significant portion of the subject matter" claimed by the patent.<sup>125</sup> Yet why were the Type 1 coatings a "significant portion" of the claimed subject matter? As we have seen, all claims are infinite. While there were an infinity of non-enabled Type 1 compositions within the scope of the claims, there were also an infinity of enabled Type 2 (low-silicon) compositions within the scope of the claims. It is not apparent why the first set ought to be more significant than the second, especially given that there were also an infinite number of non-enabled *Type 2* compositions within the scope of the claims.<sup>126</sup>

Of course, the non-enablement of Type 1 embodiments was significant in *AK Steel* because the accused infringer practiced a Type 1 embodiment. Nonetheless, whether the accused subject matter is enabled by the

---

123. See *Ex parte* Porter, 25 U.S.P.Q. (BNA) 2d 1144, 1147 (B.P.A.I. 1992); *Ex parte* Moelands, 3 U.S.P.Q. (BNA) 2d 1474, 1476-77 (B.P.A.I. 1987) (Spencer, J., dissenting in part, and Lovell, J., concurring in the result).

124. *AK Steel* could well have been a trivial case had it been resolved on written description grounds. The invention was described by the specification as employing low-silicon coatings, not merely as a particular embodiment, but as a general property of the invention. Should the claims still be construed to define a genus of coatings without limitation to silicon content, they would not correspond with the specification's fixation of the invention within the definitional hierarchy. See *infra* Part IV.

125. *AK Steel*, 344 F.3d at 1245.

126. Consider, for example, Type 2 coatings dipped at extreme temperatures, or with immiscible materials. Of course, one of ordinary skill in the art would certainly know to avoid these embodiments. However, one of ordinary skill in the art would also know to avoid the Type 1 embodiments, given the specification's explicit teachings.

disclosure is ostensibly irrelevant to the question of enablement;<sup>127</sup> certainly under the formal conception of patent law, enablement is a function solely of the claim and the disclosure, and not dependent on whom the patentee sues. Perhaps the “full scope of enablement” doctrine can be expressed in terms of some calculus of infinities, in which the relative proportions of enabled and non-enabled subject matter are assessed.<sup>128</sup> We might, for example, conclude the claim is enabled if one of ordinary skill in the art could practice some proportion of the embodiments falling within the claims. Nonetheless, since nearly all claims encompass non-enabled embodiments, the difficulty is in deciding *which* non-enabled embodiments are significant in the analysis. When claims recite particular numeric ranges, we tend to focus on whether the claimed subject matter functions with parameters lying along an axis defined by the range. In reality, however, all claims encompass subject matter defined by a very large number of axes, some explicit in the claim and some not,<sup>129</sup> and it is difficult to label, a priori, one axis significant for enablement and another not. The answer given in *AK Steel* seems to be that embodiments practiced by the accused infringer are significant in the enablement inquiry. Nonetheless, if the answer to the question of significant characteristics is “those possessed by the accused subject matter” or “those brought to the court’s attention,” then it seems that we have abandoned the notion of a coherent system of patent law based on the peripheral claim.<sup>130</sup> Indeed, if we conclude that the validity or scope of a patent depends on whom the patentee asserts it against, then we may have abandoned the conception of patents as objectively defined property rights altogether.

### **B. Can Enablement Limit Claim Scope?**

The preceding section contended that, in light of the nature of patent claims, the enablement doctrine faces fundamental difficulties as a coherent or complete doctrine of patent scope. This section explores the

---

127. *See, e.g.*, *Durel Corp. v. Osram Sylvania Inc.*, 256 F.3d 1298, 1306 (Fed. Cir. 2001) (“The dispositive question of enablement does not turn on whether the accused product is enabled.”).

128. Of course, if we discard the purely hierarchical view of enablement, cases like *AK Steel* may become formally coherent; the independent claim may be enabled notwithstanding the existence of non-enabled dependent claims.

129. For example, a claim to an object encompasses a collection of things of varying size or material, while a claim to a process encompasses a collection of acts performed under various conditions such as temperature.

130. *See infra* note 268 for discussion of abandonment of the peripheral system.

consequences of that deficiency for the permissible scope of patent claims and the structure of claiming itself. Its vehicle is to postulate a patent system that lacks additional disclosure doctrines such as written description. This section demonstrates that a patentee's entitlement is largely unbounded in such a system where enablement is the only disclosure doctrine limiting patent scope.

1. *Claim Scope Without Written Description*

The limits on a patent's scope essentially derive from only two sources: the prior art at the time of the invention, and the inventor's disclosure.<sup>131</sup> If we take the reductionism of the formalist conception at face value, these limits can be embodied in only two doctrines: nonobviousness and enablement. The doctrine of nonobviousness embodies all the limitations imposed by the prior art because it functions as a superset of novelty.<sup>132</sup> Enablement limits claim scope based on the inventor's disclosure; at least nominally, this limitation embodies the quid pro quo of the patent system that an inventor's exclusive rights be commensurate with the benefits conferred on society by his disclosure.<sup>133</sup> At least since *Lilly*, however, the need for an additional disclosure doctrine to circumscribe claim scope, the "written description" aspect of section 112, has been contested. Subsidiary to this controversy has been the question of whether the written description requirement is confined to chemistry and biotechnology or is applicable to other arts as well.

The problem, I believe, can be boiled down to a very simple hypothetical devoid of reference to any particular technology. Suppose a patent applicant to file an essentially empty disclosure, with the following claim:

---

131. I assume for this discussion that there are no questions of subject matter eligibility or compliance with the technical requirements of the law.

132. There are some technical limitations to this principle. Subject matter which is in public use may anticipate a claim even if its existence or properties were unknown to those skilled in the art. Formally, such unknown subject matter might not be regarded as obvious. *See, e.g.,* *TorPharm, Inc. v. Ranbaxy Pharms., Inc.*, 336 F.3d 1322, 1327 (Fed. Cir. 2003) (discussing anticipation and obviousness in the absence of disclosure).

133. *But see* *Holbrook*, *supra* note 62, at 131-46 (arguing against disclosure function).

4. All material objects which are enabled by the prior art, excluding those which are known or obvious in light of the prior art.<sup>134</sup>

where “enabled” here means that the material object can be made and used without undue experimentation given the current state of the art. Ought the Patent Office allow this claim?

It seems self-evident that claim 4 ought not to be patentable. Still, explaining exactly why claim 4 runs afoul of the statutory requirements for patentability is not a trivial exercise. By its own terms, claim 4 only encompasses subject matter that is novel, nonobvious, and enabled as prescribed by statute.<sup>135</sup> Therefore, insofar as the *subject matter itself*, there is no bar to the patentability of claim 4 if the doctrines of nonobviousness and enablement alone limit the scope of patent claims. If claim 4 is not patentable, it must be either that there is something impermissible about drafting a claim according to the fashion of claim 4, or that doctrines beyond nonobviousness and enablement are necessary to limit patent scope.

Claim 4’s scope is admittedly limited, but it is not empty. Claim 4’s scope is limited because most things obvious in light of the prior art are enabled by the prior art, and so claim 4 encompasses only the set of objects defined by the difference between the set of enabled objects and the set of obvious objects. *If* the standards of nonobviousness and enablement were symmetrical, then this difference would be the empty set and claim 4 would cover nothing. Any object that was enabled by the prior art would also be obvious in light of the prior art and hence unpatentable. However, the standards of enablement and nonobviousness are symmetrical neither in theory nor in practice.<sup>136</sup> Most notably, the judicial

---

134. I limit these hypothetical claims to material objects for simplicity and to avoid issues of patentable subject matter under section 101, but the principles are applicable to claims for methods and other intangibles as well.

135. Of course much of the subject matter defined by claim 4 is not “useful” as required by section 101, but if one accepts definition by such concepts as “obvious” or “enabled” then it is little stretch to supplement the definition with “useful.”

136. Before the notions of obviousness and enablement were clearly differentiated, the standard may have been more symmetrical. Writing in 1873, Curtis states that a specification will render the patent void if it “create[s] a necessity for the exercise of *inventive power* on the part of the person who has [undertaken] to apply the description.” GEORGE T. CURTIS, A TREATISE ON THE LAW OF PATENTS FOR USEFUL INVENTIONS § 256 (4th ed. 1873) (emphasis in original). The term “inventive power” suggests a cognitive aspect inherent in the modern doctrine of nonobviousness. However, Curtis’s discussion

standard for enablement, that the ordinary artisan ought to be able to make and use the invention without “undue experimentation,” invokes the effort required to produce the invention given the state of the art. In contrast, the statutory standard for nonobviousness under section 103 explicitly discourages inquiry into the inventive effort, declaring that “[p]atentability shall not be negated by the manner in which the invention [would be] made.”<sup>137</sup> Therefore, if only the doctrines of enablement and nonobviousness constrain patent scope, claim 4 defines an actual slice of patentable subject matter.

One might object that claim 4 fails to satisfy the definiteness requirement of 35 U.S.C. § 112, ¶ 2. However, indefiniteness requires insoluble linguistic ambiguity,<sup>138</sup> and under this standard the metes and bounds of claim 4 are clear. We cannot envision all the entities falling within the scope of claim 4, nor can we recite all their characteristics, nor can we say with certainty at the present what future creations will fall within claim 4. This is true, however, of all patent claims, for all patent claims are infinite and cannot specify all of the characteristics of the subject matter that they encompass. If claim 4 is less indefinite than an ordinary claim, like the chair claim of claim 1, it cannot be because of the breadth of the claim or because particular subject matter falls within the claim.<sup>139</sup> Indeed, to raise such arguments against claim 4 nearly negates the argument of indefiniteness. Arguments based on breadth are predicated upon a determination that something is within the scope of the claims, which in turn is predicated upon the ability to recognize that entities are or are not within the claim.

If claim 4 is indefinite, it must be because the properties recited by the claim—“material,” “enabled,” and “known or obvious”—are qualitatively different from the properties recited by claim 1, such that they fail to give an answer to the question of whether a particular entity falls within the claim or not. One of ordinary skill in the art could certainly decide whether putative subject matter is a material object or not; the objection must lie with the use of the properties “enabled” or “known or obvious.”

Perhaps we have committed an ontological foul by including *legal* properties, rather than physical ones, as part the definition of subject

---

is mostly centered on undue experimentation as the standard for adequacy of the specification.

137. 35 U.S.C. § 103(a) (2000).

138. See text accompanying notes 31-32 *supra*.

139. See *In re Fisher*, 427 F.2d 833, 838 (C.C.P.A. 1970) (explaining that breadth of claim is irrelevant to indefiniteness).

matter represented by claim 4. However, it is difficult to see why such legal properties would render the claim indefinite. Both nonobviousness and enablement are factually premised on the judgment and ability of one of ordinary skill in the art. In the case of nonobviousness, whether the ordinary artisan would find the differences between the invention and the prior art obvious at the time the invention was made; in the case of enablement, whether the ordinary artisan could make and use the invention without undue experimentation.<sup>140</sup> In deciding whether claim language is sufficiently definite to satisfy section 112, ¶ 2, we rely on the knowledge of one of skill in the art to assign meaning to claim terms otherwise indeterminate on their face.<sup>141</sup> It would therefore be peculiar to conclude that one of ordinary skill in the art could not determine the metes and bounds of claim 4 because one of ordinary skill in the art could not assess whether subject matter is nonobvious or enabled, especially given that in patent litigation we entrust the determinations of nonobviousness and enablement to lay judges and juries.

We might instead conclude that claim 4 is indefinite not because there is something wrong with defining subject matter in terms of legal doctrines generally, but rather because the doctrines of nonobviousness and enablement are insufficiently refined to provide definitive answers to whether particular subject matter falls within claim 4 or not. Then, however, the objection to claim 4 is premised entirely on the uncertainty in our current doctrines of enablement and nonobviousness. If we posit readily ascertainable standards of enablement and nonobviousness, then it becomes untenable to argue that one of ordinary skill in the art cannot ascertain the scope of the claim. It is difficult to believe that all objections to claim 4 would disappear if the law of nonobviousness or enablement were more precise than it is today. Moreover, if the ultimate goal in demonstrating the unpatentability of claim 4 is to prove that patent law needs only the doctrines of nonobviousness and enablement to satisfactorily limit claim scope, then it seems a Pyrrhic victory to reach that conclusion on the grounds that the current doctrines of nonobviousness and enablement are hopelessly indeterminate.

---

140. To the extent that determining the bounds of claim 4 only requires a judgment about whether *particular subject matter* would be obvious or enabled to one of skill in the art, then the determination may be simpler than the determination of whether *a claim* is obvious or enabled.

141. *See, e.g., Orthokinetics, Inc. v. Safety Travel Chairs, Inc.*, 806 F.2d 1565, 1575-76 (Fed. Cir. 1986) (stating that claim reciting “so dimensioned” is definite if one of ordinary skill in the art can obtain measurements).

## 2. *The Relation Between Enablement and Nonobviousness*

Claim 4 shows us that, at least at the formal level, the doctrines of enablement and nonobviousness provide only incomplete limits on the scope of the patentee's claims. In particular, without additional disclosure requirements, there is no necessary relation between the inventor's disclosure and the scope of the rights granted to the inventor. What are the implications for patent law? The implications depend on how we respond to the problem posed by claim 4. If we remain within our current system of defining the inventor's rights by peripheral claims, then it seems that there are three possible responses. The first is to simply concede that the patentee is entitled to the full scope of claim 4. The second is to modify our doctrines of enablement or nonobviousness to eliminate the sliver of subject matter lying between the two doctrines. The third is to invoke an additional disclosure doctrine as a limitation on the scope of the patentee's rights.

The first response to the problem posed by claim 4 would be to declare that it is not a problem. No one files patent applications with empty disclosures, but claim 4 can be made more realistic by modifying it slightly. Suppose the applicant files an application with some disclosure, and the following claim:

5. All material objects which are enabled by the combination of my disclosure and the prior art, excluding those which are known or obvious in light of the prior art.

Claim 5 represents a more plausible situation than claim 4, but does not resolve the question of scope: all things that were within the scope of claim 4 are also within the scope of claim 5.

Still, claims 4 and 5 may not be problematic at all. One might argue that at any given point in time, all inventions that are enabled by current technology will be quickly disclosed, either by being claimed in another patent application or otherwise made known to the public.<sup>142</sup> Therefore, the sliver of claimable subject matter lying in the gap between enabled and obvious subject matter will not exist in the absence of new information contributed by the inventor. On this view, since the standards of

---

142. This argument is similar to the one advanced by Judge Rader against the need for a separate written description doctrine. Judge Rader argued that inventions enabled by a technological advance at a particular point in time will inevitably be disclosed and claimed by some inventor in a patent application. *See Univ. of Rochester v. G.D. Searle & Co., Inc.*, 375 F.3d 1303, 1312 (Fed Cir. 2004) (Rader, J., dissenting from denial of rehearing en banc).

enablement and nonobviousness evolve with advances in the art, the subject matter defined by the gap between them will remain insignificant. We will encounter problems only when decisions of the courts have legally cemented the standards of enablement or nonobviousness rather than let them flow with technological advance. For if legal fixation has significantly decoupled the standards of enablement and nonobviousness, then a minimal disclosure by an inventor may render a large swath of subject matter enabled but also nonobvious.

To illustrate, we might account for *Lilly* under this theory as follows.<sup>143</sup> The development of a new general technology, recombinant DNA, opened the door to a large category of inventions based on the recovery of human genetic sequences, although the patent in *Lilly* was not one disclosing or claiming this general technology. Rather, making use of this new technology, the patentee in *Lilly* had isolated and disclosed a DNA molecule encoding rat insulin. In addition to claiming the rat insulin DNA, the patentee also claimed human insulin DNA, whose sequence was not yet known at the time of filing, and the broader genus of vertebrate insulin DNA molecules. Although enablement was not litigated in *Lilly*, it is arguable that under the prevailing standard, the isolation of human insulin DNA (or other vertebrate insulin DNA molecules) by sequence homology from the rat molecule would not have required undue experimentation and therefore was enabled.<sup>144</sup> Thus, general technological advance had rendered the invention, along with many other human DNA molecules, enabled. The same technological advance should have rendered a correspondingly large swath of subject matter obvious. However, Federal Circuit precedent, particularly *In re Deuel*, suggested that because the prior art did not suggest the structure of the claimed molecule itself,

---

143. Professor Holman makes a similar point on the asymmetry of enablement and nonobviousness in his work on written description. See Holman, *supra* note 48, at 65-67. In the *Lilly* litigation, Lilly raised neither obviousness nor lack of enablement as grounds for invalidity. The rationale behind such a strategy is unclear from the judicial opinions of the case, although Lilly may have been reluctant to raise an enablement challenge: Lilly's argument that one of the patents at issue was anticipated rested on the premise that a disclosure of insulin polypeptide sequence in the prior art enabled one of skill in the art to make an insulin cDNA molecule. See *Regents of Univ. of Cal. v. Eli Lilly & Co.*, 39 U.S.P.Q. 2d (BNA) 1225 (S.D. Ind. 1995), *aff'd in part, rev'd in part*, *Regents of Univ. of Cal. v. Eli Lilly & Co.*, 119 F.3d 1559 (Fed. Cir. 1997).

144. Under *In re Wands*, routine screening of human cDNA library may have sufficed to retrieve a human insulin cDNA. *In re Wands*, 858 F.2d 731, 736 (Fed. Cir. 1988); see Arti K. Rai, *Intellectual Property Rights in Biotechnology: Addressing New Technology*, 34 WAKE FOREST L. REV. 827, 835 (1999).

the claimed species of human insulin DNA would be nonobvious.<sup>145</sup> Human insulin DNA was therefore, more by case law than by technology, within the zone of things enabled but not obvious in light of prior art. By decoupling the standard of nonobviousness from considerations such as the ease of obtaining the chemical entity, the Federal Circuit drew the boundaries of nonobviousness such that all of the incipient information about genetic sequences fell into the gap between enablement and obviousness. The invocation of the written description doctrine in *Lilly*, and subsequent tightening of the written description and utility standards by the courts or the Patent Office, has been an attempt to limit the subject matter otherwise made patentable by the artificial divergence of the enablement and nonobviousness standards. If the standards of enablement and nonobviousness were permitted to properly float with the advancement of the technological arts, there would be no need to invoke other doctrines to circumscribe claim scope.<sup>146</sup>

There are nonetheless significant difficulties with the argument that only in unusual circumstances will patentable subject matter lie in the gap between enablement and nonobviousness. For one, even if the standards of nonobviousness and enablement freely advance with the art, the scope of inventions defined by claim 4 is decidedly nontrivial. Consider patents on simple mechanical inventions. Such inventions, not involving any radical technological advance, have been enabled by the state of the art for years. *Yet they were not made*, presumably because they were not obvious.<sup>147</sup> Likewise, there exist many inventions in which the inventive activity consists of recognizing a problem; once the problem is recognized, the solution is well within the technological capabilities of the art without further contribution from the inventor.<sup>148</sup> Perhaps inventions that do not

---

145. See Holman, *supra* note 48, at 65-66. The genus claim might nonetheless have been obvious under *Durel Corp. v. Osram Sylvania Inc.*, 256 F.3d 1298 (Fed. Cir. 2001) if the structure of at least some insulin polypeptides encoded by the claimed genus of DNA molecules was known.

146. Note, however, that under this account the patentee might not have been entitled to claim the rat insulin cDNA either, if the cDNA molecule encoding rat insulin was considered obvious over the known rat insulin polypeptide sequence. The account of *Lilly* relying on only enablement and nonobviousness would therefore have precluded the patenting of many, if not most, of the first-generation biotechnological inventions.

147. Of course, many inventions are not made because they are not perceived as worthwhile, though the recognition itself that an invention would in fact be worthwhile is a form of nonobviousness.

148. See, e.g., *In re Spinnoble*, 405 F.2d 578, 585 (C.C.P.A. 1969) ("It should not be necessary for this court to point out that a patentable invention may lie in the discovery of

open up new technological possibilities ought not to be patentable. Justice Douglas argued in his concurrence in *Great Atlantic & Pacific Tea Co. v. Supermarket Equipment Corp.*<sup>149</sup> that the constitutional mandate to promote the useful arts demanded patents that “push[ed] back the frontiers of chemistry, physics, and the like,”<sup>150</sup> rather than mere “gadgets,” and the Supreme Court recently reminded us that *Great A&P* remains relevant to the law of nonobviousness.<sup>151</sup> If we continue, however, to grant patents on inventions that are possible with current technology, we must preserve a zone of patentability between subject matter that is currently enabled and subject matter that is currently nonobvious. It follows that the standards of enablement and nonobviousness ought not to be perfectly symmetrical, and that prior inventors cannot be entitled a priori to all things enabled by their disclosures in combination with the prior art.

### C. Rethinking Enablement

If we are not content to ignore the problems posed by claims such as claims 4 and 5, then we must either rethink enablement (and possibly nonobviousness as well), or turn to additional disclosure doctrines to limit claim scope. Before considering modifications to the enablement doctrine, it is worth re-emphasizing the conceptual problems of relying solely upon enablement as a doctrine of claim scope. The uncertainty inherent in determining whether one of ordinary skill in the art could make and use a particular thing without undue experimentation is unremarkable. The true uncertainty in enablement lies in the absence of a defined relationship between the question of whether one of ordinary skill in the art could make and use *a particular thing*, and the question of whether one of ordinary skill in the art could make and use *the claimed invention*. As noted above, the scope of the inventor’s possible rights is not synonymous with the scope of things enabled by the inventor; not only is the boundary between enabled claim and non-enabled claim indeterminate in practice, but there is no defined relationship between the intellectual framework of

---

the source of a problem even though the remedy may be obvious once the source of the problem is identified.”).

149. 340 U.S. 147 (1950).

150. *Id.* at 154-55 (Douglas, J., concurring). Of course, true discoveries in chemistry, physics, and the like would be unpatentable as natural laws; moreover, at the time the Constitution was drafted, one suspects the inventors of the young republic were more concerned with incremental improvements in agriculture or manufacturing than landmark scientific discoveries.

151. *See* *KSR Int’l Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1739 (2007) (emphasizing that cases like *Great A&P* remain good law).

the enablement inquiry (whether one of ordinary skill could make and use a *thing*) and the legal framework (whether one of ordinary skill could make and use *the claimed invention*).<sup>152</sup> Such inquiries may or may not be commensurable, but our current difficulty in relating them would seem to make enablement, at least as now conceived, a fragile foundation on which to place the full burden of determining permissible claim scope.<sup>153</sup>

1. *Converging Enablement and Nonobviousness*

These considerations aside, can the doctrines of nonobviousness or enablement be modified to constrain a patentee's rights without resort to a doctrine like written description? If we think that claim 4 fairly represents the problem, that the standards of nonobviousness and enablement are not symmetric, then perhaps convergence of the two doctrines will solve our problems.<sup>154</sup> The scope of claim 4 arises because the standards for enablement and nonobviousness are not symmetric. If the set of subject matter enabled by the prior art is the same as that made obvious by the prior art, then the scope of a claim like claim 4 is non-existent.<sup>155</sup>

Conceptually, enablement and nonobviousness seem difficult to merge. At least in theory, the doctrine of nonobviousness reserves for

---

152. See *infra* Part IV.C., on the distinction between thing and invention.

153. *Amgen* may be instructive here as well. *Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313 (Fed. Cir. 2003). In *Amgen*, the patentee, who had been the first to synthesize human erythropoietin, asserted claims to essentially all synthetic erythropoietin molecules regardless of how they were synthesized. Had the patentee sought claims explicitly directed to “[a] method of producing synthetic EPO,” the enablement requirement would likely have dictated that the claims be limited to the method of synthesis disclosed by the patentee. However, because the claims were drafted with a source limitation, “non-naturally occurring,” rather than an explicit reference to the process of producing EPO, the claims were considered “composition” claims for which a patentee need disclose only a single method of making to claim the composition no matter how made. The point is not whether the patentee in *Amgen* deserved broad claims. The point is that the indeterminate relationship between claim scope and disclosure characteristic of the enablement doctrine makes the entire scope question highly sensitive to fine points of the law of enablement.

154. Professor Durham discusses a proposal to make infringement by equivalents symmetrical with nonobviousness. See generally Durham, *supra* note 75. Because both the doctrine of equivalents and enablement are in some sense scope doctrines, some of Durham's analysis not tied to the particular concerns of nonliteral infringement might be invoked to support symmetrical standards of enablement and nonobviousness. See *id.* at 1013-19 (arguing that gauging scope of patentee's rights by nonobviousness comports with economic theories of patent law).

155. See also Holbrook, *supra* note 62, at 169-73 (arguing that standards of nonobviousness and enablement ought to converge on possession).

patent protection those inventions which would not have been made at all or in a timely manner, absent the incentives provided by the future grant of the patent monopoly. Enablement (or any other scope doctrine) defines in the end what share of present economic activity over which the inventor may exercise exclusive rights. To be sure, more extensive claim scope provides larger incentives, but unless we believe inventors can perfectly forecast the future then there is no precise relationship between the current value of technology and the incentive necessary to bring that technology into being.

Notwithstanding teleological distinctions between the doctrines, we might start by defining obviousness in terms of enablement: *All things enabled by the prior art are obvious*. However, while this symmetry would eliminate the scope of claim 4, it would also eliminate the category of inventions discussed above: the inventions which are feasible with current technology but have not yet been invented. Unless we are willing to exclude all such inventions from patentability,<sup>156</sup> nonobviousness cannot be defined solely in terms of enablement.<sup>157</sup>

Perhaps we can instead define enablement in terms of obviousness: *All things obvious from the prior art and the inventor's disclosure are enabled*. On this formulation, the inventor is entitled to claim all things which are obvious from the combination of the prior art and his disclosure. Since things obvious from the prior art alone are unpatentable, the inventor's rights would be defined in terms of what we might call "marginal obviousness:" those things which were not obvious from the prior art alone, but are obvious once the inventor's disclosure is considered.<sup>158</sup> This solution has considerable formal appeal. Though

---

156. A strictly utilitarian analysis might question whether we need to offer the patent incentive to things already enabled by the prior art. However, if we believe that the patent system, at least in part, functions to protect the investments needed to bring products to market even after a technological breakthrough has been achieved, then patent protection ought not to be refused solely on the grounds that the invention required no special technological advance. From a natural rights perspective, if an inventor has created an invention that would not exist at the present time but for his or her inventive power, then the entitlement to an exclusive right does not seem to depend on whether the invention required something we define as beyond the current skill in the art.

157. Such definition would also tend to complicate the nonobviousness inquiry; given how frequently such issues arise in patent procurement and litigation, this drawback is significant.

158. This is in part the approach advocated by Professor Feldman, though she does not label it as such. Professor Feldman proposes that, for instances in which the inventor did not disclose an accused embodiment but such an embodiment is information knowable at the time of the invention, the scope of the patentee's rights should depend on

nonobviousness is subject to factual uncertainty, formulating enablement in terms of nonobviousness makes it possible to define permissible claim scope while avoiding the difficulty enablement ordinarily faces with infinite claims. The inventor would be entitled to a halo of subject matter surrounding his or her disclosure, the extent of that halo being determined by what is obvious or not based on the disclosure and the prior art. Under this formulation, the inventor is entitled to a claim akin to the following:

6. All material objects which are obvious in light of the combination of my disclosure and the prior art, excluding those which are known or obvious in light of the prior art.

With claim 6, we have succeeded in defining a doctrine of claim scope that, while subject to factual uncertainty, is not subject to the conceptual uncertainty of current enablement doctrine. We have also succeeded in turning back the clock more than a century. What claim 6 defines is, in essence, a central claiming system, in which the inventor describes a core and the scope of his rights extends in a diminishing penumbra around the core. A quasi-central system, in fact, may be the only solution to the problems of scope raised by this Article.<sup>159</sup> But if we wish to adhere to our peripheral claiming system, we must declare claim 6 to be indefinite under 35 U.S.C. § 112, and encourage the inventor to draft peripheral claims that approximate claim 6 in scope. It is not certain whether claim 6 can be dismissed as indefinite: if one of ordinary skill in the art can recognize what is obvious and what is not obvious, then it appears that one of ordinary skill in the art can ascertain the boundaries of claim 6.<sup>160</sup> If one of skill in the art cannot recognize what it is obvious and what is not, then patent law seems headed for some difficulties, given that it is the one of ordinary skill in the art who decides whether subject matter is obvious or not under 35 U.S.C. § 103.<sup>161</sup>

---

whether the step from the disclosure to the accused subject matter is routine or “requires creativity, imagination or experimentation to derive.” Robin Feldman, *The Inventor’s Contribution*, 9 UCLA J.L. & TECH. 1, 35 (2005). She describes this inquiry as having “the indirect effect of measuring the inventive leap of the accused product.” *Id.* at 39. Measuring the level of mental or inventive activity required to create something seems very firmly rooted in the nonobviousness inquiry, though here framed in terms of the inventor’s disclosure rather than the prior art alone.

159. See *infra* Part IV.C.

160. See also *supra* text accompanying notes 139-142.

161. This conundrum may highlight the difficulty in employing the perspective of one of ordinary skill in the art, logically employed to make technological judgments such as nonobviousness, to decide essentially legal matters concerning the definition of the

Formal questions of claim structure aside, defining permissible claim scope along the lines of claim 6 also carries significant policy implications. By its terms, claim 6 excludes from patent scope all technological developments occurring after the date of the invention, except those which are obvious in light of current technologies. Some commentators have advocated this result by proposing that the inventor's rights be fixed in terms of the state of the art at the time of the invention.<sup>162</sup> However, while this limitation might seem appropriate for technologies following discontinuous patterns of technological improvement,<sup>163</sup> it is more difficult to apply it to technologies characterized by continuous and cumulative development. For ordinary technologies, few would agree that any nonobvious improvement upon a patented invention should escape infringement altogether. But whether one regards this as the optimal result on policy grounds, aligning enablement and nonobviousness excludes future technologies from patent protection.<sup>164</sup> To achieve a coherent scope doctrine within the confines of the peripheral claiming system, we must consider other modifications of the enablement doctrine or look beyond it altogether.

## 2. "Enablement Plus"

In fact, it is unnecessary to look further than current case law to find that such modifications to the enablement doctrine have already been made. Certain aspects of existing enablement law, while difficult to square with the nominal conception of enablement as a "make and use" requirement, can best be explained as responses to the problem of untethered claim scope epitomized by claim 4. The underlying objection to

---

patentee's exclusive rights. In this light, one may question whether the notion that claims ought to be interpreted from the perspective of one of ordinary skill in the art is truly tenable.

162. See, e.g., Robin Feldman, *Rethinking Rights in Biospace*, 79 S. CALIF. L. REV. 1, 40-41 (2005). Alternatively, greater reliance on the doctrine of equivalents may provide coverage for future developments. See generally Cotropia, *supra* note 100.

163. Feldman proposes this rule in the context of "uncertain arts" such as biotechnology. Feldman, *supra* note 162, at 40-41. The cases may be viewed as instances in which new technologies allowed the accomplishment of old results by radically different means. The true effect of "uncertainty" may be that technology proceeds erratically, with unpredictable leaps that open or revisit large areas of subject matter, and give new meanings to old words within the lifetime of a patent.

164. One might circumvent this difficulty by relying more heavily on the doctrine of equivalents to cover future developments. See, e.g., Holbrook, *supra* note 62, at 158. Ultimately heavier reliance on the doctrine of equivalents points towards abandonment of the peripheral claim system.

claim 4, and to a lesser extent claims 5 and 6, is not that the scope of protection conferred by those claims is not calibrated to the policy goals of patent law, but rather that the scope of the claim has little or nothing to do with *what the inventor actually invented*.<sup>165</sup> Two aspects of enablement doctrine embody the requirement of a nexus to what the inventor actually made or disclosed, a consideration irrelevant to the question of whether one of ordinary skill in the art could make and use the invention without undue experimentation, but relevant to an underlying concern that the inventor be entitled to claim only that which he invented.

Consider the principle that if a feature described by the disclosure as critical for the invention is not recited in the claim, the claim is invalid for lack of enablement.<sup>166</sup> Whether a feature described by the patentee as critical appears in the claim is not relevant to whether one of ordinary skill in the art could make or use the invention defined by the claim, but is relevant to the question of whether the claim is connected to what the inventor actually invented. This principle, though little applied in recent years,<sup>167</sup> shows that current enablement doctrine incorporates limitations

---

165. If we are completely convinced by Kitch's prospect theory, we might assert that assigning the patent to the "true inventor" is not strictly necessary; what matters is the existence of the property right, not to whom it is initially assigned. Of course, the long-term effects of denying patents to those who create would be corrosive if we believe that creation is motivated by the hope of a patent. Moreover, the patent and copyright clause of the Constitution gives Congress the power to grant exclusive rights to "inventors," thereby restricting exercise of power under the clause to actual inventors. In this light one might imagine that claim 4 could be invalid under 35 U.S.C. § 102(f), which denies patentability if the inventor "did not himself invent the subject matter sought to be patented." However, section 102(f) requires that the claimed invention be derived from someone else. *See, e.g.,* *Oddzon Prods., Inc. v. Just Toys, Inc.*, 122 F.3d 1396, 1401-02 (Fed. Cir. 1997) (explaining that section 102(f) renders invalid claims in which named inventor derived invention from another).

166. *See In re Mayhew*, 527 F.2d 1229, 1233 (C.C.P.A. 1976).

167. No subsequent majority opinion of the Court of Customs and Patent Appeals or the Federal Circuit has relied upon *Mayhew*. However, the *Mayhew* principle remains enshrined in the MPEP [§ 2164.08(c)], and the PTO and the Board of Patent Appeals and Interferences have relied upon *Mayhew* to reject claims failing to recite elements described by the inventor as essential. *See, e.g., Ex parte Araki*, No. 2003-1926, 2004 WL 4979022 (B.P.A.I. Mar. 2, 2004); *Ex parte Zacharias*, No. 2002-0741, 2002 WL 32346094 (B.P.A.I. Nov. 6, 2002). Arguably, *Mayhew* has been misinterpreted; the specification's emphasis on the omitted feature may only have been evidence tending to show that the broader claim lacking the feature was not enabled.

on claim scope beyond the requirement that the disclosure teach how to make and use the invention without undue experimentation.<sup>168</sup>

More recently, the Federal Circuit has held that the novel aspect of a claimed invention must be enabled by a specific disclosure in the specification, rather than by resort to the knowledge of one skilled in the art. In *Automotive Technologies International, Inc. v. BMW of North America, Inc.*,<sup>169</sup> the Federal Circuit found enablement lacking because the patent's specification disclosed only mechanical side impact sensors, and not electronic side sensors. Dismissing the patentee's argument that the knowledge of one skilled in the art could supply the information required to construct electronic side sensors, the court held that "[i]t is the specification, not the knowledge of one skilled in the art, that must supply the novel aspects of an invention in order to constitute adequate enablement."<sup>170</sup> Under the formal model of patent scope, in which the claim limitations define the category of subject matter to which the patentee is entitled, the "novel aspects" of the invention have no significance for enablement. Either the specification enables one of ordinary skill in the art to make and use the subject matter defined by the claim limitations, or it does not; the question of which aspects of the invention are novel is only relevant to novelty and nonobviousness.

---

168. Judge Baldwin of the Court of Customs and Patent Appeals, concurring in *Mayhew*, regarded the case not as a failure to meet the enablement requirement of section 112, ¶ 1, but as a failure to meet the requirement of ¶ 2 that the claims define what "the applicant regards as his invention," that is, a connection between the claim and the inventor's subjective view of the invention. *Mayhew*, 527 F.2d at 1237-39 (Baldwin, J., concurring). For those critics of the written description doctrine who object that the doctrine lacks statutory foundation, the requirement that claims correspond to what the inventor regards as his invention would seem to provide more than adequate basis. However, the Court of Customs and Patent Appeals, with considerable internal debate, seems to have established that whether the claims define what the inventor regards as his invention is a *subjective* question, answerable only by extrinsic evidence of the inventor's intent, and not the specification. See *In re Ehreich*, 590 F.2d 902, 906-07 (C.C.P.A. 1979); *id.* at 910 (Baldwin, J., concurring); *In re Cormany*, 476 F.2d 998, 1002-03 (C.C.P.A. 1973) (Baldwin, J., concurring, and Lane, J., concurring); *In re Prater*, 415 F.2d 1393 (C.C.P.A. 1969). The Federal Circuit, no doubt reluctant to endorse a validity doctrine dependent on the inventor's subjective view of his invention, has attempted to confine doctrine by the somewhat implausible notion that the statutory requirement is applicable only during prosecution but not in infringement litigation. See *Solomon v. Kimberly-Clark Corp.*, 216 F.3d 1372, 1377-79 (Fed. Cir. 2000).

169. *Auto. Techs. Int'l, Inc. v. BMW of N. Am., Inc.*, 501 F.3d 1274 (Fed. Cir. 2007).

170. *Id.* at 1283 (quoting *Genentech, Inc. v. Novo Nordisk A/S*, 108 F.3d 1361, 1366 (Fed. Cir. 1997)).

These aspects of enablement law cannot be explained in terms of a doctrine that entitles an inventor to claim everything that his specification permits one of ordinary skill in the art to make and use without undue experimentation.<sup>171</sup> They can only be explained by a doctrine that limits the inventor's rights to subject matter he actually invented, or the subject matter described in the disclosure. The *BMW* doctrine resolves the problem posed by claims like claim 4: because the novel and nonobvious subject matter falling within the scope of the claim does not appear in the (empty) disclosure, the claim would not meet the enablement requirement.

Thus the difficulties in formulating a complete system of patent law in terms of formal axioms are more than theoretical concerns. The underlying inability of the "make and use" inquiry to satisfactorily constrain claim scope has already shaped enablement law. The question, therefore, is not whether a disclosure requirement beyond "make and use" is necessary, but whether such a requirement fits better in the law of enablement or the law of written description. The *BMW* opinion emerged from the Federal Circuit without outward controversy, presumably because it framed the requirement for disclosure in terms of enablement rather than written description.<sup>172</sup> A requirement that claimed subject matter be explicitly disclosed in the specification seems to be nothing more than written description masquerading under another name.<sup>173</sup> It makes little sense to

---

171. That particular features of the invention are essential or novel might be *evidence* relevant to enablement. For example, the fact that a particular feature was not known in the prior art might be relevant to the question of whether one of ordinary skill in the art could implement that feature without undue experimentation. However, if a skilled artisan could have easily implemented a novel feature, a premise frequently invoked in the law of nonobviousness, then the novelty of the feature would be immaterial.

172. The panel of the Federal Circuit that decided *BMW* included Judge Lourie, the most vocal proponent of applying the written description requirement to originally-filed claims, and Judge Rader, its most ardent opponent. The district court had held certain of the claims invalid under written description as well. The Federal Circuit took great pains to avoid having to decide the written description issue, which no doubt would have fractured the panel's unanimity. *Id.* at 1281-82 notes 1-2.

173. In the context of a written description priority determination, the Federal Circuit in *Vas-Cath* denied that the "novel or important" aspects of the invention held any special significance. *Vas-Cath Inc. v. Mahurkar*, 935 F.2d 1555, 1565 (Fed. Cir. 1991). But the court's point was that the written description must describe *the claimed invention*, not the novel or essential aspects in particular. *See id.* (" 'The invention' is defined by the claims on appeal. "). One presumes if the inventor has sufficiently described the claimed invention, novel or essential aspects of the claims appear as a matter of course. *See also* *Cooper Cameron Corp. v. Kvaerner Oilfield Prods., Inc.*, 291 F.3d 1317, 1323 (Fed. Cir. 2002) (explaining that written description does not require claims be limited to what inventor considers essential elements of invention).

lodge in the law of enablement a requirement that the claims correspond to the invention described by the specification. To do so would distort and confuse the law of enablement, by grafting upon it an inquiry unrelated to the core notion of whether one of ordinary skill in the art could make and use the invention. The doctrine of written description already centers on the question of whether the invention defined by the claim corresponds with the invention described in the specification. Let us therefore turn to the written description requirement to see if it can provide the limits that enablement cannot.

#### IV. WRITTEN DESCRIPTION REVISITED

##### A. What Written Description is Not

###### 1. *Written Description as Possession*

Much of the confusion about the written description doctrine derives from its unfortunate formulation as “possession of the invention.”<sup>174</sup> *Vas-Cath Inc. v. Mahurkar*, a case dealing with priority issues, provides a canonical statement of the doctrine for those who disfavor applying the doctrine to originally filed claims:

The purpose of the “written description” requirement is broader than to merely explain how to “make and use”; the applicant must also convey with reasonable clarity to those skilled in the art that, as of the filing date sought, he or she was in possession of the invention. The invention is, for purposes of the “written description” inquiry, *whatever is now claimed*.<sup>175</sup>

Expressing the written description doctrine in terms of “possession” may have been sensible in the early stages of American patent law. Prior

---

174. Professor Holbrook has suggested an entirely different view of possession than the one currently embodied in written description doctrine and identifies possession of the invention as the central touchstone of patent law. See Holbrook, *supra* note 62. In Holbrook’s structure, rather than possession being the predicate for adequate disclosure, enablement is instead the predicate or best evidence of possession: if the inventor has physically created the invention or provided an enabling description of how to do so, possession of the invention has been proven. See *id.* at 147. Accordingly, there is no need to lodge a separate possession requirement in the doctrine of written description, where it has previously resided. Holbrook is correct, I think, to argue that it makes little sense to characterize possession of the invention as a written description matter. Depending on exactly what is meant by possession, a disclosure that enables the invention may indeed be the best way to demonstrate possession. We are still left, however, with the question of what it means to enable the infinite genus we call “the invention.”

175. *Vas-Cath*, 935 F.2d at 1563-64.

to the full development of the peripheral claiming system, claims were not entities that defined a category of subject matter by listing its properties. Rather, claims were drawn directly to the inventive principle itself, and established not the inventor's right to exclude, but his right to the grant of a patent.<sup>176</sup> Even after claiming assumed primary importance, "the invention" and "the claims" were distinct concepts in American patent law.<sup>177</sup> One could discuss "the invention" in terms of the inventor's physical or mental creation, entirely apart from the question of the scope of the inventor's legal rights. Under such a regime, questions of whether the inventor physically possessed an embodiment of the invention, or whether the inventor mentally possessed the idea behind the invention, are sensible questions. But once the concepts of "invention" and "claim" became essentially synonymous in patent law, the notion of "possessing the invention" became a logical impossibility except as a rephrasing of the ultimate legal conclusion.

The expression of the written description requirement as a "possession" test is a relatively recent development, rooted in decisions of the Court of Customs and Patent Appeals that distinguished written description as a separate disclosure requirement in section 112 of the current patent statutes.<sup>178</sup> However, possession cannot by itself serve as a coherent expression of the written description doctrine unless we are to take a very literal view of possession: that the inventor possesses only those physical entities he actually created, or exactly described in the specification.

Take *Vas-Cath* as an example. The patentee filed a design patent depicting a new design of a double-lumen catheter before filing a utility patent claiming the catheter. The claims of the utility patent recited a catheter with a narrowed end, having a diameter between 50% and 100% of the remainder of the catheter. Because of a question of intervening prior

---

176. For example, in a case like *Vas-Cath*, the invention might have been defined as something like *forming the inner lumen 50 to 100% of the diameter of the outer lumen*.

177. See *infra* Part IV.C.

178. See *In re Smith*, 481 F.2d 910, 915 (C.C.P.A. 1973) ("The specification as a whole conveys possession of the claimed invention as of the filing date."). Before that time, the notion of possession more frequently described whether the *public* had received the benefit of the invention, either through the disclosure of the prior art or the patentee's disclosure. See, e.g., *Evans v. Eaton*, 20 U.S. 356, 400 (1822) (object of the specification is to "put the public in possession of the invention"); *In re Arkley*, 455 F.2d 586, 590 (C.C.P.A. 1972) (Baldwin, J., concurring) (citing cases stating that the standard of anticipation is whether public was in possession of the invention).

art,<sup>179</sup> the patentee had to establish that the utility application was entitled to claim priority from the design patent, meaning that the design patent had to satisfy the written description requirement with respect to the invention defined by the claims of the later-filed utility patent. The Federal Circuit framed the question in terms of whether the drawings showed possession of the invention defined by the claims:

[T]he proper test is whether the drawings conveyed with reasonable clarity to those of ordinary skill that [the patentee] had in fact invented the catheter recited in those claims, having (among several other limitations) a return lumen diameter substantially less than 1.0 but substantially greater than 0.5 times the diameter of the combined lumens.<sup>180</sup>

Exactly what did the patentee possess or invent? The invention is what is claimed. The claims defined a genus of catheters, with an infinite variety of measurements and materials. The genus is constrained by the ratio of diameters defined by the claims, but like all patent claims it is an infinite genus. One could perhaps decide whether the patentee's drawings enabled a person of ordinary skill in the art to make and use the genus of catheters, but is it meaningful to ask whether the patentee "possessed" the genus?<sup>181</sup>

The question of "possession of the invention" is simply not a meaningful inquiry under our current claiming system. In the peripheral claiming system, "the invention" is a bundle of properties recited by the claims, defining the perimeter of the patentee's legal right to exclude. It is not syntactically sensible to ask whether an inventor "invented" or "possessed" an abstract bundle of properties defining a legally cognizable right. Inventors create ideas and things, not abstract legal entities or infinite sets of subject matter.<sup>182</sup> One can of course make the ultimate

---

179. Which was the patentee's own Canadian design patent.

180. *Vas-Cath*, 935 F.2d at 1566.

181. The problem did not arise when possession was used in the context of anticipation, because a claim is anticipated if even a single species can be shown to have been described in the prior art. The Federal Circuit in *Vas-Cath*, referring to Court of Customs and Patent Appeals precedent, noted the disconnect between the concept of disclosure in anticipation and in sufficiency of a specification. *See id.* at 1562 (citing *In re Lukach*, 442 F.2d 967 (C.C.P.A. 1971)). The *Lukach* court noted that "[t]he matter of what language constitutes sufficient description to support a claim of given breadth has been a troublesome question." *Lukach*, 442 F.2d at 969.

182. Judge Giles Rich, the author of the *Vas-Cath* opinion, recognized the disconnection between what the inventor actually did and the legal conception of the invention as defined by the claims: "Claims are frequently a far cry from what the

legal determination that the inventor “invented” or “possessed” the abstract rights defined by the claim, but one cannot ask the question as a factual premise to that ultimate legal determination.

## 2. *Written Description as a Priority Doctrine Only*

The other major misconception about the written description doctrine is that it is a novel doctrine, or that only since *Lilly* has the written description doctrine been understood to limit the scope of originally-filed claims. According to this view, the Court of Customs and Patent Appeals first invoked a separate “written description” doctrine in *In re Ruschig* in 1967, to prevent a patent applicant from adding “new matter” to a patent application in the form of amended claims to an earlier-filed patent application.<sup>183</sup> Because originally filed claims constitute their own description, written description “simply has no application to claims without priority problems,”<sup>184</sup> and *Lilly* deviated sharply from established custom by applying the doctrine to originally-filed claims.

The idea that the disclosure limits claims, independently of the enablement requirement, was not invented by the Federal Circuit in *Lilly* in 1997 or by the Court of Customs and Patent Appeals in 1967 in *Ruschig*. In William Robinson’s monumental and influential 1890 treatise on patent law, Robinson distinguished between the description of the invention and the disclosure of how to make and use the invention:

According to the statutes, the Description must contain full explanations of three different subjects: *the invention itself*; the manner of making it; and the mode of putting it into practical use,—a complete knowledge upon all these points being

---

inventor invented. In a suit, claims are construed to find out what the patentee can exclude the defendant from doing. CLAIMS ARE CONSTRUED TO DETERMINE THE SCOPE OF THE RIGHT TO EXCLUDE, regardless of what the inventor invented.” Janice M. Mueller, *A Rich Legacy*, 14 BERKELEY TECH. L.J. 895, 899 (1999) (quoting an e-mail from Judge Giles Rich); *see also* Holbrook, *supra* note 62, at 146 (“The invention is not necessarily a particular embodiment necessarily but more the idea of the invention.”).

183. *See, e.g.*, *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 977-78 (Fed. Cir. 2002) (Rader, J., dissenting from denial of rehearing en banc) (“In 1967, in *In re Ruschig*, this court’s predecessor created for the first time a new WD doctrine to enforce priority.”) (citation omitted).

184. *Id.* at 979-80 (“In 1997, for the first time, this court purported to apply WD as a general disclosure doctrine in place of enablement, rather than as a priority doctrine.”) (citing *Regents of the Univ. of Cal. v. Lilly*, 119 F.3d 1559 (Fed. Cir. 1997)).

necessary to render the invention available to the public without further experiment or exercise of inventive skill.<sup>185</sup>

One commentator has argued that this passage merely expresses the modern doctrine of enablement (i.e., how to make and use the invention).<sup>186</sup> While that interpretation might be plausible from the isolated passage, it is difficult to sustain in context. Robinson includes in the description of “the invention itself,” information not necessary to make and use the invention,<sup>187</sup> and elsewhere distinguishes between a description of the “intrinsic character” of the invention and a description necessary to practice the invention.<sup>188</sup>

Although the notion of defining the invention by the disclosure originated when American patent law did not require claims, the written description requirement is not an obsolete relic of the time before claiming. Robinson was writing at a time when claims were not only required by the Act of 1870 but had long been common practice. The notion that claims were distinct from the disclosure, and defined the boundaries of the patentee’s legal rights, was well-developed and

---

185. WILLIAM C. ROBINSON, *THE LAW OF PATENTS FOR USEFUL INVENTIONS* § 484 (1890) (emphases added). The mode of making the invention and putting it into practical use was the inventor’s best mode. *See id.* § 485 (mode of making the invention “must be the best one known to the inventor”); *id.* § 486 (“The mode explained must be the best within the knowledge of the applicant. . .”). As compared to section 112 of the current patent statute, the Act of 1870 differed primarily by requiring that “in case of a machine, [the inventor] shall explain the principle thereof, and the best mode in which he has contemplated applying that principle so as to distinguish it from other inventions. . . .” Patent Act of 1870, ch. 230, 16 Stat. 198, 201 (repealed 1952).

186. *See Janis, supra* note 49, at 63-64 n.31 (2000). Professor Janis criticized reliance on the quoted passage to support an independent written description requirement by the court in *In re Barker*, 559 F.2d 588, 592 n.5 (C.C.P.A. 1977).

187. For example, Robinson believes that the inventor must describe the state of the prior art and how the invention differs from the prior art. ROBINSON, *supra* note 185 at § 484.

188. *See id.* § 487 (“The sole object of the Description is to confer knowledge upon the public concerning the intrinsic character of the new device or process and the mode of making it available in practice. . . .”). In fairness, Robinson was not always consistent; in one passage he sequentially states the test of a complete description as a thoroughly metaphysical inquiry (whether it “embraces every essential part and attribute of the thing described”) and then as a practical inquiry (whether “a person skilled in the art could make and use the invention.”). *Id.* § 491. The metaphysical foundations of Robinson’s system are a fascinating topic in their own right.

abundantly attested by Robinson and other treatise-writers.<sup>189</sup> Robinson recognized that the claim defined the inventor's legal right, but also asserted that claims could not embrace subject matter not described by the specification, even if such subject matter was within the knowledge of one skilled in the art:

Features of the invention not delineated in the Description cannot be inserted in the Claim, even though a mechanic in endeavoring to construct or employ the invention would inevitably discover them.<sup>190</sup>

So the notion that claim limitations must be supported by the written description, notwithstanding the ability of one skilled in the art to make and use an invention with the claimed limitations, was certainly held by one of the most influential of all patent law scholars well after the development of the claim system.

Even if an independent written description requirement germinated only in 1967, it hardly has a less-esteemed pedigree than our modern enablement doctrine. The doctrine of enablement, in the sense of a requirement of section 112 that the disclosure teach how to make and use the claimed invention without undue experimentation, was itself created by the Court of Customs and Patent Appeals contemporaneously with its articulation of the written description requirement. Prior to that time, Patent Office practice was to issue "undue breadth" rejections that encompassed what we would now term enablement, written description,

---

189. A caveat to this argument is that omnibus claims, claiming the invention "substantially as described" in the disclosure, were still permissible, though the practice of claiming by essential properties was established by this time. *Id.* § 511.

190. *Id.* § 515. Robinson may not have been on the firmest ground for this statement. He cites as authority *Needham v. Washburn*, 17 F. Cas. 1276 (C.C.D. Mass. 1874) (No. 10,082), and *Kelleher v. Darling*, 14 F. Cas. 223 (C.C.D. Me. 1878) (No. 7,653). *Kelleher* concerned new matter in a reissue, though the court did reason by analogy that a claim reciting that feature in the original patent would have been invalid for failure to comply with the written description statute, notwithstanding the ability of one skilled in the art to discover the feature. *Kelleher*, 14 F. Cas. at 228. *Needham* did concern an original claim, and the court stated: "Much reason exists for holding, that the second feature of the claim is invalid, because not embraced in the description of the method or process used by the complainant, as required by the act of congress. . . ." *Needham*, 17 F. Cas at 1278. But the court declined to rest its holding entirely on that ground, because the claimed feature was the omission of a welding flux employed by the prior art process. *Id.* Both cases were decided by Justice Clifford riding the First Circuit. But the point is not whether Robinson's assertion was decisively settled law at the time; the point is that a notion of a description requirement beyond enablement and applicable to original claims was current.

and indefiniteness rejections;<sup>191</sup> only starting in about 1970 did the Court of Customs and Patent Appeals begin rigorously distinguishing the separate aspects of section 112 that we recognize today.<sup>192</sup> Thus, strictly speaking, our modern doctrines of enablement and written description were both crystallized from the undifferentiated mass of “undue breadth” at the same time. Certainly the requirement that the patent’s disclosure teach how to make and use the invention has long been central to patent law; it is nonetheless mistaken to say for purposes of interpreting our current patent statute that the written description doctrine is an abnormally novel development.

If *Lilly* was the first modern case to articulate a doctrine of written description applicable to originally-filed claims, it was not necessarily a radical innovation. Commentary contemporary with *Lilly* describes the alternative priority-policing function of written description inclusively rather than exclusively.<sup>193</sup> Moreover, certain pre-*Lilly* Federal Circuit

---

191. See Paul M. Janicke, *Patent Disclosure—Some Problems and Current Developments* (pt. 2), 52 J. PAT. & TRADEMARK OFF. SOC’Y 757 (1970) (discussing confusion in Patent Office’s “undue breadth” rejections).

192. See *id.* at 761-63; Brian P. O’Shaughnessy, *The False Inventive Genus: Developing a New Approach for Analyzing the Sufficiency of Patent Disclosure Within the Unpredictable Arts*, 7 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 147, 172, 199 (1996); *In re Mayhew*, 527 F.2d 1229, 1235 (C.C.P.A. 1976) (Baldwin, J., concurring):

Beginning in 1970, we departed from a vast line of authority which permitted the PTO to reject claims under the second paragraph of § 112 for “undue breadth.” Up to that time, examiners quite frequently determined what they felt the invention was and rejected all claims which were broader than their conception of the invention, using the second paragraph of § 112 as the statutory basis.

Before the 1952 Act, at least in “unpredictable” arts like chemistry, rejections for “undue breadth” focused on the lack of description and utility rather than the inability to make and use the invention. See generally *In re Langmuir*, 62 F.2d 93, 95 (C.C.P.A. 1932). (holding that claims may not be broader than the disclosure “in chemical cases and cases where the properties of materials are concerned.”); Samuel S. Levin, *Broader than the Disclosure in Chemical Cases*, 31 J. PAT. & TRADEMARK OFF. SOC’Y 5 (1949). An example of the “properties of materials” doctrine appears to be *In re Marshall*, 54 F.2d 421, 423 (C.C.P.A. 1932). (affirming rejection because applicant had not disclosed metals with hardness or general physical characteristics recited by application’s claims).

193. See IRAH H. DONNER, *PATENT PROSECUTION: PRACTICE AND PROCEDURE BEFORE THE U.S. PATENT OFFICE* (1996) at 503 (citing amendment as example of general principle that claims must encompass subject matter disclosed in the description); *id.* at 500-01 (characterizing *In re Fisher*, 427 F.2d 833 (C.C.P.A. 1970) as a written

precedent seems hard to reconcile with the notion that original claims require no disclosure beyond enablement. If claims themselves satisfactorily disclose the subject matter they encompass, then *prior art patents* necessarily must disclose all enabled subject matter falling within their claims. This appears not to have been the case.<sup>194</sup> Moreover, prior to *Lilly*, the Federal Circuit expressed the “priority-policing” test as whether later claims would have been rejected if filed with the original application.<sup>195</sup> If there is no written description requirement applicable to originally filed claims, then such an expression is absurd: originally-filed claims cannot be rejected for lack of written description. Yet the Federal Circuit expressed the doctrine in those terms, implying either very careless word choice or recognition that original claims do not ipso facto constitute their own description.<sup>196</sup>

## B. What Written Description Is: A Doctrine of Definition

Written description, then, is not about possession or priority alone. It is instead a general doctrine of disclosure. Critics have maintained that written description as articulated by *Lilly* is a special biotechnology rule

---

description case); O’Shaughnessy, *supra* note 192, at 180-81 (stating that written description compliance issues “might” arise from claim amendments).

194. See *In re Benno*, 768 F.2d 1340, 1346 (Fed. Cir. 1985) (“The scope of a patent’s claims determines what infringes the patent; it is no measure of what it discloses.”). In *Benno*, Judge Rich reasoned by analogy that Samuel Morse’s infamously overbroad eighth claim would have anticipated the Telex if claims disclosed everything within their scope. On this analogy alone one could argue that patent claims only fail to disclose *non-enabled* subject matter within their scope. However, *Benno* was a simple mechanical case, and there was no allegation that the subject matter alleged to be within the prior art was not enabled by the prior art patent.

195. See *U.S. Steel Corp. v. Phillips Petroleum Co.*, 865 F.2d 1247, 1251 (Fed. Cir. 1989) (explaining that test for whether claim in continuation application was entitled to benefit of filing date of parent application is whether claim would have been rejected for lack of support if filed with parent application). It is true that *U.S. Steel* discusses enablement and written description together under the question of “lack of support.” However, Judge Markey in *U.S. Steel* does not suggest that lack of enablement would be the only grounds for rejecting the claim if filed with the parent application, and discusses at length a written description case. See *id.* (discussing *In re Koller*, 613 F.2d 819 (C.C.P.A. 1977)). Judge Markey, it should be noted, was at least initially opposed to the notion of a separate written description doctrine. See *In re Barker*, 559 F.2d 588, 594-95 (C.C.P.A. 1977) (Markey, C.J., “heartily” dissenting).

196. Even if *Lilly* had instituted for the first time a disclosure doctrine applicable to originally-filed claims, it is peculiar to criticize it on those grounds. Most statutory patent law is codification of judicial innovations, and modern patent law retains extremely significant common-law doctrines having no statutory basis, such as the Doctrine of Equivalents and the doctrine of inequitable conduct.

requiring nucleotide-by-nucleotide disclosure of DNA molecules, or a rigid rule limiting patentees to disclosed embodiments alone.<sup>197</sup> Such interpretations miss the point of *Lilly*. The written description requirement is a general requirement that the applicant for a patent *define the invention* according to traditional principles of logic. Consider the language of *Lilly*:

A description of a genus of cDNAs may be achieved by means of a *recitation of a representative number* of cDNAs, defined by nucleotide sequence, falling within the scope of the genus or of a *recitation of structural features common to the members of the genus*, which features constitute a substantial portion of the genus.<sup>198</sup>

Most analyses and criticisms of this key language from *Lilly* have focused on the Federal Circuit's demand for structure or sequence information to satisfy the written description requirement.<sup>199</sup> Such analyses miss the more important point. *Lilly*'s importance does not lie in how one describes a DNA molecule, but in how one describes a *genus*. The Federal Circuit demanded that the claimed genus, i.e., a genus of DNA molecules, be described either by disclosure of a representative number of species in the genus, or by disclosure of properties that are common to members of the genus. These two modes correspond precisely to the two modes of *definition* articulated in formal logic. Recitation of the features or properties of a genus corresponds to definition by intension, or definition *per genus et differentiam*. In this classical mode of definition, a thing is defined by specifying the proximate genus to which it belongs, and those properties which differentiate it from other members of the

---

197. See, e.g., Mueller, *supra* note 53, at 651 (arguing that *Lilly*'s "per se rule that a claim to a cDNA must be described in terms of its specific nucleotide sequence" runs contrary to tradition that the patent system "provided more in terms of patent scope than merely those embodiments expressly disclosed by the inventor in her application.").

198. Regents of the Univ. of Cal. v. Lilly, 119 F.3d 1559, 1569 (Fed. Cir. 1997) (emphasis added). A "cDNA," or complementary DNA molecule, is a synthetic DNA molecule produced by reverse transcription of a messenger RNA encoding a protein such as human insulin.

199. See, e.g., Holman, *supra* note 48, at 19 n.89 (collecting structural criticisms); Dan L. Burk & Mark A. Lemley, *Biotechnology's Uncertainty Principle*, 54 CASE W. RES. L. REV. 691, 697-98 (2004) (stating that the written description requirement demands precise sequence data); see also Amgen Inc. v. Hoechst Marion Roussel, Inc., 314 F.3d 1313, 1332 (Fed. Cir. 2003) (suggesting that *Lilly* was inapplicable because challenged patent disclosed a DNA sequence).

genus.<sup>200</sup> The alternative mode of description suggested by *Lilly*, enumeration of a representative number of members of the genus, corresponds to definition by extension, or definition by type. It proceeds by designating some individual or group of individuals as central or typical members of the genus and determining membership in the genus by degree of resemblance.<sup>201</sup>

Yet every claim is in the end a genus claim.<sup>202</sup> Therefore, if *Lilly* provides a method to define and describe a genus, then *Lilly* provides a method to define and describe any claim. The Federal Circuit clearly understood itself to be promulgating a general doctrine of definition in *Lilly*. Holding that the inventors had not sufficiently described the genus of DNA molecules encoding mammalian insulin by the phrase “mammalian insulin cDNA,” the standard the court employed was one of definition:

It does not specifically *define* any of the genes that fall within its definition. It does not *define* any structural features commonly possessed by members of the genus that distinguish them from others. One skilled in the art therefore cannot, as one can do with a fully described genus, visualize or recognize the identity of the members of the genus.<sup>203</sup>

---

200. See, e.g., PETER COFFEY, THE SCIENCE OF LOGIC 94 (1912) (“In order, therefore, to define any object of thought, we must find out and indicated its *proximate genus*—the next highest class into which it *naturally* falls—and the attribute or group of attributes which distinguishes it from other *cognate species* of the same *genus*.”) (citation omitted).

201. See *id.* at 98. In linguistics the notion of a family gathered around a type is often attributed to Wittgenstein, but the idea was in circulation well before. See JOHN NEVILLE KEYNES, STUDIES AND EXERCISES IN FORMAL LOGIC 34 (1884). Note that in some respects the intensional and extensional modes of definition also correspond to the peripheral and central modes of patent claiming. The topic of definition in patent law and its implication for the structure and interpretation of claims deserves far fuller treatment than can be accorded here.

202. See *supra* Part III.A.1.

203. *Lilly*, 119 F.3d at 1568 (emphases added). The theme of *mental* process—visualization or recognition—in the doctrine of written description seems to have derived from *Amgen, Inc. v. Chugai Pharm. Co.*, 927 F.2d 1200 (Fed. Cir. 1991). *Amgen* was an infringement action in which the question of priority of invention for a DNA molecule encoding erythropoietin was in dispute. As set forth by 35 U.S.C. § 102(g), an inventor may initially establish priority by demonstrating earlier conception of the invention. Conception in patent law is an entirely mental act, though it must be demonstrated by some objective disclosure. The court in *Amgen* held that an inventor who had failed “to envision the detailed chemical structure of the gene so as to distinguish it from other materials” could not establish conception until he had actually isolated the claimed DNA

By expressing the written description doctrine as a doctrine of definition, the Federal Circuit provided, at least in theory, both a coherent rationale and a coherent test for application of the written description doctrine. Had *Lilly*'s lead been followed, the true role of the written description doctrine, and how it differs from that of enablement, might have become clear.

### C. Losing the Path

Unfortunately, since *Lilly*, the written description doctrine has gravitated back to the quixotic notion of "possession."<sup>204</sup> This trend is especially evident in the evolution of the PTO's Guidelines<sup>205</sup> for assessing patent applications for compliance with the written description requirement. The initial Guidelines, issued in response to *Lilly*, explained the doctrine of written description in terms of possession. However, the Guidelines also framed written description as a doctrine of definition as articulated in *Lilly*. For generic claims, the Guidelines suggested that the specification must allow "one skilled in the art [to] readily envision a sufficient number of members of the claimed genus to provide written description support for the genus."<sup>206</sup> Thus, the written description inquiry was to proceed by asking whether the inventor had conveyed enough information for one of ordinary skill in the art to define the genus by its intension. Likewise, satisfying the requirement by disclosure of common characteristics was judged by whether one of skill in the art could "reasonably predict sufficient identifying characteristics of the other members of the genus and, thus establish possession of the genus."<sup>207</sup>

In subsequent revisions of the Guidelines, the PTO eliminated the definitional aspect of the written description doctrine and focused entirely on the notion of possession. The result was an essentially tautological expression of the doctrine:

---

molecule. *Fiers v. Revel*, 984 F.3d 1164, 1168-69 (Fed. Cir. 1993) (discussing *Amgen*, 927 F.2d 1200).

204. The notion of possession may well have utility in other contexts; the point is that possession is not sensible in the context of written description.

205. See Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112, 1, "Written Description" Requirement, 66 Fed. Reg. 1099, 1104 (Jan. 5, 2001) [hereinafter Guidelines].

206. Request for Comments on Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. 112 1 "Written Description" Requirement, 63 Fed. Reg. 32,639, 32,641 (June 15, 1998) (citing *Regents of the Univ. of Cal. v. Lilly*, 119 F.3d 1559 (Fed. Cir. 1997)).

207. *Id.* at 32, 642.

To satisfy the written description doctrine, a patent specification must describe the claimed invention in sufficient detail that one skilled in the art can reasonably conclude that the inventor had possession of the claimed invention. An applicant shows possession of the claimed invention by describing the claimed invention with all of its limitations.<sup>208</sup>

In other words, one describes the invention by showing possession, and one shows possession by describing the invention. With respect to genus claims, the revised Guidelines discarded the notion that the specification must convey enough information to permit one of skill in the art to envision or predict characteristics of members of the genus. Rather, the question was whether “the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus in view of the species disclosed.”<sup>209</sup> It may be difficult enough to understand how to “possess” a genus of inventions, which are at least discrete objects or processes. It seems even more difficult to understand how to “possess” attributes or features of the genus. The patentee may possess a thing such as a red ball, and perhaps somehow a genus of red balls, but by what means do we assess whether he possesses “red”?<sup>210</sup>

Further muddying the nature of the doctrine, the revised Guidelines added that possession could be shown if the specification described an actual reduction to practice, or sufficiently disclosed to indicate that the invention was “ready for patenting.”<sup>211</sup> These standards were imported from the Supreme Court’s opinion *Pfaff v. Wells Electronics, Inc.*, a case that decided at what point of development inventions could be considered “on sale” for purposes of novelty.<sup>212</sup> *Pfaff* held that an invention was

---

208. Revised Interim Guidelines for Examination of Patent Applications Under the 35 U.S.C. § 112, P 1 “Written Description” Requirement, 64 Fed. Reg. 71427, 71434 (Dec. 21, 1999) [hereinafter Revised Guidelines]. The final revision of the Guidelines added that “describing the claimed invention with all of its limitations” could be achieved by “using such descriptive means as words, structures, figures, diagrams and formulas that fully set forth the claimed invention.” Guidelines, 64 Fed. Reg. at 1104.

209. Revised Guidelines, 64 Fed. Reg. at 71436; Guidelines, 66 Fed. Reg. at 1106.

210. By framing the inquiry as whether the inventor showed possession of *properties* of the genus, the Patent Office seems to have committed itself to a form of metaphysical realism, the position descending from Plato that universals have an independent and objective existence outside of the particulars that instantiate them.

211. Revised Guidelines, 64 Fed. Reg. at 71429 (adopting suggestions to incorporate *Pfaff* analysis).

212. See *Pfaff v. Wells Elecs., Inc.*, 525 U.S. 55 (1998).

“ready for patenting” if it was actually reduced to practice, or if the inventor had prepared drawings or other descriptive material sufficient to enable practice of the invention.<sup>213</sup> The Guidelines incorporated the *Pfaff* standards in response to comments suggesting that the *Pfaff* analysis was pertinent to the written description analysis.<sup>214</sup> Yet the relevance of the *Pfaff* standard to the written description doctrine is hard to fathom. In order to encourage prompt filing of patent applications, the *Pfaff* standard was set to trigger the on-sale bar no earlier than the point when the inventor could patent the invention; if the invention was not yet ready for patenting then the law ought not to penalize the inventor. However, the *Pfaff* standard seems intended to identify the point in time at which the inventor *could* describe the invention to satisfy the standard of section 112, not to determine whether the inventor *did* describe the invention within the meaning of section 112.<sup>215</sup> The logic of incorporating the *Pfaff* standard appears to rest on the notion of written description as possession: if reduction to practice shows possession of the invention, and if the Supreme Court mentioned reduction to practice as an alternative to “drawings or other descriptions,” then a reduction to practice is a form of description. But such reasoning, apart from ignoring the Supreme Court’s emphasis on an *enabling* disclosure, commits an elementary error of logic by assuming that if both description and reduction to practice indicate “ready for patenting,” then reduction to practice equates to description.<sup>216</sup>

More importantly, the *Pfaff* inquiry, a tool for assessing whether the statutory bar to patenting has been triggered, is unhelpful for a doctrine of claim scope. As explained in Part II, the statutory bar (like other provisions of section 102) is triggered by *any* overlap between the set defined by the claim and the set of prior art. If any one embodiment of the claimed invention was sold and was ready for patenting under the *Pfaff*

---

213. *See id.* at 67.

214. *See* Revised Interim Guidelines, 64 Fed. Reg. at 71429.

215. Before the approach of the Guidelines, Professor Holbrook suggested that the “ready for patenting” inquiry explicitly incorporate an enablement analysis. *See* Timothy R. Holbrook, *The More Things Change, the More They Stay the Same: Implications of Pfaff v. Wells Electronics, Inc. and the Quest for Predictability in the On-Sale Bar*, 15 BERKELEY TECH. L.J. 933, 968-74 (2000). I have no quarrel with incorporating disclosure notions into the “ready for patenting” analysis; what seems awkward is the reverse.

216. The logic appears to be of the form: Socrates is a Greek. Plato is also a Greek. Therefore, Plato is Socrates. Ironically, one commentator argues that *Pfaff* is incompatible with the existence of a separate written description requirement. *See* Limin Zheng, Note, *Purdue Pharma L.P. v. Faulding Inc.*, 17 BERKELEY TECH. L.J. 95, 109 (2002).

standard, then the claim is invalid under section 102(b). In contrast, a mere intersection between what was described and what was claimed does not satisfy the disclosure requirements of section 112. Satisfaction of the *Pfaff* standard may show that the inventor described *something* within the scope of the claims, but that sheds little light on whether the inventor described the set of all things encompassed by the claim.

The Federal Circuit has endorsed the PTO's Guidelines,<sup>217</sup> while simultaneously emphasizing that a patent specification may demonstrate "possession" but still fail to provide a written description of the invention.<sup>218</sup> However, the court also phrased the written description inquiry as whether the applicant has "demonstrate[d] possession of the generic scope of the claims."<sup>219</sup> In subsequent opinions, the court has continued to assess the adequacy of support for a generic claim by asking whether the written description demonstrates that "the patentee possessed the full scope of the invention."<sup>220</sup> The possession inquiry, at least as currently constituted, cannot support or explain how written description functions as a limitation on claim scope.

---

217. *See Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 964 (Fed. Cir. 2002) ("We are persuaded by the Guidelines on this point and adopt the PTO's applicable standard for determining compliance with the written description requirement."). Formally, the court's endorsement of the Guidelines might be read to extend only to the point addressed in that section of the *Enzo* opinion—the use of correlated structural and functional properties to describe claimed subject matter. *Id.* However, the *Enzo* court's remand instructions explicitly directed the district court to judge the broader questions whether the genus was adequately described according to the Guidelines. *Id.* at 967-68.

218. *Id.* at 969 (explaining that possession is ancillary to the statutory requirement of written description).

219. *Id.* at 966.

220. *See LizardTech, Inc. v. Earth Res. Mapping, Inc.*, 424 F.3d 1336, 1345 (2005); *see also Univ. of Rochester v. G.D. Searle & Co.*, 375 F.3d 1303, 1307 (Fed. Cir. 2004) (Lourie, J., concurring in the denial of rehearing en banc) ("[T]he issue may still remain in a given case, especially with regard to generic claims, whether an original claim conveys that one has possession of and thus has invented species sufficient to constitute the genus."). However, there are encouraging signs that at least the Board of Patent Appeals and Interferences has moved away from reliance on possession. For example, in *Ex parte Porro*, No. 2008-0184, 2008 WL 2259960 (B.P.A.I. Mar. 11, 2008), the Board assessed the compliance of claims to a method of making vitamin C in genetically engineered yeast. The Board's analysis focused on whether one of ordinary skill in the art could visualize or recognize members of the claimed genus; "possession" was invoked only as the ultimate legal conclusion.

#### D. Anchoring the Definitional Hierarchy

Once we recognize written description as a method of logical definition, then its function in determining claim scope becomes clear. The system of definition in classical logic postulated hierarchical trees or chains of categories, each category being differentiated from the category above it by some necessary and essential characteristic property.<sup>221</sup> In the classical example, a human is defined and distinguished from all other things by successively narrower genera, until we reach the level of the individual person:

- Objects
- Material Objects
- Living Objects
- Animals
- Humans
- (Socrates)

We could construct a similar chain focusing on the rat insulin DNA molecule at issue in *Lilly*:

- DNA
- Vertebrate DNA
- Vertebrate insulin DNA
- Mammalian insulin DNA
- Rat insulin DNA
- (Particular variant of rat insulin DNA)

The written description requirement anchors claim scope to a particular level within the chain of definition. The inventor who has discovered and disclosed only rat insulin DNA is not entitled to claim higher categories, such as “vertebrate insulin DNA,” because the inventor has defined the genus neither by properties that distinguish it from other genera, nor by a set of types by which the genus can be recognized by degree of resemblance. Nor is the inventor entitled to claim a particular variant of rat insulin DNA, unless the differentia—in this case, the structural distinction between the variant and the type—of that species can be derived from the inventor’s disclosure.

---

221. This scheme is generally known as the Tree of Porphyry, as it was set out explicitly in Porphyry’s *Isagogue*, a commentary on Aristotle’s *Categories*. See COFFEY, *supra* note 200, at 78-9.

By anchoring claim scope within the hierarchy of definitional genera, written description deals directly with the question of claim scope and has the potential to resolve formal questions of claim scope in a way that enablement cannot. If written description was necessary solely to rationalize the formal structure of claiming, that would be little reason to maintain the doctrine. But of course the question of properly locating a patent's scope within the definitional hierarchy is critical to the policies of the patent system. In its traditional role in chemical practice, the written description doctrine prevented the inventor of a broader genus from reaching down the definitional chain to claim enabled but undisclosed members of that genus. Such function is necessary if we are to preserve the incentive for later inventors to develop improved or otherwise favorable members of the known genus.<sup>222</sup> If description of the genus necessarily described every member of the genus, patents on favorable members of the genus would either be unobtainable or the property of the inventor of the genus.<sup>223</sup> To use the example based on *Lilly*, suppose that particular synthetic variations of the rat insulin gene have properties making them more valuable than ordinary rat insulin. By forbidding the patentee who has disclosed the structure of rat insulin DNA from claiming those improved variants, inventors can patent subsequently discovered improved variants.<sup>224</sup>

Likewise, written description limits the inventor from reaching too far upwards on the definitional chain. This, according to some commentators, is the novel and heretical aspect of *Lilly*. However, once the role of written description is seen as properly locating claim scope in the definitional hierarchy, restrictions on upward reach seem as reasonable as restrictions on downward reach. Again, the goal must be to preserve incentives for

---

222. Such members of the genus must still be nonobvious over the genus in order to be patentable.

223. If the generic disclosure sufficed to disclose the members of the genus, then the patentee would be entitled to claim them as they were described in his original specification. If not claimed by the patentee, they would be thereafter be unpatentable, having been made "prior art" by the provisions of section 102. This balance further illustrates the necessity of the written description doctrine. Given that all things enabled by a disclosure do not become prior art, it would be curious to conclude that a species enabled but not described by the specification is always disclosed for purposes of section 112, the patentee's right to claim it, but not for purposes of section 102, a subsequent inventor's right to claim it.

224. Such variants would still infringe the generic inventor's patent, of course. Nonetheless, patents on a favorable embodiment are valuable, though less valuable than they would be in the absence of the generic patent.

later inventors but the doctrine has more bite: no one<sup>225</sup> is entitled to a patent on the broader genus such as “mammalian insulin DNA,” though patents on cognate genera (such as “human insulin DNA”) are still available. If the original inventor has indeed enabled the broader genus,<sup>226</sup> there would seem to be little difficulty in the inventor accumulating the information necessary to define the genus under the written description doctrine. The argument that the inventor who has enabled the broader claim ought to be entitled to it regardless of his ability to describe it seems to carry the seeds of its own destruction: if accumulating the information needed to describe the genus is difficult and time-consuming, though “enabled,” then perhaps enablement is doing a poor job of implementing the quid pro quo of the patent system.<sup>227</sup>

Phrasing the scope problem in terms of the definitional hierarchy makes the *Lilly* upward-limiting aspect of the written description doctrine a natural extension of the traditional downward-limiting aspect of the doctrine. Likewise, once we understand that all claims are genus claims, we can understand that (1) the doctrine of written description is applicable to all categories of inventions, not just biotechnological inventions, and (2) satisfaction of the requirement is usually a matter of course for categories like simple mechanical inventions. The genus of “chairs with four legs” is much larger and more variable than a genus like “mammalian insulin DNA.”<sup>228</sup> Yet, if presented with a disclosure embodying the novel

---

225. The disclosure of a species anticipates the broader genus. This doctrine is necessary to prevent the generic inventor from removing species from the public domain.

226. Assuming we have a coherent way to answer this question.

227. This argument in some ways resembles the one made by Kitch’s prospect theory: that even once the point of patentability has been reached (here, enablement of the genus) significant investments may be necessary to identify commercially useful embodiments of the genus. Edmund W. Kitch, *The Nature and Function of the Patent System*, 20 J.L. & ECON. 265 (1977). Kitch’s solution was to grant the original inventor a broad patent, enabling him to coordinate the process of commercial development. Critics who doubt the ability of patentees to coordinate development would deny the broad patent to maximize incentives for others to develop commercial embodiments. If those critics are correct, then the inventor who ‘enables a genus’ should not necessarily obtain a patent covering the entire genus.

228. We may postulate an infinite variety of dimensions, materials, decorative qualities, etc. for chairs. In contrast, there are 4,000 to 5,000 mammalian species; if the insulin gene is different in each species then there are at most 4,000 to 5,000 mammalian insulin genes. Of course one can postulate an infinite number of DNA molecules that encode a mammalian insulin polypeptide, by adding non-coding sequences or varying the sequence of the insulin polypeptide, but such changes in DNA sequence are qualitatively much simpler than the transformations that can be imagined of chairs.

inventive idea of placing four legs on a seating surface, one of ordinary skill in the art would readily grasp the concept of the genus of all chairs with four legs, and envision any given member of the genus.<sup>229</sup> Viewed in this light, the application of the written description doctrine to “ordinary” inventions should be uncontroversial.

Take the case of *Gentry Gallery Inc. v. Berkline Corp.*,<sup>230</sup> viewed by many commentators as a prime example of the written description requirement run amok. The disclosure of the patent in *Gentry* described a sofa with two reclining seats, controls for the recliners being located on a console between the seats. The Federal Circuit held that a claim which did not fix the location of the controls on the console was invalid for failure to satisfy the written description requirement, because the patent disclosed only controls mounted on the console.<sup>231</sup>

Had the court treated the problem as one of broader and narrower genera, the claims might have been held valid. The question would have been whether the disclosure of a narrower genus, sofas with controls mounted on a console, would have permitted one of ordinary skill in the art to envision the broader genus of sofas with controls mounted elsewhere on the sofa. If locating the controls on the console was not necessary for the function of the sofa, then one of ordinary skill in the art would likely have been able to envision a genus of sofas with controls located elsewhere than the console. If the sofa arts are predictable arts, then based on the properties of the sofa with console-mounted controls, one skilled in the sofa arts would likely have been able to predict the properties of sofas belonging to the broader genus. *Gentry Gallery* therefore represents not an aberrant application of a biotechnology doctrine to a mechanical patent, but a failure to recognize that the principles of genus and species explicit in chemical and biotechnological practice are inherent in every category of invention.

---

229. This is not the same as being able to make and use any member of the genus. One of skill in the art can envision a chair made of neutronium, and perhaps predict its properties, but cannot make and use one. Even if the fashioning of furniture from neutronium is known in the art, actually making a chair from neutronium may require unforeseeable advancements that by definition could not have been envisioned by one of ordinary skill in the art.

230. *Gentry Gallery, Inc. v. Berkline Corp.*, 134 F.3d 1473 (Fed. Cir. 1998).

231. *Id.* at 1480.

## V. THE SIGNIFICANCE OF DEFINITIONAL INFORMATION

It would be overly ambitious, in this Article, to attempt to provide a comprehensive methodology for assessing the sufficiency of definitional information, or circumscribe precisely the proper spheres of written description and enablement.<sup>232</sup> Such questions as whether the patentee has provided sufficient representative members for a definition by extension, or what constitutes degree of resemblance in a particular instance, will be complex factual inquiries in some cases. There is no escaping such difficulties in patent law. But at least treating the question as one of definition provides a principled intellectual framework to decide questions of genus and scope, rather than an ad hoc approach. This final section briefly considers the significance of a definitional information requirement, why it ought to be lodged in the doctrine of written description, and how treating written description as a doctrine of definition changes our view of other facets of patent law.

### A. Consequences of an Obligation for Definitional Information

At the most general level, the patent system's disclosure doctrines control the balance between initial and follow-on inventors by dictating how broadly an earlier inventor may claim under a given disclosure.<sup>233</sup> Preferences for that set-point depend largely on whether one views broad or early patent rights as conducive to innovation.<sup>234</sup> A patent system might set that balance via enablement, written description, or other doctrines. For the purposes of this Article, the question is not where that balance should be set but what doctrinal theories and tools are best suited to achieve that result.<sup>235</sup>

The enablement inquiry tends to be fact-intensive, requiring evidence of what one of ordinary skill in the art could or could not accomplish with

---

232. In "easy" disclosure cases the same information will likely satisfy the definitional requirement and enable at least some of the subject matter of the patent.

233. See Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1024-46 (1989) (discussing theories of innovation and patent scope).

234. See generally *id.* (analyzing effect of broad patents on biomedical research).

235. Burk and Lemley favor increased use of judicial "policy levers" to adjust a statutorily uniform patent law to the technologically heterogeneous innovation economy. See generally Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575 (2003). They identify the written description doctrine as an existing policy lever, see *id.* at 1652-54, though they disagree with how the Federal Circuit has pulled it. See *id.* at 1682-83; *id.* at 1688-89 (criticizing Federal Circuit application of disclosure requirements in biotechnology and software cases).

certain efforts given the state of the art. The inquiry may require extensive expert testimony and may not be amenable to early judicial intervention. In contrast, written description is a question of what the patent specification discloses: does the text of the patent disclose the invention defined by the claims? No less than most other inquiries in patent law, this question is resolved from the perspective of one of ordinary skill in the art. Nevertheless, the underlying question in written description, what information is conveyed by the patent specification, may be more capable of judicial resolution than questions about the behavior or thought processes of technological artisans.<sup>236</sup> To the extent that we desire more judicial control over patent scope and desire such control not be confined by the particular testimony adduced in a given case,<sup>237</sup> written description may be a more appealing doctrine than enablement.

---

236. This argument is not supported by the current standards of appellate review. Enablement and nonobviousness are treated as issues of law with underlying factual components, whereas written description is treated as an issue of fact. However, in practice, the nature of the inquiry—content of a text versus the mindset or capability of one of ordinary skill in the art—may well be more important than the nominal standard of appellate review. The current “factual” state of written description is in any event curious. It derives from the Court of Customs and Patent Appeals’ statement in *In re Ruschig*:

[W]e doubt that the rejection is truly based on section 112, at least on the parts relied on by appellants. If based on section 112, it is on the requirement thereof that “The specification shall contain a written description of the invention.” . . . We have a specification which describes appellants’ invention. The issue here is in no wise a question of its compliance with section 112, it is a question of fact: Is the compound of claim 13 described therein?

*In re Ruschig*, 379 F.2d 990, 995-96 (C.C.P.A. 1967). *Ruschig* was subsequently cited for the proposition that written description issues were questions of fact. *See In re Wertheim*, 541 F.2d 257 (C.C.P.A. 1976) (quoting *Ruschig* for proposition that written description requirement is issue of fact). Moreover, the Court of Customs and Patent Appeals was not particularly deferential to the PTO on issues of fact. *See In re Zurko*, 142 F.3d 1447, 1454-55 (Fed. Cir. 1998) (discussing the Court of Customs and Patent Appeals review of Patent Office decisions), *rev’d*, 527 U.S. 150 (1999).

237. Burk and Lemley argue that judges should have substantial discretion to adjust policy levers. *See Burk & Lemley, supra* note 235, at 1668. They seem to suggest that industry-tailored policy discretion should be informed by legal and economic scholarship rather than emphasis on the facts of the particular case. *See id.* at 1671. Certainly on the question of disclosure the Federal Circuit has explicitly eschewed policy-based rulings and chosen instead to decide cases based on the evidence before the district court and the standard of review. *See, e.g., Amgen Inc. v. Hoechst Marion Roussel, Inc.*, 314 F.3d 1313, 1337 (Fed. Cir. 2003) (“The dissent, however, does not directly challenge the court’s factual findings, nor does it mention the decisions relied on by the district court. Instead, it finds fault in the absence of discussion of [section 112 precedent], and makes

The conception of written description this Article advances emphasizes the significance of definitional information provided by the patentee. Setting aside for a moment the question of which doctrine demands such information, we may observe that patent theorists have already recognized the importance of definitional information, even if it has long been overshadowed by other aspects of patent disclosure. Edmund Kitch, in his seminal work on prospect theory, maintained that the patent system was superior to a government auction system because the patent system provided incentives for private parties to identify and define claims.<sup>238</sup> The premise of Kitch's argument is that definitional information is costly and does not appear of its own accord. Indeed, with Kitch's emphasis on the patent as a prospect for future development, the primary role of the patent specification was to provide definitional information. Believing that patentees had independent incentives to disclose enabling information and better ways to do so than within the formal constraints of the patent document, Kitch argued, "The purpose of the description in the patent is not to disclose the commercially relevant technology, but to provide a context in which the legal limits of the claim acquire meaning."<sup>239</sup> Likewise, more recent theories of the patent system that look beyond the incentive or reward functions of the patent also emphasize the importance of definitional information, rather than technological disclosure *per se*, in setting the scope of the patentee's rights.<sup>240</sup>

---

broader arguments seemingly based upon policy considerations."). It is somewhat difficult to reconcile this sort of reasoning with Burk and Lemley's identification of "more intrusive appellate review" as the reason for lack of policy direction. *See* Burk & Lemley, *supra* note 235, at 1671. In any event, the argument for increased judicial discretion seems to support the existence of an independent written description requirement, if only as an additional lever to adjust aspects of the patent system that enablement has difficulty controlling.

238. Kitch, *supra* note 227, at 265, 266 n.4.

239. *Id.* at 287. Although the patentee's disclosure of his or her technological advance lies at the heart of many theories of the patent system, there is little known about whether such disclosure is significant for technological progress. *See* Rebecca S. Eisenberg, *Analyze This: A Law and Economics Agenda for the Patent System*, 53 VAND. L. REV. 2081, 2093-94 (2000) (noting lack of data on disclosure functions of patent specification); *see also* Holbrook, *supra* note 62, at 131-46 (arguing against "teaching" function of enablement).

240. *See* F. Scott Kieff, *The Case for Registering Patents and the Law and Economics of Present Patent-Obtaining Rules*, 45 B.C. L. REV. 55, 99 (2003) (arguing that independent written description requirement decreases social costs under a registration theory).

The economic significance of definitional information is also inherent in Meurer and Nard's game-theoretic analysis of the Doctrine of Equivalents.<sup>241</sup> Meurer and Nard suppose that, for a certain level of investment, an inventor discovers a quantum of information that suffices to *enable* a broad set of embodiments, but only permits the inventor to *claim* a more limited set of embodiments. In the model, the inventor chooses between investing further "refining" so as to be able to claim the broader set, or resting upon a claim to the narrower set and relying upon the DOE for the rest.<sup>242</sup> The analysis therefore assumes that the initial investment provides enough information to enable a genus, but not to claim it literally.

In Meurer and Nard's treatment, the additional "refinement costs" that the patentee may spend to expand his literal claim scope are directed towards identifying and claiming the subject matter enabled by the patent disclosure.<sup>243</sup> These costs might represent the time necessary to imagine additional embodiments of the invention, and to formulate more expansive claim language to cover additional embodiments.<sup>244</sup> However, the analysis is equally valid if we view "refinement" as the process of developing more definitional information necessary to describe claims of broader scope.<sup>245</sup> What the inventor may claim literally on a given disclosure is determined by the disclosure doctrines of section 112. Thus an inability to claim literally the broader set of embodiments might reflect inability to formulate appropriate claim language without further investment, or it might reflect an inability to satisfy the disclosure obligations of section 112 with respect to known claim language or embodiments. If additional claim drafting and additional disclosure are interchangeable from the perspective of refinement costs, then the economic consequences of a weak written description regime versus a strong written description

---

241. Michael J. Meurer & Craig Allen Nard, *Invention, Refinement and Patent Claim Scope: A New Perspective on the Doctrine of Equivalents*, 93 GEO. L.J. 1947 (2005).

242. *See id.* at 1983-84 (setting forth model in which investment *W* enables embodiment set *F* but only permits inventor to literally claim set *E*).

243. *See id.* at 1984.

244. *See id.* at 1985 (assuming inventor may literally claim larger set after "mentally identifying the embodiments" in the set). One suspects that a large portion of this cost would end up being the cost of patent attorneys.

245. The Court of Customs and Patent Appeals seems to have been cognizant of this process early on. *See In re Clarke*, 356 F.2d 987, 992 (C.C.P.A. 1966) (describing situation where inventor has recognized that invention was generic, and "was endeavoring to determine by exercise of reasonable diligence the precise scope of the invention").

regime generally follow those of a strong DOE versus weak DOE regime.<sup>246</sup>

This parallelism should hold beyond the costs and benefits of permitting patentees to assert patents against broader ranges of subject matter: just as a strong DOE regime tends to erode the notice function of patent claims by permitting a patentee to reach beyond the literal boundaries of the patent, a regime with weak obligations of definitional information is likely to yield less certain patent claims.<sup>247</sup> Defining the true genus of the invention not only anchors the patent right at a particular level of generality, but also serves to more precisely define the boundaries of the patent.

Given that definitional information is real and costly, how important is it to demand such information from the patentee? It is perhaps very important. Certainty in the boundary of the patent right is central to the purposes of the patent system,<sup>248</sup> and recent critiques attribute many of the problems of our current system to our inability to precisely define the scope of patent claims.<sup>249</sup> We might therefore regard an obligation to provide definitional information as part and parcel of the patent quid pro quo, along with the more familiar obligation to disclose the technological details of the invention's advance.

---

246. Meurer and Nard show that the behavioral consequences of these regimes depend on the magnitude of the cost of refinement versus the cost of invention, and the returns expected from the monopoly or duopoly returns expected from the broader or narrower scope. *See* Meurer & Nard, *supra* note 241.

247. The question of the interplay between disclosure and DOE is more complex than can be fairly treated in this Article. For a general treatment, though focused on the doctrine of prosecution history estoppel, see R. Polk Wagner, *Reconsidering Estoppel: Patent Administration and the Failure of Festo*, 151 U. PA. L. REV. 159 (2002). Two observations on the DOE are pertinent here. First, the problems of disclosure identified in this Article emanate from the peripheral claim system. To the extent that the DOE represents a retreat from the primacy of peripheral claiming, one response to the issues presented here might be to admit that the peripheral claim system has done all that we can reasonably ask of it, and to rely more heavily on the DOE for an integrated analysis of scope, validity, and infringement. Second, current case law hinders use of the DOE as a doctrine of claim scope. The public dedication doctrine of *Johnson & Johnston* holds that a patentee who discloses but fails to claim subject matter is prohibited from reaching that subject matter under the doctrine of equivalents. *See supra* note 21. The doctrine discourages patentees from including definitional information, or at least representative members, in their disclosures for fear that DOE coverage will be compromised.

248. *See, e.g.*, Craig Allen Nard, *Certainty, Fence Building, and the Useful Arts*, 74 IND. L.J. 759, 785-95 (1999) (discussing importance of certainty in patent system).

249. *See generally* JAMES BESSEN & MICHAEL J. MEURER, *PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK* (2008).

## B. Meaning and the Place of Definitional Information in Patent Doctrine

However, if definitional information is something that contributes to the certainty of patent claims, are we not referring to the definiteness requirement of section 112, ¶ 2, rather than the disclosure requirement of section 112, ¶ 1? In a sense we are if by “definiteness” we mean the ability to recognize the extent of legal rights deriving from disclosure of an invention. The current doctrine of claim definiteness was, along with enablement and written description, part of the vague category of “undue breadth” before the Court of Customs and Patent Appeals began differentiating the doctrines of section 112 as recognized today.<sup>250</sup> Since that time, the doctrine of claim definiteness has been associated almost entirely with the lexical and syntactical clarity of claim language rather than the claimed subject matter or the disclosure.<sup>251</sup> Likewise, in construing patent claims, we turn to the disclosure as a *resource* of meaning for claim language but we do not *demand* such meaning from the disclosure. Despite the Federal Circuit’s recent emphasis on a contextual interpretation of claim language,<sup>252</sup> the court is reluctant to turn to the disclosure as a source of definition unless particular claim terms are considered in need of interpretation.<sup>253</sup> Nor is possible invalidity of the claim for overbreadth (or for other reasons) to be considered in claim interpretation except in extremis.<sup>254</sup> Thus, in modern practice, the questions of scope, meaning, and clarity are, in the style of classical legal orthodoxy, largely independent inquiries governed by distinct analytical frameworks.

The notion of definitional information, however, partakes of scope, meaning, and clarity together. The existence of a concept straddling our current doctrinal boundaries may show us that the Court of Customs and Patent Appeals drew the boundaries between the doctrines of section 112 incorrectly, in which case much of this Article could be read as a call for a

---

250. See *supra* text accompanying notes 191-192.

251. See *supra* text accompanying notes 31-32.

252. See *Phillips v. AWH Corp.*, 415 F.3d 1303, 1313 (Fed. Cir. 2005) (emphasizing role of context in claim interpretation).

253. See *MBO Labs., Inc. v. Becton, Dickinson & Co.*, 474 F.3d 1323, 1330-31 (Fed. Cir. 2007) (restricting analysis to claim words disputed by parties).

254. See *Phillips*, 415 F.3d at 1327 (“While we have acknowledged the maxim that claims should be construed to preserve their validity, we have not applied that principle broadly, and we have certainly not endorsed a regime in which validity analysis is a regular component of claim construction.”).

radically transformed doctrine of definiteness rather than a defense of the existing written description doctrine. Certainly a renewed emphasis on the definitional content of the disclosure would lead to increased reliance on the disclosure as a source of meaning for claim interpretation, even if the nominal doctrines of claim construction saw no changes. However, we are unlikely to find such meaning in the disclosure unless we begin to demand it from patentees. If we seek meaning only in the words of the claim, then the current doctrine of indefiniteness may impose an adequate obligation of definitional information upon the patentee. On the other hand, if we believe that the meaning of the inventor's rights must be found within a larger context,<sup>255</sup> then the disclosure of the invention must become a more significant source of definitional information.<sup>256</sup>

Current enablement doctrine is not well-suited to enforce an obligation of definitional information. Enablement has difficulty grappling with the problems of genus and infinite claims. Instead, enablement is concerned with purely physical relations between entities in the physical world: the question of whether one of ordinary skill in the art could make or use a thing without undue time or effort is a *physical* function, even if its arguments (persons, things, and the acts of making and using) are hypothetical entities.<sup>257</sup> In contrast, the relation between scope and definition is very much one of description; logicians and linguists have been grappling with the problems of description in these terms at least since Aristotle. The longevity of that effort may give us pause, but if we

---

255. See Christopher A. Cotropia, *Patent Claim Interpretation Methodologies and Their Claim Scope Paradigms*, 47 WM. & MARY L. REV. 49, 105-24 (2005) (evaluating specification-oriented versus claim-language-oriented interpretive methodologies); Nard, *supra* note 21, at 43-64 (emphasizing importance of contextual information in claim interpretation). Professor Nard goes beyond the specification-oriented approach, arguing that to rely on the text of the patent alone disregards the social essence of language. *Id.* at 53-54.

256. This should not be read as a blanket endorsement of reading limitations from the disclosure into the claims; rather, viewing written description as a doctrine of definition may clarify when it is appropriate to limit the invention based on features disclosed in the description. If the patentee has employed the mode of definition by intension, then the features or properties disclosed do indeed characterize "the invention" and should be implied into all its embodiments. If the patentee has employed definition by extension, the question becomes to what extent that particular feature is present in the genus derived by degree of resemblance to the type members. Again, the question may be a complex factual one, but perhaps the intellectual framework of definition provides a more rational guide than the current system for invoking limitations from the specification.

257. See Durham, *supra* note 75, at 995-96 (noting problem of applying obviousness inquiry to claim in abstract, rather than subject matter within the claim).

treat scope problems in patent law at least in part as problems of definition, we will find a heritage of thought dedicated to the relationship between logical categories, language, and the physical world.<sup>258</sup>

## VI. CONCLUSION: THE PAST AND FUTURE OF CLAIMING

The problems posed by the necessity to define a genus did not arise prior to the development of the peripheral claiming system in its modern form. It is well-appreciated that the United States formerly followed a central system, in which the patentee described an embodiment of the invention and infringement was assessed by comparison between what the inventor had made or described and the accused subject matter. What is perhaps less appreciated is that, in their original form, claims were not considered to define subject matter in the same sense as modern claims. Early claims defined only the novel *inventive principle* the inventor created, not a category of objects or processes.<sup>259</sup> Indeed, in early practice, a claim defining the structure of an operative machine with a novel inventive feature, as modern claims do, was invalid, for the inventor had included old features over which he had no rights along with the new and inventive feature to which he was entitled.<sup>260</sup> If the “invention” is the novel principle discovered or created by the inventor, defined at the

---

258. See generally Collins, *supra* note 100 for a recent application of such thought to the problem of meaning of claims over time. A note of caution is in order here: there are perhaps deep difficulties with the classical notions of categories, at least as applied to human thought and language. See generally GEORGE LAKOFF, *WOMEN, FIRE, AND DANGEROUS THINGS: WHAT CATEGORIES REVEAL ABOUT THE MIND* (1987). Such discrepancies between the system of classical categories, upon which claims are based, and the structure of human language and thought, may have significant implications for claim interpretation as well as the task of definition.

259. See, e.g. *Winans v. Denmead*, 56 U.S. 330, 343 (1853) (stating that invention “did not consist in a change of form, but in the new employment of principles or powers, in a new mode of operation, embodied in a form by means of which a new or better result is produced”). Today *Winans* is characterized as an early case on the DOE. Certainly the policy arguments for or against infringement in *Winans* are the same ones invoked for or against the DOE today, but in historical context *Winans* reflects a nonstructural methodology of defining “the invention.” See Meurer & Nard, *supra* note 241, at 1961-66 (discussing early “principle of the invention” cases and *Winans*). Lutz notes that in 1863, a Patent Office rule expressly permitted a patentee to claim the inventive principle directly. See Karl B. Lutz, *Evolution of the Claims of U.S. Patents*, 20 J. PAT. OFF. SOC’Y 457, 465 (1938) (quoting PTO’s 1863 pamphlet, “Guide to Practice of the Patent Office”).

260. See N. J. Brumbaugh, *History and Purpose of Claims in United States Patent Law*, 14 J. PAT. OFF. SOC’Y 273, 279-81 (1932) (discussing early cases that required claim to distinguish between old and new).

appropriate level of generality,<sup>261</sup> then it is possible to ask whether an accused infringer is implementing that principle without having to decide whether the inventor enabled the category of all things that employ that principle.

Under this regime that German patent law retained until the late 20th century,<sup>262</sup> questions of claim scope, nonobviousness, and infringement were resolved as an integrated inquiry. The inventive principle contributed by the inventor was assessed in light of prior art, and claim scope and infringement were determined according to whether they embodied the inventive principle disclosed by the inventor.<sup>263</sup> Under such a system, it is not necessary to define *ex ante* a category of objects over which the inventor may exercise exclusive rights. The disadvantage of such a regime is that a patent has no definite bounds, and the public cannot be certain what is infringing and what is not.<sup>264</sup> Hence the development of the

---

261. Thus, for example, in characterizing an English patent, the Supreme Court in *O'Reilly v. Morse* stated that the inventor had not discovered the principle that hot air will promote the ignition of fuel better than cold air, but “he had invented a mechanical apparatus, by which a current of hot air, instead of cold, could be thrown in. . . . The interposition of a heated receptacle, in any form, was the novelty he invented.”). *O'Reilly v. Morse*, 56 U.S. 62, 116 (1853). In *O'Reilly* itself the deficiency of Morse’s eighth claim was that it was not limited to the recording structures disclosed in the specification, but the Court’s discussion of the Neilson patent suggests it was breadth of the principle, rather than failure to limit by particular structure, that doomed Morse’s claim. Then again, the similar division of the Justices in the contemporaneous *Winans* and *O'Reilly* cases may simply reflect disagreement on the Court between Justices who viewed claims as defining principles and those who viewed claims as defining structures. See Karl B. Lutz, *Evolution of the Claims of U.S. Patents*, 20 J. PAT. OFF. SOC’Y 377, 388 (1938) (discussing division of the Court).

262. Germany retained a central claiming system until 1973, when it joined the European Patent Convention. See Heinz Winkler, *The Scope of Patent Protection: Past, Present, and Future*, 10 INT’L REV. INDUS. PROP. & COPYRIGHT LAW 296 (1979).

263. This approach persisted well into the late 19th century, at least in the language of the decisions. See *Swain Turbine & Mfg. Co. v. Ladd*, 102 U.S. 408, 414-15 (1880) (noting in context of interpretation that prior art waterwheels “contained the fundamental element . . . which the appellant claims as the principle of Swain’s invention.”); see also *In re Cawood Patent*, 94 U.S. 695, 701 (1876) (discussing novelty and infringement in terms of “principle of the invention” rather than structure alone).

264. The uncertainty of central claiming did not seem to impede technological development in Germany during the late 19th and 20th centuries. However, the German economy, at least prior to World War II, was characterized by industrial concentration and infrequent patent litigation, rather than vigorous competition and commonplace litigation. See Heinrich Kronstein, *The Dynamics of German Cartels and Patents*, 9 U. CHI. L. REV. 643 (1992). These characteristics may have blunted any chilling effect of uncertain patent scope.

modern American system of claiming: in which claims recite properties precisely defining a set of objects or processes over which the patentee asserts exclusive rights.<sup>265</sup> As the notions of invention, claim, and legal right converged, the concepts of scope, infringement, and nonobviousness crystallized into distinct doctrines. However, in fixating upon a system in which legal rights were precisely delineated by a system of claims that defined not what the inventor had created, but to what he was entitled, patent law lost at least two advantages of the central claiming system: the ability to integrate information about patent validity and the technological advance represented by the invention into the infringement inquiry,<sup>266</sup> and the ability to define the inventor's *permissible* entitlement in a theoretically rigorous manner.

So long as the structure of patent law remained less formal, the looseness of patent doctrine may have concealed the lack of a coherent theory of permissible claim scope.<sup>267</sup> If so, then modern efforts to condense patent law into a conceptually ordered system have brought this deficiency into sharp relief. Perhaps we cannot resolve the problem of patent scope without abandoning, at least in part, the peripheral claiming system that lies at the heart of modern patent law.<sup>268</sup> Yet if patent law is to

---

265. For a snapshot of the transition, see Brumbaugh, *supra* note 260, at 283-84 (describing rejection of forms in participial form in the early 1870s).

266. The reverse doctrine of equivalents, by considering the relationship between the patentee's disclosed invention and the accused subject matter, retained some of the integrative features of the earlier system. But the doctrine has been essentially abandoned in favor of the modern practice of quarantined inquiries. *See supra* note 23.

267. Judge Learned Hand described resort to the doctrine of equivalents as a means "to temper unsparing logic." *Royal Typewriter Co. v. Remington Rand*, 168 F.2d 691, 692 (2d Cir. 1948).

268. Commentators who advocate modification of the enablement requirement, and the elimination of written description as a limit on originally filed claims, ultimately arrive at a rejection of the formal peripheral claiming concept, though they do not describe it as such. The foundation of the formal system is that claim scope, validity, and infringement are all independent entities; though claims are construed to preserve their validity, the scope of the claim is in theory fixed and the question of whether accused subject matter falls within the patent's claims is resolved without reference to the question of whether the patentee's disclosure enables that particular embodiment. Professor Holbrook, however, concludes that (as a matter of claim construction) "[i]n order to literally infringe the patent, the patent would have to enable *the accused device*, thus showing that the patentee had placed the PHOSITA in possession of it." Holbrook, *supra* note 62, at 158 (emphasis added). Likewise, Professor Feldman would frame the inquiry as "the leap that it will take to get from what the inventor actually disclosed to the product that the inventor is trying to reach." Feldman, *supra* note 158, at 40. She

be a conceptually ordered system founded upon the peripheral claim, then the doctrine of written description, conceived as a doctrine of definition, will remain a necessary aspect of the law.

---

recognizes that this approach departs from the traditional notion that patent scope is determined without reference to the allegedly infringing material. *Id.*

# EMOTION, DILUTION, AND THE TRADEMARK CONSUMER

By *Laura R. Bradford*<sup>†</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1228
II. TRADEMARKS, INFORMATION, AND COMPETITION .....	1240
A. THE INFORMATIONAL PURPOSE OF BRANDS .....	1240
B. TRADEMARK DILUTION AND INTERNAL CONFUSION .....	1242
III. THREE CONCEPTIONS OF EMOTION IN CONSUMER DECISION-MAKING .....	1251
A. IRRATIONAL WEIGHERS: CONSUMERS MISLED BY EMOTION MAKE BAD CHOICES.....	1253
B. RATIONAL MAXIMIZERS: USEFUL PRODUCT INFORMATION INFLUENCED BY EMOTION .....	1255
C. COGNITIVE MISERS: USING EMOTION TO MINIMIZE COGNITIVE DECISION-MAKING COSTS.....	1260
1. <i>The Intertwined Roles of Emotion and Reason</i> .....	1260
2. <i>Emotion as a Heuristic in Consumer Decision-Making</i> .....	1262
3. <i>Familiarity as Trademark Selling Power</i> .....	1271
IV. IMPLICATIONS .....	1278
A. BLURRING IS A FORM OF TARNISHMENT .....	1278
B. INFORMATION QUALITY AND PROPERTY RIGHTS IN MENTAL PROCESSES .....	1279
1. <i>Lowering Brand Credibility does not necessarily frustrate Consumer Preferences</i> .....	1281
2. <i>Property Rights in Positive Feelings is Beyond the Scope of IP Law</i> .....	1283

---

© 2008 Laura R. Bradford

<sup>†</sup> Assistant Professor of Law, George Mason University School of Law. I am grateful for comments on earlier versions from Robert Brauneis, Eric Claeys, Bruce Kobayashi, Mark McKenna, Roger Schechter, Rebecca Tushnet, Samson Vermont, and participants at the 2007 Trademark Dilution conference at Santa Clara University School of Law. Thanks also to the library staff at the George Washington University School of Law and George Mason University School of Law for research and other assistance. All errors are mine. Comments welcome at [lbradfor@gmu.edu](mailto:lbradfor@gmu.edu).

3.	<i>Enforcing Different Advertising Standards for Established Producers and Market Newcomers Promotes Neither Accuracy nor Competition</i> .....	1284
C.	FREE-RIDING AS INFORMATION .....	1286
1.	<i>Blurring Eases Barriers to Entry</i> .....	1287
2.	<i>Blurring Can Increase Useful Product Information Available to Consumers</i> .....	1291
3.	<i>Other Alternatives to Dilution Regulation Exist</i> .....	1296
D.	LIMITING DILUTION TO USE OF THE EXACT MARK .....	1297
V.	<b>BEYOND DILUTION: DECOUPLING CONSUMER SURVEY RESULTS AND THE EFFICIENT MARKET</b> .....	1297

*The protection of trade-marks is the law's recognition of the psychological function of symbols.*

—*Mishawaka Rubber & Woolen Manufacturing v. S. S. Kresge*<sup>1</sup>

## I. INTRODUCTION

The law generally ignores the role of emotions in consumer decision-making, although emotions are widely acknowledged to play a dominant role in shaping preferences concerning risk,<sup>2</sup> borrowing,<sup>3</sup> consumption<sup>4</sup>

1. 316 U.S. 203, 205 (1942).

2. See, e.g., Dan M. Kahan, *Two Conceptions of Emotion in Risk Regulation*, 156 U. PA. L. REV. 101 (2008); George F. Loewenstein et al., *Risk as Feelings*, 127 PSYCHOL. BULL. 267 (2001); Rachel F. Moran, *Fear Unbound: A Reply to Professor Sunstein*, 42 WASHBURN L.J. 1 (2002); Paul Slovic, *What's Fear Got to Do With it? It's Affect We Need to Worry About*, 69 MO. L. REV. 971 (2004).

3. Michael Holtje et al., *Psychology and BAPCPA: Does Greater Disclosure Affect Consumer Credit Behavior?*, 26 AM. BANKR. INST. J. 20 (2007); George Loewenstein & Ted O'Donoghue, *"We Can Do this the Easy Way or the Hard Way": Negative Emotions, Self-Regulation and the Law*, 73 U. CHI. L. REV. 183, 195-200 (2006); Cass R. Sunstein, *Boundedly Rational Borrowing*, 73 U. CHI. L. REV. 249 (2006); cf. Oren Bar-Gill, *Bundling and Consumer Misperception*, 73 U. CHI. L. REV. 33 (2006) (referring to consumer "misperception" of credit risks but not exploring the sources of consumer myopia).

4. See generally Tim Ambler et. al., *Saliency and Choice: Neural Correlates of Shopping Decisions*, 21 PSYCHOL. & MARKETING 247 (2004); Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: The Problem of Market Manipulation*, 74 N.Y.U. L. REV. 630 (1999) ("[O]ur affective responses to products more often than not determine the purchasing decision, regardless of whether we experience the decision as having resulted from 'reasons.'"); Loewenstein & O'Donoghue, *supra* note 3, at 188-96; Girish N. Punj & Clayton L. Hillyer, *A Cognitive Model of Consumer-Based Brand Equity for Frequently Purchased Products: Conceptual Framework and Empirical Results*, 14 J. CONSUMER PSYCHOL. 124 (2004).

and choice.<sup>5</sup> Trademark law has been especially suspicious of the role that emotion plays in increasing demand for branded goods.<sup>6</sup> Some have argued that emotional advertising causes consumers irrationally to pay a premium for trademarked products that are not functionally superior to generic substitutes.<sup>7</sup> Accordingly, what one thinks about the emotional influence of popular brands tends to dictate one's views about the proper scope of trademark protection.<sup>8</sup>

---

5. See generally Ambler et al. *supra* note 4; George Loewenstein & Jennifer S. Lerner, *The Role of Affect in Decision Making*, in HANDBOOK OF AFFECTIVE SCIENCE 619 (Richard J. Davidson ed., 2003); Barbara Mellers et al., *Emotion-Based Choice*, 128 J. EXPERIMENTAL PSYCHOL.: GEN. 332 (1999); Michel Tuan Pham et al., *Affect Monitoring and the Primacy of Feelings in Judgment*, 28 J. CONSUMER RES. 167 (2001); Punj & Hil-lyer, *supra* note 4; R. B. Zajonc, *Feeling and Thinking: Preferences Need No Inferences*, 35 AM. PSYCHOLOGIST 151, 155 (1980) [hereinafter Zajonc, *Feeling and Thinking*]; David J. Arkush, *Situating Emotion: A Critical Realist View of Emotion and Nonconscious Cognitive Processes for the Law* (Aug. 20, 2007) (unpublished manuscript), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1003562](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1003562); cf. Colin F. Camerer, *Wanting, Liking, and Learning: Neuroscience and Paternalism*, 73 U. CHI. L. REV. 87 (2006) (finding that human learning functions coordinate liking and wanting but noting that in cases of behavioral disorders we want things even though we don't like them).

6. See, e.g., *Smith v. Chanel, Inc.*, 402 F.2d 562, 567 (9th Cir. 1968). The court stated:

The primary value of the modern trademark lies in the "conditioned reflex developed in the buyer . . ." To the extent that advertising of this type succeeds, it is suggested, the trademark is endowed with sales appeal independent of the quality or price of the product to which it is attached; economically irrational elements are introduced into consumer choices . . . .

*Id.*; see also *Triangle Publ'ns, Inc. v. Rohrlich*, 167 F.2d 969, 980 n.13 (2d Cir. 1948) (Frank, J. dissenting) ("[Broad trademark protection] enables one to acquire a vested interest in a demand 'spuriously' simulated through 'the art of advertising' . . .").

7. See, e.g., EDWARD CHAMBERLAIN, *THE THEORY OF MONOPOLISTIC COMPETITION* (5th ed. 1946); Ralph S. Brown, Jr., *Advertising and the Public Interest: Legal Protection of Trade Symbols*, 57 YALE L.J. 1165 (1948).

8. Compare Robert G. Bone, *Hunting Goodwill: A History of the Concept of Goodwill in Trademark Law*, 86 B.U. L. REV. 547 (2006) [hereinafter Bone, *Hunting Goodwill*] (arguing that protecting positive feelings as part of the sellers goodwill is unnecessary); Glynn S. Lunney, Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 437-39 (1999) (arguing that protection for the value of trademarks beyond the basic identification of source, such as the protection of trademark prestige or "selling power," exceeds the costs of competition and is socially wasteful) with Shahar J. Dilbary, *Famous Trademarks and the Rational Basis for Protecting "Irrational Beliefs"*, 14 GEO. MASON L. REV. 605 (2007) (finding that consumers derive value from the intangible auras of branded products and arguing for strong trademark protection); William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J.L. & ECON. 265, 270, 274-75 (1987) (rejecting the irrational characterization of consumer response to

This Article argues that trademark law can benefit from an updated understanding of the influence of emotion on consumer decision-making. Psychological research in this area is relatively new.<sup>9</sup> Yet it already reveals a number of insights that can be useful for trademark theory. Specifically, research on emotion and choice can shed light on one of trademark law's most elusive and controversial doctrines: trademark dilution. Modern consumer emotion research provides a sturdier justification for dilution protection than economic or cognitive science doctrines standing alone. This is because trademark fame, or familiarity, signals information about risk and quality to consumers through quick and efficient innate emotional response mechanisms. However, while reliance on fame lowers effort or "search costs" for buyers, it doesn't do much to promote trademark law's other stated aim: the efficient exchange of information about consumer preferences concerning specific product features, functions, or quality. Instead, overly strict protection of mark familiarity through the law of trademark dilution can burden competitors who signal product quality and reliability using communication strategies besides fame. Therefore, this Article argues in favor of a very narrow dilution regime that will conserve the signaling value of mark fame for consumers but avoid unduly burdening competitor efforts to communicate product value through other means.

---

trademark advertising and condemning any free-riding on the allure of strong trademarks).

9. In the last twenty years, advances in the understanding of human emotional processes have led to re-evaluations of long-standing assumptions in fields as diverse as criminal law, commercial law, and securities regulation. For representative criminal law articles, see, e.g., Katherine K. Baker, *Gender and Emotion in Criminal Law*, 28 HARV. J.L. & GENDER 447 (2005); Dan M. Kahan & Martha C. Nussbaum, *Two Conceptions of Emotion in Criminal Law*, 96 COLUM. L. REV. 269 (1996); Victoria Nourse, *Passion's Progress: Modern Law Reform and the Provocation Defense*, 106 YALE L.J. 1331, 1331-35 (1997); Arkush, *supra* note 5, at 78. For sources related to emotions and commercial law, see, for example, sources cited *supra* at note 3. For representative literature on securities regulation and investor behavior, see generally, Peter H. Huang, *Emotional Impact Analysis in Financial Regulation: Going Beyond Cost-Benefit Analysis* (Institute for Advanced Study, School of Social Science, Economics Working Paper No. 62, 2006), available at [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=870453](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=870453) (arguing that financial regulators should consider the impact of policies on investor confidence and overall market mood); Peter H. Huang, *How Do Securities Laws Influence Affect, Happiness, & Trust?*, 3 J. BUS. & TECH. L. 257 (2008); Peter H. Huang, *Regulating Irrational Exuberance and Anxiety in Securities Markets*, in THE LAW AND ECONOMICS OF IRRATIONAL BEHAVIOR 501 (Francesco Parisi & Vernon L. Smith eds., 2005); Peter H. Huang, *Moody Investing and the Supreme Court: Rethinking the Materiality of Information and the Reasonableness of Investor*, 13 SUP. CT. ECON. REV. 99 (2005); Peter H. Huang, *Trust, Guilt, and Securities Regulation*, 151 U. PA. L. REV. 1059, 1075-89 (2003).

Trademark dilution is a relatively recent innovation.<sup>10</sup> Trademark law traditionally aims to improve the quality of market information through prohibitions on deceptive uses of trade symbols.<sup>11</sup> Symbols that confuse consumers reduce the efficiency of the market by causing consumers to purchase the wrong good. More broadly, protection for the informational integrity of trade symbols allows consumers to spend less time and effort searching for desired goods and so lowers “search costs.”<sup>12</sup>

Trademark dilution law extends these prohibitions to interferences with the uniqueness of a trademark. For example, consumers may or may not think that “Chevrolet shoes” were made by the car company, but their presence in the marketplace would diminish or “dilute” the singularity of the original Chevrolet mark. The harm protected against, as classically described, is “the gradual whittling away or dispersion of the identity and hold upon the public mind of the mark or name by its use on non-competing goods.”<sup>13</sup> Examples of recent uses of famous marks found to be dilutive include the use of the mark “Perfume-bay” for E-bay, the mark “Nikepal” for Nike, and the name “Hot Rigz” for “Hot Wheels” toy cars.<sup>14</sup>

Trademark owners think that dilution is harmful but have had trouble explaining their reasoning. Proponents of dilution regulation have linked the dilution cause of action to the goal of lowering search costs.<sup>15</sup> They

---

10. The first federal dilution law was passed in 1995. Federal Trademark Dilution Act of 1995, Pub. Law. 104-98, codified at 15 U.S.C. § 1125(c). Before this trademark dilution protection was available on a patchwork basis from state law. The first state anti-dilution law was passed in 1947 in Massachusetts.

11. Stacey L. Dogan, *What is Dilution, Anyway?*, 105 MICH. L. REV. FIRST IMPRESSIONS 103, 106 (2006) (“Trademark law has never aimed to provide exclusive rights in marks, but has focused on preserving informational clarity in the marketplace.”); see also William P. Kratzke, *Normative Economic Analysis of Trademark Law*, 21 MEM. ST. U. L. REV. 199, 216-217 (1991) (arguing that law should grant exclusive rights in trademark interests to facilitate the transmission of informational and identifying messages); Lunney, *supra* note 8, at 431-32.

12. See, e.g., *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159, 163-64 (1995) (“[T]rademark law . . . ‘reduce[s] the customer’s costs of shopping and making purchasing decisions’”); WILLIAM LANDES & RICHARD POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* at 168 (2003).

13. Frank I. Schechter, *The Rational Basis of Trademark Protection*, 40 HARV. L. REV. 813, 825 (1927).

14. *Perfumebay.com Inc. v. eBay, Inc.*, 506 F.3d 1165 (9th Cir. 2007); *Jada Toys, Inc. v. Mattel, Inc.*, 518 F.3d 628 (9th Cir. 2008); *Nike Inc. v. Nikepal Int’l Inc.*, 84 U.S.P.Q.2d (BNA) 1820 (E.D. Cal. 2007).

15. See, e.g. Jerre B. Swann, Sr., *Dilution Redefined for the Year 2002*, 92 TRADEMARK REP. 585, 585 (2002); Jacob Jacoby, *Dilution in Light of Victoria’s Secret: The Psychology, Varieties and Measurement of Trademark Dilution*, 9-11 (NYU Ctr. for Law & Bus. Working Paper No. CLB-03-020, 2008), available at <http://w4.stern.nyu.edu/>

argue that promiscuous use of well-known symbols will cause the mark's meaning and significance to decline, and therefore also the mark's utility as a means for quickly locating goods.<sup>16</sup> In this formulation, free-riding on the familiarity of well-known marks increases "internal search costs," or the amount of mental time and effort consumers must expend to connect the mark to its original owner and larger goodwill.<sup>17</sup>

Critics are skeptical that a few extra seconds of cogitation, assuming they are required, justifies a race to the federal courthouse if consumers are not actually confused about who makes a particular good.<sup>18</sup> Famous

---

emplibrary/03-020.pdf [hereinafter Jacoby, *Dilution in Light of Victoria's Secret*]; see also Dogan, *supra* note 11, at 104; Stacey L. Dogan & Mark A. Lemley, *The Merchandising Right: Fragile Theory or Fait Accompli?*, 54 EMORY L.J. 461, 493 (2005) ("[P]roperly understood, dilution is targeted at reducing consumer search costs, just as traditional trademark law is."); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1704 n.90 (1999) ("The information consumers can obtain and process is in part a function of how clear the association between mark and product remains in their minds; 'clutter' therefore imposes real costs on consumers."); J. Thomas McCarthy, *Proving a Trademark Has Been Diluted: Theories or Facts?*, 41 HOUS. L. REV. 713, 727-28 (2004). McCarthy argues:

[T]here is potential harm to both consumers and mark owners if a once-unique designation loses its uniqueness. The argument is that this makes it harder for consumers to link that designation with a single source--the hallmark of a strong trademark. Under this theory, dilution increases the consumer's search costs by diffusing the identification power of that designation.

*Id.*; Richard A. Posner, *When Is Parody Fair Use?*, 21 J. LEGAL STUD. 67, 75 (1992). Posner states:

A trademark seeks to economize on information costs by providing a compact, memorable, and unambiguous identifier of a product or service. The economy is less when, because the trademark has other associations, a person seeing it must think for a moment before recognizing it as the mark of the product or service.

*Id.*

16. See Swann, *supra* note 15, at 598-611 (arguing that dilution lowers a mark's "communicative clarity" and ability to cut through clutter for consumers); Frank I. Schecter, *The Rational Basis of Trademark Protection*, 40 HARV. L. REV. 813, 830 (1927); see also *Amstar Corp. v. Domino's Pizza, Inc.*, 615 F.2d 252, 259-60 (5th Cir. 1980).

17. LANDES & POSNER *supra* note 1212, at 207; Jacob Jacoby, *The Psychological Foundations of Trademark Law: Secondary Meaning, Genericism, Fame, Confusion and Dilution*, 91 TRADEMARK REP. 1013, 1047 (2001) [hereinafter Jacoby, *Psychological Foundations*].

18. See, e.g., Bone, *Hunting Goodwill*, *supra* note 8, at 559 (arguing that uses that cause a consumer to reflect a bit longer but do not confuse consumers do not interfere with a trademark's core purpose of protecting the informational clarity of marks); Christine Haight Farley, *Why We Are Confused About the Trademark Dilution Law*, 16 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1175, 1184 (2006) (arguing that the harm of dilu-

marks have always existed side-by-side with lesser-known siblings. Ford Motor Company and Ford Modeling Agency are both market leaders and neither seems to suffer from the presence of the other, though the consumer presumably must work at the margin to distinguish them.<sup>19</sup> Protecting the appeal of advertising symbols from dilution also seems to squarely conflict with free speech interests in promoting criticism of, discussion about, and comparison with well-known brands.<sup>20</sup> Not surprisingly, courts have found the harm threatened by dilution “dauntingly elusive” to comprehend; accordingly, they have been reluctant to enforce the law as written.<sup>21</sup> Nonetheless, trademark owners have successfully lobbied Congress to shore up and reinforce the trademark dilution action in the Trademark Dilution Revision Act of 2006.<sup>22</sup> Part II of this Article explains the traditional justifications for trademark regulation and demonstrates how dilution proponents have attempted to reconcile the law with these aims, with limited success.

Part III of this Article argues that trademark dilution law is so difficult to understand because it aims at emotion and only indirectly at information.<sup>23</sup> The emotion referred to here is of the most basic kind: “affect”

---

tion is illusory and the law is based on an unjust enrichment rationale); David S. Welkowitz, *Reexamining Trademark Dilution*, 44 VAND. L. REV. 531, 542-44 (1991) (claiming that trademark owners have failed to demonstrate any actual harm flowing from dilution).

19. Barton Beebe, *A Defense of the New Federal Trademark Antidilution Law*, 16 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1143, 1150 (2006) [hereinafter Beebe, *A Defense*]; see also Transcript of Oral Argument, *Moseley v. Victoria's Secret Catalogue*, 537 U.S. 418 (2003) 2002 WL 31643067 at \*10-11 (posing the questions whether the two uses of Ford, and also Delta Airlines, Delta Plumbing and Delta Dental caused the same kind of harm as dilution).

20. See, e.g., Rebecca Tushnet, *Gone in Sixty MilliSeconds: Trademark Law and Cognitive Science*, 86 TEX. L. REV. 507, 548-52 (2008).

21. *Ringling Bros.-Barnum & Bailey Combined Shows, Inc. v. Utah Div. of Travel Dev.*, 170 F.3d 449, 451 (4th Cir. 1999) (labeling the dilution cause of action as “dauntingly elusive”); Clarisa Long, *Dilution*, 106 COLUM. L. REV. 1029, 1031 (2006) (finding that the judicial enforcement of dilution law has diminished over time).

22. See, e.g., Beebe, *A Defense*, *supra* note 19 at 1155 (noting that the Trademark Dilution Revision Act was the result of extensive work by the International Trademark Association, a trade association of mark holders and their attorneys and the American Intellectual Property Law Association, a bar association of corporate intellectual property attorneys).

23. The role of emotion in trademark law has been the subject of some controversy, and it is important to clarify what I mean when I refer to consumer emotions. See *infra* text accompanying notes 116-118, 158-180 (describing traditional and modern account of emotion's impact on decision-making). Traditionally, commentators have referred to “emotions” as the feelings that arise in reaction to the content of advertising. However, this paradigm, like most trademark theory, focuses on seller communications and intentions and assumes that buyers understand such communications at face value. See, e.g.,

or the automatic negative or positive response that a mark generates when viewed by a consumer.<sup>24</sup> Research on brands and emotion teaches that familiar trademarks offer value to consumers by lowering the amount of effort required in purchasing decisions.<sup>25</sup> Consumers experience positive feelings from the ease of evaluating familiar marks and accordingly infer positive qualities for the brand and the underlying product. Positive and negative emotions in this way serve as second-order sources of information about the quality and risk of the underlying good. Those who use famous marks in ways inconsistent from the owner risk making the marks more costly to evaluate, and thus may cause consumers to automatically feel more negatively toward the original brand and affiliated products.<sup>26</sup> Dilutive conduct can thus undercut the reliability of trademarks as sources of information for buyers. This, in a nutshell, is the harm caused by “blurring” famous marks.

Understanding the dilution cause of action through the lens of emotion clarifies an elusive doctrine, but reveals reasons to be wary of over-enforcement. Properly understood, dilution regulation aims to preserve the signaling value of brand familiarity for consumers. Because consumer markets suffer from information asymmetries in which sellers possess better information than buyers, buyers depend on signals from sellers and other third parties in making purchasing decisions.<sup>27</sup> Brand advertising

---

Brown, *supra* note 7, at 1182; Swann, *supra* note 15, at 591-94. I am interested in emotion in a different sense, namely, as a mediator of information exchange between sellers and buyers. Contrary to the assumptions of most trademark theory, buyers do not passively process advertising on its face, but use a variety of strategies, including automatic emotional response, to extract information from advertising without having to allocate too much attention to it. *See infra* text accompanying notes 171-213. The problem with dilution then is not, as some overzealous commentators have argued, that it changes the meaning of the famous mark, *e.g.* Swann, *supra* note 15, at 598, 601 (arguing that dilution protects a mark’s singular identity and meaning), but that it changes how reliable that mark is to buyers as a source of information about quality. The distinction is important because any legal regime that sought to ensure that consumers only experienced positive reactions to famous brands would face significant conflicts with the First Amendment and free competition, and I do not wish to advocate in favor of any such regime.

24. *See, e.g.*, Slovic, *supra* note 2 at 971-72 for similar definition of “affect.”

25. *See infra* text accompanying notes 214-249.

26. *See infra* text accompanying notes 54-56.

27. For example, buyers might use the relative price, or warranty information, third party rating systems or comparative advertising by competitors as signals about the underlying quality of goods that cannot be effectively evaluated before purchase. *See generally* Tülin Erdem & Joffre Swait, *Brand Equity as a Signaling Phenomenon*, 7 J. CONSUMER PSYCHOL. 131, 132-33 (1998) (discussing different kinds of market signals); Amna Kirmani & Akshay R. Rao, *No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality*, 64 J. MARKETING 66, 66-69, 75-76 (2000)

and the resulting brand familiarity can be one such type of signal. Familiar brands offer lower risk and decreased evaluation costs. Consumers rationally rely on familiarity as a signal that the seller has a substantial investment in the brand that will be forfeited if it fails to adequately police quality. Familiarity also can convey that others in the marketplace have found the seller's goods satisfactory. Consumers therefore can rely on low-cost emotional responses to familiarity as sources of information about goods rather than using more costly methods of direct evaluation.<sup>28</sup> Dilution regulation seeks to shore up the signaling value of brand advertising by ensuring that only the seller responsible for making a brand familiar to consumers can capture the rewards of positive consumer response.<sup>29</sup> It also prevents illicit follow-on uses that muddy the clarity of this signal.<sup>30</sup>

Situating the dilution doctrine within the larger universe of strategies that sellers use to communicate about product quality also readily reveals its costs. Part IV of this Article explores some of these costs. The emotion literature suggests, in contradiction to the claims of dilution regulation proponents, that much of the "selling power" of famous marks is due primarily to their familiarity and not any specific benefit, tangible or intangible, of the product. Brand familiarity as a signal is convenient for consumers, but it also is anti-competitive. Privileging the signaling value of familiarity disadvantages market newcomers, and therefore reduces the choices available to buyers.<sup>31</sup> In addition, encouraging consumer reliance on brand

---

(presenting a typology of different market signals); Phillip Nelson, *Advertising as Information*, 82 J. POL. ECON. 729 (1974) [hereinafter Nelson, *Advertising as Information*] (advertising as a seller signal); George J. Stigler, *The Economics of Information*, 69 J. POL. ECON. 213, 224 (1961); Joseph E. Stiglitz, *The Causes and Consequences of the Dependence of Quality on Price*, 25 J. ECON. LITERATURE 1, 1-48 (1987) (price signals).

28. See *infra* text accompanying notes 171-180.

29. Cf. Ill. High Sch. Ass'n v. GTE Vantage Inc., 99 F.3d 244 (7th Cir. 1996).

[Trademark dilution laws] protect the trademark owner from the erosion of the distinctiveness and prestige of a trademark caused by . . . a proliferation of borrowings that, while not degrading the original seller's mark, are so numerous as to deprive the mark of its distinctiveness and hence impact . . . , even though there is no confusion of source.

*Id.* at 247; see also LANDES & POSNER, *supra* note 12, at 172 (free-riders who appropriate the strong brands of others destroy the informational capital built up in the brand).

30. McCarthy, *supra* note 15, at 727-28.

31. See *infra* Part IV; see also Howard Beales, et al., *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 508 (1981) (arguing that use of market share as a quality signal may cause the first entrant in a product class to maintain its high share by force of historical monopoly rather than the superiority of its product). This would be the case when the consumer values avoiding risk and conserving effort over optimal product selection. Studies suggest that many consumers approach product choice

familiarity can discourage product innovation because familiarity has been shown to provide a competitive advantage even if the specific attributes of the underlying product or brand are not objectively superior.<sup>32</sup> Legal protection for familiarity also may reduce the kinds of information available to consumers about existing products. Sellers have a variety of avenues available with which to communicate information about unobservable product characteristics. These include price, warranty, third-party certifications, and advertising about product features. Granting proprietary rights in the consistency of brand signaling when these arguably more informative forms of quality signaling are not protected from copying adds to already lopsided incentives to increase brand advertising at the expense of these other methods.<sup>33</sup>

Even the conduct sought to be regulated, trademark blurring, can serve as a useful signal of quality. The law already accepts that some unauthorized uses of familiar marks are more helpful than harmful; the paradigmatic case is comparative advertising. The statement “better than brand X,” made by a competitor is helpful because it alerts consumers to new choices.<sup>34</sup> In the same way, a slogan, name, logo, or design that communicates that a product is “X-like,” especially in a distinct market, may also offer useful information. Use of a familiar symbol to call attention to a new product is nothing new. Many brand leaders who now zealously condemn blurring previously relied on a blurring strategy to ease their own entry into the market. For example, Tiffany’s chose its singular blue color and McDonald’s its clown spokes-character to free-ride on the appeal of well-known products of the time.<sup>35</sup> This kind of advertising may lower the immediate signaling value of strong brands, but consumers benefit in the long-term from learning easily about new options.<sup>36</sup>

In addition, free-riding can serve as a signal about the quality of the famous, blurred mark for the same reasons claimed for brand advertising. Just as familiarity of a mark can convey information about producer in-

---

with these preferences. *See, e.g.*, Wayne D. Hoyer, *An Examination of Consumer Decision Making for a Common Repeat Purchase Product*, 11 J. CONSUMER RES. 822 (1984).

32. William E. Baker, *When Can Affective Conditioning and Mere Exposure Directly Influence Brand Choice?*, 28 J. ADVERT. 31, 44-45 (1999); *cf.* Beales et al., *supra* note 29, at 508.

33. *See infra* text accompanying notes 309-314.

34. *Smith v. Chanel, Inc.*, 402 F.2d 562, 567 (9th Cir. 1968); *see also* Landes & Posner, *supra* note 8, at 307-09 (arguing that dilution protection should not be used to prevent competitors from advertising that they have a comparative product).

35. *See infra* text accompanying notes 286-298.

36. *Cf.* Kratzke, *supra* note 8, at 211 (stating that the paramount objective of trademark law should be to promote inter-brand competition).

vestment in quality, the use of a famous mark by an interloper, even if not overtly critical or parodic, can alert consumers to risks inherent in the dominant brand. Typically interlopers that evoke famous marks in advertising subtly reveal to consumers alternative understandings such as that some segment of the public thinks the brand is pretentious, a bad value, or simply over-exposed.<sup>37</sup> Even if consumers process the new message only subconsciously, automatically, and involuntarily, as they do with much authorized brand advertising, a resulting increase in negative feelings about the dominant brand may well be welfare-enhancing and efficient. Whether consumers respond to advertising “rationally” or “emotionally” is a red herring; what matters is whether consumers can easily find reliable information about the risks and benefits of different products. Thus, the optimal dilution regime will protect consumer interests in the reliability of branding signals, but avoid undermining competition or incentives to provide reliable information through other means.

Acknowledging consumer interests on both sides of the dilution debate facilitates a more tailored application of the dilution cause of action. Part IV of this Article proposes revising trademark dilution law to prohibit only commercial uses of exact copies of famous marks.<sup>38</sup> Trivial alterations or exact uses of famous marks on unrelated goods increase consumer costs without adding new information, but nontrivial alterations are likely to add more than they detract from the quality of information exchanged in the marketplace, and therefore should not be subject to dilution liability.<sup>39</sup>

---

37. See *infra* text accompanying notes 321-325. As detailed *infra*, non-competitors may have greater incentives than direct competitors to expose these kinds of flaws because competitors risk undermining the appeal of the product category if they undermine a dominant seller, or at the very least, will have to share any market gains from lowering the dominant brand’s appeal with other sellers of the same product. See *infra* notes 317-318.

38. Subject to the usual defenses for expressive, descriptive, and critical use. One might argue that a formulation this narrow renders dilution law irrelevant since use of a mark identical to a famous one is likely to be confusing in any product market. However, use of a dilution analysis in such situations can lower enforcement costs. See, e.g., Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 VA. L. REV. 2099 (2004) [hereinafter Bone, *Enforcement Costs*] (arguing that dilution prohibitions lower the costs of enforcing famous marks in situations where confusion is likely).

39. Of course, if such a use is likely to cause confusion, it will be subject to a trademark infringement suit. Others have advocated variations of a cost-benefit social welfare analysis for trademark dilution. See, e.g., Long, *supra* note 21, at 1057 (arguing for a different approach to low-value uses such as counterfeiting of famous marks and high-value uses such as parody and satire). *But cf.* Tushnet, *supra* note 20 at 566 (endorsing strict fame, uniqueness, and identity requirements as one possible way to limit the dilution regime). An identity requirement does find support in the caselaw. See, e.g., Mo-

More broadly, a greater understanding of how consumers rely on brand signals reveals both the potential and the limitations of cognitive psychology in shaping trademark regulations. Trademark law has historically relied on judicial assumptions about consumers to determine trademark rights and liabilities in a variety of contexts. Behavioral research offers empirical data against which to test these assumptions.<sup>40</sup> However,

---

seley v. V Secret Catalogue, Inc., 537 U.S. 418, 434 (2003) (suggesting a lighter burden of proof for demonstrating dilution where a defendant has used an identical mark). Interestingly, although some courts claim to require identity or near identity between marks for dilution liability, in practice they require merely that the junior mark be similar enough to call to mind the famous mark. *See, e.g., Nike Inc. v. Nikepal Int'l Inc.*, 84 U.S.P.Q.2d (BNA) 1820 at \*6 (E.D. Cal. 2007) (finding mark Nikepal to be functionally identical to Nike because a majority of consumers think of the latter when viewing the former); *see also Thane Int'l, Inc. v. Trek Bicycle Corp.*, 305 F.3d 894, 907 (9th Cir. 2002) (finding that a reasonable fact finder could conclude that the marks "TREK" and "OrbiTrek" were nearly identical to one another); *Nabisco, Inc. v. PF Brands, Inc.*, 191 F.3d 208, 218 (2d Cir. 1999) (stating that marks must be similar enough that the junior mark conjures an association with the senior); *accord Fed. Express Corp. v. Fed. Espresso, Inc.*, 201 F.3d 168, 176 (2d Cir. 2000) (opining that the marks "Federal Express" and "Federal Espresso" were sufficiently similar to support a dilution cause of action); *cf. Eli Lilly & Co. v. Natural Answers, Inc.*, 233 F.3d 456, 469 (7th Cir. 2000) (evaluating confusion and dilution using the same similarity standard).

40. The use of behavioral research in the context of trademark law has recently been strongly criticized. Tushnet, *supra* note 20, at 510, 528-546 (2008). And indeed, much of the existing scholarship in this area is worthy of criticism. Trademark advocates seeking to explain trademark dilution to a wary judiciary have mined the cognitive literature for studies that purport to show dilutive conduct's more negative effect on subsequent consumer recognition and recall of famous marks. *See, e.g., Swann, supra* note 15, at 606-14; Jacoby, *Psychological Foundations, supra* note 17 at 1049; Jacoby, *Dilution in Light of Victoria's Secret, supra* note 15, at 9, 20-21; *see also* Maureen Morrin & Jacob Jacoby, *Trademark Dilution: Empirical Measures for an Elusive Concept*, 19 J. PUB. POL'Y & MARKETING 265 (2000). Such attempts have rightly drawn criticism for being methodologically imprecise and devoid of real-world context. *See* Tushnet, *supra* note 20, at 528-46. Not all behavioral studies suffer from these specific flaws. Although consumer survey data is notoriously imprecise and subject to manipulation, as Tushnet details, new research techniques offer the potential of more precise methods to measure consumer motivation in purchasing. *Id.* at 544-46; *see, e.g., Ambler et al., supra* note 4 (using neuroimaging techniques to examine relative activity in brain areas associated with different processing tasks during shopping activity); Punj & Hillyer, *supra* note 4 (same). These techniques have been especially useful for gauging the influence of emotion on economic decision-making. *See* Colin Camerer et al., *Neuroeconomics: How Neuroscience Can Inform Economics*, 43 J. ECON. LITERATURE 9, 14 (2005) (stating that because different parts of the brain are more or less associated with affective or cognitive processing, brain imaging studies done while people engage in different kinds of economic tasks can provide insights about the mix of affective and cognitive processes in these tasks); P. Kenning & H. Plassmann, *NeuroEconomics: An Overview from an Economic Perspective*, 67 BRAIN RES. BULL. 343, 343, 352 (2005). Economic studies using historical purchase data

consumers are complex creatures with divergent and sometimes competing motivations. Trademark law has suffered in recent years from an automatic equation of consumer uncertainty or negative response to a legally cognizable injury. Consumers value convenience in the short term and greater choice in the long term. There are two types of search costs: (i) the costs of finding goods from a known producer, and (ii) more broadly, the costs of identifying options for any consumptive choice. Sometimes, as with dilution regulation, lowering the first kind of search cost may increase the second.<sup>41</sup> Acknowledging dueling consumer interests calls into question trademark law's reliance on momentary snapshots of consumer

---

following changes in branding strategy also offer relatively unbiased sample sets through which to examine consumer behavior. *See, e.g.,* Tülin Erdem, *An Empirical Analysis of Umbrella Branding*, 35 J. MARKETING RES. 339, 347 (1998) (using scanner data for toothbrush and toothpaste purchases to show a negative effect on parent brand purchases after the introduction of a poor quality brand extension); Vanitha Swaminathan, *Sequential Brand Extensions and Brand Choice Behavior*, 56 J. BUS. RES. 431 (2003) [hereinafter, Swaminathan, *Sequential Brand Extensions*] (household scanner panel purchase data); Vanitha Swaminathan et al., *The Impact of Brand Extension Introduction on Choice*, 65 J. MARKETING 1, 12-14 (2001) [hereinafter, Swaminathan, *The Impact of Brand Extension Introduction*] (same). Although no research method is flawless and uncertainties remain even with these relatively more objective measures, lawmakers need not wait for perfect certainty to gain insights from cognitive and behavioral studies. For one thing, because trademark laws rest on the utilitarian justification of improving consumer welfare, some model of consumer behavior is necessary to judge the effects of different proposals. Models informed by credible research are superior to naked assumptions drawn from the personal experience of a reviewing judge. Indeed, Professor Tushnet is not even uniformly critical of the cognitive and behavioral research so much as she is critical of how they have been used by advocates of dilution protection. In several places, she herself relies on cognitive studies to refute assertions by dilution proponents. Tushnet, *supra* note 20, at 536-42 (citing studies showing a positive effect on memory and a liking for multiple and varied uses of familiar terms, called "reaffirmation effects"); *id.* at 543 n.177 (arguing that poor quality brand extensions are unlikely to harm sales of the senior brand because of research studies demonstrating an absence of negative spillover effects to parent brands where consumers were given a way to distinguish the extension from the senior brand, such as through sub-branding).

41. A second example would be prohibitions on uses of trademarks in keyword advertising. Some trademark owners have sought to prohibit unrelated companies from purchasing ads designed to appear when the owner's trademark is used as a search term on an internet search engine. For example, a search for "Delta Airlines" might call up sponsored links for United and Southwest. Purveyors of complementary products, such as travel insurance or travel guides, might advertise as well. So long as the ads are labeled as such, most consumers find the links more helpful than distracting. They do, however, provide more clutter on a user's screen after a search and can momentarily delay the consumer's clicking over to the desired vendor. *See, e.g.,* Brookfield Commc'ns v. W. Coast Entm't Corp., 174 F.3d 1036, 1064 (9th Cir. 1999) (finding a momentary diversion of keyword advertising actionable as an initial interest confusion).

perception by themselves as conclusive arbiters of long-term welfare. An emphasis on fleeting reactions, which tend to reflect short-term interests, fails to reconcile convenience with long-term interests in enhanced competition. Accordingly, this Article argues that behavioral insights are properly understood as data to be weighed in constructing rules, but that consumer perceptions by themselves can no longer remain the sole indicators of legal injury in trademark law.

## II. TRADEMARKS, INFORMATION, AND COMPETITION

### A. The Informational Purpose of Brands

The goal of trademark law is broadly accepted as improving the quality of information in the marketplace.<sup>42</sup> Trademarks are an efficient and simple means of communicating information.<sup>43</sup> Sellers use advertising and trade symbols to inform likely buyers about desirable qualities and characteristics of their goods.<sup>44</sup> Trademarks ensure that consumers associate these characteristics with the right product.

For the consumer, one brand name can serve as a repository for “a complex constellation of associations and images that comprises a consumer’s knowledge of the brand and his attitudes towards it.”<sup>45</sup> Any relevant information about the underlying good, including advertising claims, community reputation, and the individual’s previous experience with that product becomes associated with the trademark and is easily accessible to the consumer upon encountering the mark in commerce.<sup>46</sup> Consumers can rely on source indicators to quickly locate goods and services that match their tastes for quality and price.<sup>47</sup>

---

42. Dogan, *supra* note 11, at 106 (“[T]rademark law has never aimed to provide exclusive rights in marks, but has focused on preserving informational clarity in the marketplace.”); *see also* Kratzke, *supra* note 11, at 216-17 (trademark law should grant exclusive rights in trademark interests to facilitate the transmission of informational and identifying messages); Lunney, *supra* note 8, at 431-32.

43. Mark Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1688 (1999).

44. Kratzke, *supra* note 11, at 216; Lunney, *supra* note 8, at 432; Nelson, *Advertising as Information*, *supra* note 27, at 735.

45. Kratzke, *supra* note 11, at 204-05 (citing John F. Coverdale, Comment, *Trademarks and Generic Words: An Effect-On-Competition Test*, 51 U. CHI. L. REV. 868, 875 (1984)); *see also* Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 16 (“For consumers, a brand name functions as a core node around which other information in memory is connected and around which new information can be organized.”).

46. *See* LANDES & POSNER, *supra* note 12, at 167; Kratzke, *supra* note 11, at 207.

47. LANDES & POSNER, *supra* note 12, at 166-68.

This interest in informational clarity serves both a public and private purpose. Glynn Lunney has explained how trademarks improve allocative efficiency<sup>48</sup> in the marketplace by allowing consumers to efficiently express preferences to producers:

By enabling consumers to connect information to precise product[s] more accurately, trademarks help consumers express more accurately their preferences and tastes for the varying mix of product features, quality and prices that each finds desirable. Trademarks can, therefore, help ensure that the pricing signals received by producers from the market (or “expressed demand”) more accurately reflect consumers’ actual tastes and preferences (or “actual demand”).<sup>49</sup>

This feedback loop ideally facilitates an efficient marketplace that produces the socially optimal types and amounts of goods at the best prices.<sup>50</sup> In this way, the consumer acts as a proxy for the interests of society in distributing resources efficiently.<sup>51</sup>

Because of the importance of this information exchange, trademark law prohibits behavior that confuses the information or identification function of trade symbols in the mind of the consumer. Symbols that deceive consumers reduce the efficiency of the market by causing consumers to unwittingly purchase a different good.<sup>52</sup> A system that does not protect the

48. Allocative efficiency is the market condition whereby resources are allocated in a way that maximizes the net benefit attained through their use. Allocative efficiency refers to a situation in which the limited resources of a country are allocated in accordance with the wishes of consumers. An allocatively efficient economy produces an optimal mix of commodities from the general consumer perspective. See Lunney, *supra* note 8 at 444 & n.275; Glynn S. Lunney, Jr., *Reexamining Copyright's Incentives-Access Paradigm*, 49 VAND. L. REV. 483, 489, 598 (1996).

49. Lunney, *supra* note 8, at 432; see also *Smith v. Chanel, Inc.*, 402 F.2d 562 (9th Cir. 1968). The court stated:

[Trademark] makes effective competition possible in a complex interpersonal marketplace by providing a means through which the consumer can identify products which please him and reward the producer with continued patronage. Without some such method of product identification, informed consumer choice, and hence meaningful competition in quantity could not exist.

*Id.*

50. Kratzke, *supra* note 11, at 216 (noting that advertising helps to make the market efficient because it enables consumers to transmit accurate messages concerning their choices of resource allocation); Lunney, *supra* note 8, at 432.

51. Kratzke, *supra* note 11, at 212.

52. *Id.* at 272; see also Stacey L. Dogan & Mark A. Lemley, *Trademarks And Consumer Search Costs on the Internet*, 41 HOUS. L. REV. 777, 788-89 (2004) (stating that

informational integrity of trade symbols causes consumers to expend extra resources in searching for desired goods, and thus increases “search costs.”<sup>53</sup>

Regulation of trade symbols is not without its own costs. Trademark protection increases information costs for competitors and new entrants who may lack significant advertising capital. In the absence of protected trademarks, consumers would need to re-evaluate each product choice with every purchase.<sup>54</sup> If they sought a quality cereal with a pleasing shape, for example, they could not rely on the shorthand of the General Mills and Cheerios, and their previous experience with these brands to guide them. They would have to evaluate the price, ingredients, nutrition, shape, color and whatever other attributes were easily ascertainable for each product every time they went to the store.<sup>55</sup> This would cost consumers in time, but would allow a new competitor more easily to capture consumer attention to promote attributes of the new product that may be preferable to more established choices. We allow this disadvantage to new entrants because we believe it is outweighed by the informational benefits of trademarks in the form of lower search costs.<sup>56</sup>

## **B. Trademark Dilution and Internal Confusion**

Legal prohibitions have also been extended to behavior that lessens the “selling power” of previously popular marks, or “trademark dilution.”<sup>57</sup> Trademark dilution laws prohibit the use of famous marks on non-competitive goods when such use is likely to lessen the distinctiveness of the mark. Commentators have defined dilution’s harm as one of “internal search costs.”<sup>58</sup> Internal search costs are created when the presence of additional users of distinctive marks forces consumers to work harder to remember the original mark and to connect it with its associated goodwill.<sup>59</sup>

---

trademark law maximizes efficiency by lowering search costs and has “historically limited itself to preventing uses of marks that ‘defraud[ed] the public’ by confusing people into believing that an infringer’s goods were produced or sponsored by the trademark holder).

53. *See, e.g.*, *Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159, 163-64 (1995) (“Trademark law . . . ‘reduce[s] the customer’s costs of shopping and making purchasing decisions.’”); LANDES & POSNER, *supra* note 12, at 168.

54. Landes & Posner, *supra* note 8, at 269.

55. *See id.*

56. Kratzke, *supra* note 11, at 208; LANDES & POSNER, *supra* note 12, at 168.

57. Schecter, *supra* note 16, at 831.

58. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 20; *see* Tushnet, *supra* note 40, at 518 (describing prevailing cognitive model for dilution law as resting on the notion of increased “internal” search costs).

59. LANDES & POSNER, *supra* note 11, at 207.

By distracting consumers from their original associations with branded goods, dilution thus diminishes the efficiency of the marketplace.

Trademark dilution can happen in two ways, either through blurring or tarnishment. The recent Trademark Dilution Revision Act defines dilution by tarnishment as an “association arising from the similarity between a mark or trade name and a famous mark that harms the reputation of the famous mark.”<sup>60</sup> Tarnishment has typically been understood as the connection of a famous mark with unsavory products and services, such as those promoting sex, drugs, or violence.<sup>61</sup>

Of the two, blurring has caused more head-scratching among courts and commentators.<sup>62</sup> Blurring is defined in the Trademark Dilution Revision Act as “association arising from the similarity between a mark or trade name and a famous mark that impairs the distinctiveness of the famous mark.”<sup>63</sup> The cause of action is deceptively simple on its face: it is aimed at the use of famous marks (like Google) by unrelated users on a new class of goods or services (e.g., Google Petroleum). Consumers are unlikely to think that the oil producer is run by the search engine. Yet the presence of the second mark may lessen the ability of the first to serve as a distinctive identifier for the search engine.

Blurring has proven difficult to define explicitly because of the lack of specificity in the statutory definition as to what kind of “association” will be enough to impair a trademark’s “distinctiveness.” Distinctiveness is a term of art in trademark law that refers to a word’s or a symbol’s ability to uniquely identify one source or producer for goods or services.<sup>64</sup>

---

60. Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, § 2, 120 Stat. 1730, 17301 (to be codified at 15 U.S.C. § 1125(c)(32)(B)).

61. Long, *supra* note 21, at 1057.

62. See, e.g., Beebe, *A Defense*, *supra* note 19, at 1148-49 (discussing problems in the interpretation of trademark blurring); Farley, *supra* note 18, at 109-110 (discussing problems in defining dilution by blurring).

63. Trademark Dilution Revision Act, § 2 (to be codified at 15 U.S.C. § 1125(c)(2)(B)).

64. It is a relative term. An invented word like “Kodak” is more inherently “distinctive” than a descriptive word like “pretty” to designate the source of a product. This is because consumers are more likely to recognize it as a designator of source than a description of the product’s features. However, even descriptive words and pictures can “acquire” distinctiveness by becoming widely associated in the public mind with a specific commercial source. Thus “American Airlines” is a distinctive trademark despite the fact that its terms are entirely descriptive of any U.S. consumer aviation company. This understanding of distinctiveness does little to illuminate what kind of conduct the blurring cause of action aims to discourage. Trademark’s taxonomy of “distinctive” marks suggests an impairment of distinctiveness occurs when a mark is rendered less capable of serving as a unique source identifier. This might suggest uses of a mark to refer to an

Some courts have held that the mere fact of association of one mark with another is enough to “impair distinctiveness.”<sup>65</sup> The distinctiveness of the senior mark is diminished if it no longer brings to mind the senior user alone.<sup>66</sup> The Supreme Court in *Moseley v. V Secret Catalogue Inc.* criticized this simplistic equation by stating that “[b]lurring’ is not a necessary consequence of mental association.”<sup>67</sup> The Court required some showing that the association was likely to “reduce the capacity of the famous mark to identify the goods of its owner.”<sup>68</sup> However, in 2006, Congress revised the dilution act in the wake of the *Moseley* decision, and seemingly returned to an emphasis on pure mental association by including it in two of the six factors used to gauge the existence of “blurring.”<sup>69</sup>

---

entire class of goods or services rather than those of just one producer. *E.g.*, *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4 (2d Cir. 1976) (defining a taxonomy of distinctiveness in which generic terms for classes of goods have the least amount of distinctiveness). The statutory definition of dilution negates any such reading; the cause of action is limited to uses as “designations of source.” Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, § 2, 120 Stat. 1730, 1730 (to be codified at 15 U.S.C. § 1125(c)(1) & (3)). This statutory formulation has historically been aimed at uses of marks on non-similar goods. *See* Schechter *supra* note 16, at 831 (proposing a new dilution cause of action aimed at uses on non-competing goods). Or the statute might aim at uses that undercut “acquired” distinctiveness by lessening the public’s learned understanding that the mark signifies one unique producer. Reading distinctiveness to refer to the public’s acquired understanding fits well with the statute’s mandate to only protect only already “famous” marks. However, trademark law has not historically wrestled with this notion of distinctiveness from others, as opposed to distinctiveness of source per se. Trademark law’s tools for measuring source distinctiveness are ill-designed for this new kind of inquiry. *See* Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. REV. 621, 702 (2004).

65. *Mattel, Inc. v. MCA Records, Inc.*, 296 F.3d 894, 904 (9th Cir. 2002) (“The distinctiveness of the mark is diminished if the mark no longer brings to mind the senior user alone.”).

66. *Id.*

67. *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 434 (2003).

68. *Id.*

69. Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, § 2, 120 Stat. 1730, 1730 (to be codified at 15 U.S.C. § 1125(c)(32)(B)). The other factors include: (i) degree of similarity between the marks, (ii) degree of inherent or acquired distinctiveness of the famous mark, (iii) the extent to which the owner of the famous mark is engaging in substantially exclusive use of the mark, (iv) the degree of recognition of the famous mark. The two factors concerning association are (v) whether the defendant intended to create an association with the famous mark, (vi) any actual association between the marks. *Id.*; *see also* *The Hershey Co. v. Art Van Furniture*, No. CV-14463, 2008 WL 4724756 at \*15 (E.D. Mich. Oct. 24, 2008) (finding a likelihood of blurring where defendant moving van company intended to create an association with the Hershey bar trade dress by portraying a couch wrapped in similar-looking trade dress on the side of its vans).

Efforts to explain why a mental association between two marks will lessen the source-identifying potency of the senior mark have tended to focus on what Professor Frank Schechter, the original proponent of dilution regulation, called a mark's "selling power."<sup>70</sup> Schechter equated selling power with uniqueness.<sup>71</sup> Modern trademark theorists have refined this definition to account for a mark's ease of recall in memory and ability to convey product-specific information.<sup>72</sup> Dilution thus occurs when a "mark's propensity to bring to mind a particular product or source is weakened."<sup>73</sup> Proponents claim that blurring interferes with the speed and accuracy of recall of a senior mark because encumbering a brand with divergent associations can impair its memorability.<sup>74</sup> Consumers will no longer register an immediate impression related to the original source when they see the mark.<sup>75</sup>

Judge Posner offered what is perhaps the most influential explanation of trademark blurring harm in *Ty Inc. v. Perryman*:

There is concern that consumer search costs will rise if a trademark becomes associated with a variety of unrelated products. Suppose an upscale restaurant calls itself "Tiffany." There is little danger that the consuming public will think it's dealing with a branch of the Tiffany jewelry store if it patronizes this restaurant. But when consumers next see the name "Tiffany" they may about both the restaurant and the jewelry store, and if so the efficacy of the name as an identifier of the store will be diminished. Consumers will have to think harder—incur as it were a higher imagination cost—to recognize the name as the name of the store . . . So "blurring" is one form of dilution.<sup>76</sup>

Marketing experts have provided psychological models that explain why follow-on uses of a mark can cause an increase in internal search times. These explanations focus on the ability of trademarks to serve as

---

70. Schechter, *supra* note 16, at 819.

71. *Id.*

72. Jacoby, *Dilution in Light of Victoria's Secret*, *supra* note 40, at 1048; Morrin & Jacoby, *supra* note 40, at 274; Swann, *supra* note 15, at 624.

73. *Jordache Enters., Inc. v. Hogg Wyld, Ltd.*, 828 F.2d 1482, 1489 (10th Cir. 1987).

74. Beebe, *A Defense*, *supra* note 19, at 1148 (stating that blurring is when D's use of a mark is similar or identical to P's and blurs the link between the P's mark and the goods or services to which the mark is traditionally attached); Jacoby, *Dilution in Light of Victoria's Secret*, *supra* note 15, at 20; Swann, *supra* note 15, at 620, 624.

75. Jerre B. Swann, Sr., *Dilution Redefined for the Year 2000*, 37 HOUS. L. REV. 729, 751 (2000).

76. *Ty Inc. v. Perryman*, 306 F.3d 509, 511 (7th Cir. 2002).

“core nodes” in memory around which consumers organize all of their information about the product and brand.<sup>77</sup> Follow-on uses, even when not confusing, add new associations to this cognitive network that reduce the strength of pre-existing brand beliefs.<sup>78</sup> For example, the brand Nike quickly brings to mind “shoes,” and also perhaps qualities associated with these shoes such as “sports,” “winning,” “achievement,” and “success” for a large segment of the purchasing public. In the Nike example, if someone uses the mark for couches, then people might be less likely immediately to think of “shoes,” and all the associated positive properties of those sneakers, the next time they see the mark. Thus, although the consumer has not been misled by externally deceptive information, she will now take longer, and require more effort, to sort relevant from irrelevant associations.<sup>79</sup>

Professor Barton Beebe has called this definition “empirical in orientation.”<sup>80</sup> For a judge to find that a junior mark “blurs” a senior mark, the judge must find that the junior mark causes consumers to “think for a moment” before recognizing that the senior mark refers to the goods of the senior mark’s owner.<sup>81</sup> In this formulation, dilution by blurring increases “search costs” in a measurable way: consumers must take more time and expend more cognitive resources to connect the mark to the senior user’s goods and larger goodwill.<sup>82</sup>

---

77. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 16-17; Jacoby, *Psychological Foundations*, *supra* note 1740, at 1024.

78. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 20; Jacoby, *Psychological Foundations*, *supra* note 40, at 1049.

79. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 20-21.

80. Beebe, *A Defense*, *supra* note 62, at 1149

81. *Id.*

82. There is some disagreement as to whether blurring extends just to a consumer’s memory of the proper product category, e.g. Audi for cars and not shoes, or to the whole constellation of associations with a mark that make up what practitioners refer to as goodwill. In this broader formulation, use of Audi for shoes would interfere not only with recall of product category but also specific product attributes such as “speed” and high-tech engineering. Compare Beebe, *A Defense*, *supra* note 19, at 1148-49 (arguing that blurring protects only memory of product type) with Swann, *supra* note 75, at 750-53 (arguing that blurring extends brand connotations and experiential “promise”). Attempts to limit the notion of dilution just to information about product category seem like previous efforts to distinguish between “informational” and “persuasive” functions of trademarks: attractive as a theory but nearly impossible to put into practice. For example, the more narrow conception of dilution as protection of product class association would seem to exclude any protection for well-known marks such as Sony or Virgin that already are associated with a variety of product classes.

This existing “cognitive” account of trademark dilution by blurring is purely informational. As one commentator has noted about dilution by blurring:

[T]here is no evaluative component; there is no “I like” or “I dislike,” there is no “that’s good,” or “that’s bad.” The association is simply informational. Where there used to be a single commercial association, [a] name is now associated with two commercial entities . . . increasing the consumer’s mental search costs.<sup>83</sup>

However, it is implicit in this definition that increased search costs will cause consumers to change their behavior in inefficient ways. In deciding that insufficient evidence of actual dilution was presented in the *Moseley* decision, the Court focused on what Victoria’s Secret had not shown: that a consumer who saw the allegedly diluting mark “Victor’s Little Secret” did not “form any different impression” or in any way “change his conception” of the more well-known Victoria’s Secret.<sup>84</sup> The Court linked a lessening of the capacity to identify and distinguish between brands with an involuntary alteration in the consumer’s conception of the senior brand. Similarly, dilution theorists point to cognitive research studies connecting these increased internal search times with decreased sales.<sup>85</sup> Implicitly then, the harm flowing from “internal search costs” is an involuntary change in the consumer’s preference for goods bearing a certain brand. What is missing in trademark decisions and commentary is a clear explanation of why mental association of two unrelated businesses, which may infinitesimally increase consumer’s internal mental search times, is likely to cause any such change in conception or behavior.

Many commentators have questioned whether the mere association of one item with another is likely to cause either internal confusion or a

---

83. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 22. Dilution has thus been described as an impairment to the brand’s ability to convey information. Swann, *supra* note 15, at 617. At the extreme, commentators have suggested that the impairment to a brand’s informational value is worse than confusion. In a straight infringement case, consumers understand what the mark signifies, but mistakenly believe the defendant’s goods also exhibit those properties. Assuming the infringement is halted before becoming widespread, the information conveyed by the mark never changes. In the dilution case, the second user alters the specific information conveyed by the mark by introducing irrelevant associations, so that consumers can no longer rely on the mark in any situation for useful product information. *Id.*

84. *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 434 (2003).

85. Jacoby, *Dilution in Light of Victoria’s Secret*, *supra* note 15, at 20-21 (citing studies claiming that first brand awareness is best predictor of brand switching behavior).

change in consumer behavior.<sup>86</sup> Brands like Ford and Nissan have coexisted for years with twins or near twins in the modeling and bakery industries and neither seems to have suffered much for it.<sup>87</sup> It seems entirely plausible that over time, the association between the two producers would dissipate as consumers learn to distinguish two similarly named but non-competing businesses. As Professor Rochelle Cooper Dreyfuss has pointed out, most of us have friends and relatives who share similar names, and yet we are capable of distinguishing our friend Dave from our cousin Dave.<sup>88</sup> Furthermore, the context in which we encounter such names is likely to counteract any distraction offered by the second user. As Professor Rebecca Tushnet notes, when we say “Delta” in the back of a cab with our luggage on the seat next to us, the cab driver is unlikely to think we want to visit the plumbing supplier.<sup>89</sup>

Further, it’s not clear that simple association between two entities is harmful. Imitation is the sincerest form of flattery. As Professor Graeme Austin has written:

Implicit in Judge Posner’s approach is the idea that consumers care that they must think harder. . . . If dilution imposes an imagination “cost,” it follows that the ordinarily prudent consumer is somebody who prefers to have her imagination unburdened by conflicting messages about brands. But this is not necessarily so, or even more likely so. . . . [It] might . . . be more fun than costly.<sup>90</sup>

---

86. Bone, *Hunting Goodwill*, *supra* note 8, at 559; Welkowitz, *supra* note 18, at 542; Beebe, *A Defense*, *supra* note 19, at 1150, 1165-70; Farley, *supra* note 18, at 110; Ringling Bros.-Barnum & Bailey Combined Shows, Inc. v. Utah Div. of Travel Dev., 170 F.3d 449, 463-65 (4th Cir. 1999).

87. Beebe, *A Defense*, *supra* note 19, at 1150 (“No one can seriously suggest that the typicality of the trademark FORD has been significantly diminished by the coexistence in the American marketplace of a modeling agency—or of millions of people, for that matter—with the same name.”); *see also* Nissan Motor Co. v. Nissan Computer Corp., 378 F.3d 1002, 1014 (9th Cir. 2004) (noting widespread third party use of names identical or similar to “Nissan”). For example, Nissen bread was established in 1899. *See* J.J. Nissen, <http://www.jjnissenbreads.com> (last visited October 7, 2008).

88. Rochelle Cooper Dreyfuss, *Reconciling Trademark Rights and Expressive Values: How to Stop Worrying and Learn to Love Ambiguity*, in *TRADEMARK LAW AND THEORY: A HANDBOOK OF CONTEMPORARY RESEARCH* 261 (Graeme B. Dinwoodie & Mark D. Janis eds., 2008).

89. Tushnet, *supra* note 40, at 529.

90. Graeme W. Austin, *Trademarks and the Burdened Imagination*, 69 *BROOK. L. REV.* 827, 895 (2004).

Other commentators have suggested that the fact that one product references another may strengthen the association of the brand with the senior producer in the minds of consumers.<sup>91</sup>

Finally, many kinds of information are likely to cloud the informational salience of marks; dilution aims at only a trivial subset.<sup>92</sup> For example, uses of trademarks in the context of news reporting, criticism, and parody are mostly exempted from dilution liability, although one could imagine such uses introducing distracting stimuli to individual consumers about their chosen brands.<sup>93</sup> Even the use of products in everyday life is likely to stimulate individual associations that will conflict or drown out any managed brand “personality.”<sup>94</sup> Proponents of dilution law have been unable to articulate a principled rationale that would distinguish the effects of such activities from the blurring harms sought to be prevented.<sup>95</sup> In-

---

91. Tushnet, *supra* note 40, at 536-39; see Chi-Ru Jou, *The Perils of a Mental Association Standard of Liability: The Case Against the Subliminal Confusion Cause of Action*, 11 VA. J.L. & TECH. 2, 58-60 (2006).

92. See, e.g., Tushnet, *supra* note 40, at 547-552.

93. Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, § 2, 120 Stat. 1730, 1730 (to be codified at 15 U.S.C. § 1125(c)(3)) (exempting such uses except where the famous mark is used as a “designation of source” for the second user’s goods or services).

94. Tushnet, *supra* note 40, at 540.

95. This criticism, while true, does not seem fatal. These stimuli may seem distracting to producers, who would like to exclude all negative information from consumer consideration, but dilution seems aimed at only irrelevant information. That is, the baseline for dilution cannot be the idealized image that the producer presents to the consumer in hopes she will adopt it. The baseline must be all of the information that a consumer would normally consider, directly and indirectly, in making a choice. Associations that arise organically due to word-of-mouth, personal experience, or newspaper coverage are arguably relevant to the mark, and are the kinds of things that we as a society want consumers to factor in purchasing decisions. Thus, while such associations may cloud the positive marketing communications that producers would like to attach to the mark, on the whole they increase the overall quality of information in the marketplace and help consumers to make better decisions. The argument against dilution by blurring is thus that it clouds the mark’s salience without offering any relevant product or brand-specific information. The use of the mark by unrelated newcomers to grab attention can be seen as almost entirely random. The use then lowers the mark’s informational salience in a way that is unlikely to aid consumers in making better decisions. That, at least, is the case in favor of regulating dilution by blurring. For further discussion of this point, see *infra* text accompanying note 274. Alternatively, one could frame the exceptions for parody and criticism as an example of legislative social welfare balancing: the benefits to society by allowing free expression of this kind outweighs whatever indirect harms might result from blurring consumer associations. See, e.g., Long, *supra* note 21 at 1065 (advocating a social welfare balancing approach to dilution liability).

deed, trademark owners often sue satirists and critics under a dilution theory.<sup>96</sup>

Courts have also proven wary of such a seemingly untethered cause of action. Professor Clarissa Long has documented how judicial willingness to find liability under the federal statute has waned over the years.<sup>97</sup> Before the enactment of the 2006 Trademark Dilution Revision Act, courts had become so uncomfortable with the ephemeral nature of the harm in federal dilution actions that they required plaintiffs to provide some evidence of actual dilution before they would enjoin use of a mark.<sup>98</sup> Some judges have been downright hostile, claiming that trademark owners “must have had some kind of a lobby” to get such a statute passed.<sup>99</sup> Congress’ recent amendments to the Federal Trademark Dilution Act have clarified some of the definitional uncertainties, but have also eliminated many of the judicial doctrines created to cabin the statute’s reach.<sup>100</sup> Willingness to implement the revised statute remains to be seen.

This Article argues that the failure to communicate a coherent harm for dilution by blurring depends in part on the lack of explicit acknowledgement of the important role that emotion plays in shaping consumer preferences. As detailed more fully below, blurring has the potential to alter our emotional response to familiar trademarks even if it does not alter our conscious appraisal of the underlying product or producer.<sup>101</sup> Both sides of the dilution debate avoid discussion of emotion presumably because of a sense that it is an irrational and untrustworthy basis for decision-making

---

96. *E.g.*, *Louis Vuitton Malletier S.A. v. Haute Diggity Dog, LLC*, 507 F.3d 252 (4th Cir. 2007) (dismissing trademark dilution case against maker of parody dog toy luggage); *Mattel, Inc. v. MCA Records, Inc.*, 296 F.3d 894 (9th Cir. 2002) (same against song critiquing commercialism of Barbie toy).

97. Long, *supra* note 21, at 1030-31 (attributing some of the judicial unwillingness to enforce the doctrine to the ambiguous nature of the harm).

98. *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 433 (2003); *accord* *Ringling Bros.-Barnum & Bailey Combined Shows, Inc. v. Utah Div. of Travel Dev.*, 170 F.3d 449, 453 (4th Cir. 1999) (“[W]e agree with its basic points that “dilution” under the federal Act consists of (1) a sufficient similarity of marks to evoke in consumers a mental association of the two that (2) causes (3) actual harm to the senior marks’ economic value as a product-identifying and advertising agent.”).

99. Transcript of Oral Argument, *Ringling Bros.-Barnum & Bailey v. Utah Div. of Travel Dev.*, 935 F. Supp. 763 (E.D. Va. 1996) (No. 96-788-A).

100. *See generally* Beebe, *A Defense supra* note 19 at 1156- 1161 (noting that the TDRA does away with judicially-created limitations on the dilution cause of action including a requirement that the plaintiff demonstrate “actual dilution” or its mark rather than a “likelihood of dilution” and the requirement in some circuits that a mark be inherently distinctive to receive anti-dilution protection).

101. *See infra* Sections III.C.1 & III.C.2.

and thus a flimsy bedrock for a legal regime.<sup>102</sup> Yet the last twenty years of research on cognitive processing and consumer decision-making suggests that emotion and affect are vital to reasoned deliberation, and that the stigmatization of emotion's part in the trademark debate should be reconsidered.<sup>103</sup> The next Part sets out the two existing accounts of the influence of emotion in trademark use and then suggests a third model that more accurately explains trademark owners' concerns about dilution by blurring.

### III. THREE CONCEPTIONS OF EMOTION IN CONSUMER DECISION-MAKING

The concept of a consumer's voluntary and involuntary response to branded goods has its roots in a decades-old debate about the proper role of emotion in advertising and trademark law. Three schools of thought have developed to explain how consumers are influenced by emotion in their appreciation of advertising and the use of trademarks. These models—the irrational weigher, the rational maximizer, and the cognitive miser—are sketched out below, along with their respective accounts of the benefits and problems of emotional decision-making.

As advertising in the early twentieth century shifted from informative to persuasive strategies, trademarks came to symbolize not just a good's producer but the larger desire-driven claims made in the product's advertising.<sup>104</sup> As advances in communication and transportation made wide-

---

102. One recent exception is an article by Jeremy N. Sheff, *The (Boundedly) Rational Basis of Trademark Liability*, 15 TEX. INTEL. PROP. L.J. 331 (2007). Sheff argues that both confusion and dilution rules are justified as correctives against conduct that would otherwise unfairly manipulate the cognitive biases of consumers, including the pervasive tendency to rely on emotional response as a heuristic, and so render such biases pervasively inaccurate. *Id.* at 333, 358-361. Although I agree with Sheff's description of the goals of dilution law, for the reasons set out further *infra*, I think the question of which consumer preferences are "accurate" and which are manipulated is a much more complicated question.

103. See *infra* text accompanying notes 152-249.

104. Often this imagery had little direct relevance to the advertised product. It was not intended to inform about product choices directly, but to speak to people in general about "what the mass of unregenerative mankind wanted." JOSEPH J. SELDIN, *THE GOLDEN FLEECE: SELLING THE GOOD LIFE TO AMERICANS* 18-19 (Macmillan 1963); see also *Mishawaka Rubber & Woolen Mfg. Co. v. S. S. Kresge Co.*, 316 U.S. 203, 205 (1942) (the object of much modern advertising is "to impregnate the atmosphere of the market with the drawing power of a congenial symbol"). In the markets of the nineteenth century, producers targeted advertising to local consumers already in search of a product. Such consumers typically already knew the names of local producers, and primarily used trademarks to connect existing goodwill with the right seller. Advertising touted informa-

spread distribution of commercial goods possible, producers sought to compete outside local markets.<sup>105</sup> They turned to strategies to build an identity for their wares that would resonate with masses of unknown consumers. Through clever imagery, suggestive turns of phrase, and celebrity endorsers, some trademarks came to symbolize not just a specific producer, but powerful images such as “freedom,”<sup>106</sup> “youth,”<sup>107</sup> “mildness,”<sup>108</sup> or “achievement.”<sup>109</sup>

Trademark scholars of the time disagreed over how much the law should recognize the new persuasive force of trade symbols. Frank Schechter, in an influential 1927 Harvard Law Review article, was the first to argue for a cause of action that would recognize a mark’s “selling power” apart from its value as a source-identifier.<sup>110</sup> Schechter argued that the function of a trademark in modern commerce was no longer to identify a producer but was to “identify a product as satisfactory and thereby to stimulate further purchases by the consuming public.”<sup>111</sup> In Schechter’s formulation, the mark itself sells the goods by its connection to desirable qualities.<sup>112</sup> Thus, Schechter sought to expand the ambit of trademark protection to include not just deceptive uses, but non-confusing uses of the mark on unrelated goods as well.<sup>113</sup> This “dilution” would whittle away the singular identity of the mark and its hold on the public mind.<sup>114</sup> Where

---

tion—sales, promotions, specific qualities—directly relevant to a decision to buy. As railroads, telephones, radios and other devices brought regional and national competition and advertising, the function of advertising changed. Producers now needed to convince strangers to embrace their offerings. Building demand was a special challenge amidst a population accustomed to rural self-sufficiency. *See* Bone, *Hunting Goodwill*, *supra* note 8, at 580. Producers shifted away from informational advertising toward campaigns designed to evoke a more personal connection with the targeted buyer.

105. *See* Bone, *Hunting Goodwill*, *supra* note 8, at 580.

106. Marlboro, at least for men. ROBERT HEATH, *THE HIDDEN POWER OF ADVERTISING: HOW LOW INVOLVEMENT PROCESSING INFLUENCES THE WAY WE CHOOSE BRANDS* 27 (2001); *see* BRAND EQUITY & ADVERTISING: ADVERTISING’S ROLE IN BUILDING STRONG BRANDS 126-127 (David A. Aaker & Alexander L. Biel eds., 1993).

107. Mountain Dew. The brand also claims to embody energy and exhilaration. *See* Theresa Howard, *Being True to Dew*, BRANDWEEK, Apr. 24, 2000, at 28.

108. Neutrogena. *See* Barbara Loken & Deborah Roedder John, *Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact?*, 57 J. MARKETING 71, 742 (1993).

109. Nike. HEATH, *supra* note 106, at 89 (Nike slogan encapsulates the concept of achievement).

110. Schechter, *supra* note 16 note at 819.

111. *Id.* at 818.

112. *Id.* at 819.

113. *Id.* at 831.

114. *Id.* at 825.

a mark owner had spent vast resources cultivating an image of excellence for a particular product, Schechter thought that the advertising impression should be insulated from misappropriation to the same extent as the use of the producer's plant or machinery.<sup>115</sup>

#### A. Irrational Weighers: Consumers Misled By Emotion Make Bad Choices

In the 1940s, a second school of thought arose that was more skeptical of the influence of persuasive advertising on purchasing behavior. This "Irrational Weigher" school argued that emotional advertising manipulated consumers into overvaluing branded goods.<sup>116</sup> Professor Ralph Brown made the most prominent of these critiques in an influential 1948 article titled *Advertising and the Public Interest*. He distinguished between "informative" advertising (about price and attributes) and "persuasive" advertising that sought to "impregnate the brand with these powerful emotional stimulants, and so influence consumers to spend money on impulsive, unnecessary and expensive branded goods."<sup>117</sup> Brown argued that such ads used a "bewildering manipulation" of emotional urges to mesmerize consumers into assessing identical goods differently based on an illusory brand personality.<sup>118</sup> This emotional response led people to pay a premium for goods that did not differ significantly from more generic choices, and so undercut market efficiency.<sup>119</sup>

To Brown, a trademark's only legitimate function was as a designator of source. He saw no value in the irrational emotional appeal of the brand

115. *Id.* at 829-30. Schechter advocated such protection only for "coined" trade symbols, i.e., words that the producers had made up themselves instead of using existing terms in the language. In 1927, these were the only kinds of symbols that could be fully protected as "trade marks."

116. *See, e.g.*, Brown, *supra* note 7, at 1171-75; *see also* CHAMBERLAIN, *supra* note 7, at 246-50; Sigmund Timberg, *Trade-Marks, Monopoly and the Restraint of Competition*, 14 *LAW & CONTEMP. PROBS.* 323, 325-26 (1949).

117. Brown, *supra* note 7, at 1171-75.

118. *Id.* at 1182. Further, "[t]he classical economists who enthroned the consumer never dreamed that he would be making his decisions under a bombardment of stupefying symbols." *Id.* Because of manipulative power of persuasive advertising, the consumer is choosing only between "one illusion and another. . . . Persuasive advertising is, for the community as a whole, just a luxurious exercise in talking ourselves into spending our incomes." *Id.* at 1182-83. Brown undoubtedly was influenced by the then-dominant behavioral school of psychology. Behaviorists focused on environmental stimuli and stressed the power of "reinforcement learning," which involved learning by repetitious exposure to an idea. Such theories "fed fears of advertising's power to brainwash consumers during the 1940s and 1950s." Bone, *Hunting Goodwill*, *supra* note 8, at 603 n.318.

119. Brown, *supra* note 7, at 1171-75.

and opposed protection of its commercial magnetism. Brown denounced dilution laws as the protection of irrational desires stimulated by advertising that conflicted with actual consumer interest.<sup>120</sup> By legitimating the claims that advertised brands delivered benefits like cachet or sex appeal unavailable from functionally equivalent and cheaper offerings, such laws helped to insulate established producers from competition.<sup>121</sup> Brown was not alone in hoping that dilution of famous brands would occur and so force consumers to more carefully analyze how they spent their hard-earned dollars.<sup>122</sup>

This critique persists in modern scholarship.<sup>123</sup> For example, modern commentators have proposed that prohibitions against dilution are inefficient because they promote the creation of brand personalities that appeal to consumer hearts rather than minds.<sup>124</sup> These commentators have translated Brown's admonition to limit protection to the informative component of brands into claims that dilution laws should protect only against a lessening of the connection of a famous mark to its underlying product, but not extend to any lessening of the unique personality of the mark itself or its larger emotion-laden goodwill.<sup>125</sup>

---

120. *Id.* at 1187-94.

121. *Id.* at 1184-85 (urging a legal distinction between the informational and persuasive function of trade symbols); *Id.* at 1187-94 (denouncing dilution as unwarranted protection of the persuasive function).

122. *Id.* at 1204. Contemporary economists agreed that consumers, once so manipulated, were drawn by blind habit back to the same goods so that emotively charged trademarks could confer a monopoly-like power. CHAMBERLIN, *supra* note 7, at 272-74; Timberg, *supra* note 116, at 326 (arguing that trademarks influence consumers to make economically irrational purchasing decisions). In the 1930s and 1940s, some commentators supported grade marks that carefully described exact quantity and quality as a superior alternative to trademarks. *See, e.g.*, LEVERETT S. LYON ET AL., GOVERNMENT & ECONOMIC LIFE 362-78 (Victor Abramson ed., 1939); MARGARET G. REID, CONSUMERS AND THE MARKET (1939); Carl A. Auerbach, *Quality Standards, Informative Labeling, and Grade Labeling as Guides to Consumer Buying*, 14 LAW & CONTEMP. PROBS. 362 (1949).

123. *E.g.*, Robert N. Klieger, *Trademark Dilution: The Whittling Away of the Rational Basis for Trademark Protection*, 58 U. PITT. L. REV. 789, 856-63 (1997).

124. *Id.* at 857; Bone, *Hunting Goodwill*, *supra* note 8, at 620 (arguing that protecting positive feelings as part of a seller's goodwill is socially wasteful and unnecessary); Kratzke, *supra* note 11, at 222 (arguing that protection of trademark value beyond its informational function distorts competition); *see also* Lunney, *supra* note 8, at 437-39 (arguing that to the extent that protection of a firm's goodwill includes protection of the emotional content of the good, the benefits of such protection are not worth the costs).

125. Beebe, *A Defense*, *supra* note 19, at 1148-49; Bone, *Hunting Goodwill*, *supra* note 8, at 552-53, 622 (criticizing the extension of trademark protection beyond product

## B. Rational Maximizers: Useful Product Information Influenced by Emotion

More recently, the economics literature has taken up Schechter's cause and defended the protection of persuasive advertising and the emotional valence of trade symbols. This "Rational Maximizer" literature offers two reasons for accepting the emotional connection with brands. First, maximists argue that emotional appeal is part of the product or service sold under the brand.<sup>126</sup> If consumers believe themselves better off because a certain car or candy might make them more attractive, then purchase of such items will increase subjective well-being.<sup>127</sup> In this view, "emotional" preferences are not different from, and are no less valid than, utilitarian ones, and consumer welfare increases with the facilitation of either kind of choice. Second, economists have provided evidence that persuasive, emotion-based advertising provides important informational cues to consumers, even if consumers do not fully understand how they use the information. For example, advertising can raise price elasticity simply by alerting consumers to new choices.<sup>128</sup> Adherents to this "advertising as information" school then tend to support regulation of dilution as useful protection against clutter—against the devaluation of the information, both for identification and persuasion purposes, which trade symbols provide. These two reasons are explored more fully below.

First, the rational maximizer school of thought argues that little reason exists to prefer choices based on product functions and price over less tangible "emotional" responses to brands.<sup>129</sup> If consumers find more value in goods for the status they confer, for their ability to link purchasers to desired communities, or for other abstract emotional properties, then the

---

and firm goodwill to encompass the "inherent goodwill" of the mark); Lunney, *supra* note 8, at 437.

126. Swann, *supra* note 15, at 594.

127. See, e.g., Dilbary, *supra* note 8 (arguing that consumers derive value from the intangible auras of branded products).

128. Lee Benham, *The Effect of Advertising on the Price of Eyeglasses*, 15 J.L. & ECON. 337, 337-52 (1972); Phillip Nelson, *The Economic Consequences of Advertising*, 48 J. BUS. 213, 219, 224 (1975) [hereinafter Nelson, *Economic Consequences*]; Nelson, *Advertising as Information*, *supra* note 27, at 732 (hypothesizing that the more brand advertising a consumer encounters, the more likely he is to try the advertised brand).

129. Nelson, *Economic Consequences*, *supra* note 128, at 213 ("The idea that advertising changes tastes seems to have a particular appeal to advertising's critics. But this idea is consistent with advertising operating in perfectly competitive markets and with advertising improving welfare. . . . We economists have no theory of taste changes, so this approach leads to no behavioral predictions.").

market should be free to respond to these values.<sup>130</sup> People consume goods and services for more than just utilitarian reasons; we use acquired goods to scale the last three rungs of Maslow's pyramid (social needs, ego needs, and self-actualization).<sup>131</sup> A woman who believes L'Oreal hair dye makes her "worth it," or a man who believes smoking Marlboros makes him more macho may derive more satisfaction from using these products than they would from using generic substitutes.<sup>132</sup> These "emotional" perceptions are no less rationally self-interested than logical preferences tied to quality and price.<sup>133</sup> Society has no reason to value one set of preferences over another, so long as the consumer receives the satisfaction she wants from using the advertised good.<sup>134</sup>

Because emotional preferences can create utility for consumers, rational maximizers argue that the expression of these preferences is just as important to an efficient marketplace as the expression of functional preferences.<sup>135</sup> Newcomer brands that free-ride on positive brand associations may change these associations by adding new connotations.<sup>136</sup> The new brand's associations may make the senior brand less appealing to its core constituency and so may mislead those consumers into changing their preferences.<sup>137</sup>

Second, some studies suggest that advertising (regardless of the content of the advertisement) is itself a reliable indicator of quality.<sup>138</sup> Heavy

---

130. Dilbary, *supra* note 8, at 661, 664 (arguing that persuasive advertising creates a separate intangible value for consumers apart from the underlying physical product and that allowing producers to capitalize on that additional value through premium pricing is not anti-competitive).

131. Swann, *supra* note 15, at 618.

132. Jessica Litman, *Breakfast with Batman: The Public Interest in the Advertising Age*, 108 YALE L.J. 1717, 1730 (1999).

133. Swann, *supra* note 15, at 618 ("Consumers thus logically, rationally and emotionally are interested in more than quality and price.").

134. Bone, *Hunting Goodwill*, *supra* note 8, at 603; Nelson, *Economic Consequences*, *supra* note 128, at 213.

135. Dilbary, *supra* note 8, at 607-08, 622-28, 663-64; *cf.* Swann, *supra* note 15, at 618 ("Mental as well as physical expectations are entitled to insulation.").

136. Dilbary, *supra* note 8, at 664.

137. Swann, *supra* note 15, at 615 ("Any interference with consumers' associations with, or feelings about, such a brand would undermine its essence."); *see also* Dilbary, *supra* note 8, at 664-65 (dilution law protects consumer investment in externality of brand image).

138. Nelson, *Economic Consequences*, *supra* note 128, at 214; Nelson, *Advertising as Information*, *supra* note 27, at 732; LANDES & POSNER, *supra* note 12, at 174; *cf.* Benjamin Klein & Keith B. Leffler, *The Role of Market Forces in Assuring Contractual Performance*, 89 J. POL. ECON. 615 (1981) (presence of nonsalvageable capital is a means of enforcing quality promises).

advertising of a brand suggests lower risk because it signifies a level of investment in the brand by the underlying firm.<sup>139</sup> A firm that devotes substantial advertising revenue and additional product lines to a single brand uses its accumulated investment, in addition to the expected future cash flows from products under the brand, as collateral to guarantee future quality consistency.<sup>140</sup> If the firm ceases to police quality or introduces inferior products to its line, its entire investment in the brand is jeopardized.<sup>141</sup> Therefore, consumers rationally place higher confidence in brands that appear to have a greater backing because they represent a lower risk of inconsistent quality.<sup>142</sup>

The safety valve of this information model is the consumer's rational self-interest.<sup>143</sup> The neoclassical economic model assumes that consumers will respond to advertising only so long as it gives them reliable information at less cost than other search strategies, such as random sampling.<sup>144</sup> The correlation between advertising and higher sales suggests that consumers do pay attention to advertising.<sup>145</sup> A possible inference from this

139. Klein & Leffler, *supra* note 138, at 94; Cynthia A. Montgomery & Birger Wernerfelt, *Risk Reduction and Umbrella Branding*, 65 J. BUS. 31 (1992) (brand investments signify less quality variation and lower risk); Nelson, *Advertising as Information*, *supra* note 27, at 734, 752; Birger Wernerfelt, *Umbrella Branding as a Signal of New Product Quality: An Example of Signaling by Posting a Bond*, 19 RAND J. ECON. 458 (1988).

140. Klein & Leffler, *supra* note 138, at 627; Wernerfelt, *supra* note 139, at 459.

141. Klein & Leffler, *supra* note 138, at 100.

142. Peter A. Dacin & Daniel C. Smith, *The Effect of Brand Portfolio Characteristics on Consumer Evaluations of Brand Extensions*, 31 J. MARKETING RES. 229, 232 (1994); Montgomery & Wernerfelt, *supra* note 139 (noting that branded goods may not be higher in quality but will have lower quality variation). Furthermore, companies differ in their ability to use resources efficiently to produce quality goods; those that are the most efficient will have the most extra capital to spend on advertising. Nelson, *Advertising as Information*, *supra* note 27, at 732; Nelson, *Economic Consequences*, *supra* note 128, at 214.

143. Nelson, *Economic Consequences*, *supra* note 128, at 215; Nelson, *Advertising as Information*, *supra* note 27, at 734.

144. Nelson, *Economic Consequences*, *supra* note 128, at 215. *But see* HEATH, *supra* note 106, at 78 (finding advertising messages and brand associations are stored in implicit memory regardless of the level of conscious attention paid to the ad); Stewart Shapiro et al., *The Effects of Incidental Ad Exposure on the Formation of Consideration Sets*, 24 J. CONSUMER RES. 94 (1997) (arguing exposure to advertising influences a consumer's willingness to include a product in a consideration set even if the consumer does not consciously attend to the communication).

145. *See, e.g.*, HEATH, *supra* note 106, at 11 (the most successful companies spend the most on advertising); JEAN JACQUES LAMBIN, *ADVERTISING, COMPETITION AND MARKET CONDUCT IN OLIGOPOLY OVER TIME* 94-95 (1976) (reporting the results of an analysis of ten years' worth of sales data for 107 brands in 8 Western European countries).

data is that advertising provides better information at less cost than other available search strategies.

Because the rational maximizer model concentrates on the seller's incentives to maximize quality, it treats the consumer's subjective understanding of why she chooses the advertised brand as irrelevant. Whether or not she is consciously aware of the efficiency of using heavy advertising as a decision heuristic, its use is assumed to lead her to a welfare-maximizing result.<sup>146</sup> Her own self-interest will lead her to disregard advertising for brands that fail to meet her tastes.<sup>147</sup> Although the consumer may misattribute the cause of her preference for the advertised brand to persuasive claims made about the brand, members of the "advertising as information" school would defend her reliance on advertising because it leads to efficient product choices.<sup>148</sup> Signaling through brand atmospherics is thus equivalent to signaling through price premiums, discounts, warranties, or any other kind of strategy through which buyers can gain information about product quality from sellers.

Accordingly, the modern take on the emotional component of brand preference is that these expectations are entitled to insulation from confusion in the same way that utilitarian expectations are. This model posits "emotion" to be another kind of information that consumers can use to weigh the costs and benefits of specific consumptive choices. Belief in the emotional claims made about the brand will increase consumer satisfaction upon purchase, making the consumer subjectively better off.<sup>149</sup> Further, there may be reason to believe that the supposedly irrational emotive preferences linked to heavily-advertised brands may lead to objectively

---

146. Nelson, *Advertising as Information*, *supra* note 27, at 751; Nelson, *Economic Consequences*, *supra* note 128, at 215-17.

147. Nelson, *Advertising as Information*, *supra* note 27, at 734, 751; Nelson, *Economic Consequences*, *supra* note 128, at 215. Note that consumers may not be able to easily disregard advertising even if they would like to. *See, e.g.*, sources cited at note 144 *supra*.

148. *See* Nelson, *Advertising as Information*, *supra* note 27, at 751; *see also* Phillip Nelson, *Information and Consumer Behavior*, 78 J. POL. ECON. 311, 311 (1970) (consumers are only "dimly aware" of the consequences of their product choices); *cf.* Nelson, *Economic Consequences*, *supra* note 128, at 213-14 (arguing information offered by persuasive advertising is not the content of the message but the fact that the producer considers the brand worth advertising); *id.* at 217 (arguing that advertising's primary function for experience goods is to tell consumer which brands advertise the most; the substance of the advertising message is irrelevant).

149. *See supra* text accompanying notes 130-136.

better product choices, albeit not for the reasons that the consumer imagines.<sup>150</sup>

The irrational weigher and the rational maximizer models take a different view of the utility of a trademark's emotional appeal, but otherwise share a similar conception of trademarks as sources of information in decision-making. Both models embrace a Cartesian cost-benefit decision-making process in which consumers examine the attributes of different product choices and select the good that best meets their needs. The models differ solely as to the kinds of information consumers are encouraged to consider in weighing costs and benefits. The irrational weigher model values functional concerns like price and quality over emotional appeal, whereas the rational maximizer model considers emotional appeal to be just as viable a consideration. In either model, whether consumers make practical judgments about features or associative judgments about identities, they are weighing product features and choosing, in a utility-maximizing way, the brand that best meets their needs.

Much of the literature debating the wisdom of dilution protection adopts either one or the other of these two models. Recall that many proponents of dilution theory claim that blurring lessens the ability of famous marks to act as core information "nodes" around which consumers organize a variety of information.<sup>151</sup> If one believes that most of the information that consumers associate with famous marks is emotionally manipulative and functionally irrelevant, then one will not worry about non-confusing uses that might "blur" the connection to this universe of marginalia. However, if one believes that such associations offer useful, easily accessible information, then one will favor protection against opportunistic uses of the mark that could diminish the strength of these connections.

---

150. Nelson's research is characteristic of the neoclassical economic school of thought, which assumed that although people did not actually go through the calculations necessary to discover their subjective expected utility, they used decision-making strategies, such as relying on advertising that would allow them to behave as if they had. Herbert A. Simon, *Rational Decision-Making in Business Organizations*, Nobel Memorial Lecture (Dec. 8, 1978), in *ECONOMIC SCIENCES, 1969-1980: THE SVERIGES RIKSBANK (BANK OF SWEDEN) PRIZE IN ECONOMIC SCIENCES IN MEMORY OF ALFRED NOBEL* 343-348 (Assar Lindbeck ed., 1992).

151. See *supra* text accompanying notes 45-47, 77-78.

### C. Cognitive Misers: Using Emotion to Minimize Cognitive Decision-Making Costs

#### 1. *The Intertwined Roles of Emotion and Reason*

In the last twenty years, a more integrated approach to emotion in consumer decision-making has emerged that undermines both the irrational weigher and rational maximizer views. One weakness of both visions of the consumer is that neither accounts for limitations on consumer time and resources.<sup>152</sup> Each view assumes that consumers use trademarks when trying to solve the problem of optimal product selection. Yet, much research suggests that consumers often use trademarks to solve a different problem: how to use the least amount of cognitive effort in choosing a product. Humans have limited cognitive resources and so allocate them judiciously. In this respect, people have been described as “cognitive misers” who will expend only the effort required to make a satisfactory, rather than optimal, decision.<sup>153</sup> Because emotional responses arise automatically, consumers short on time, motivation, or information often rely on positive or negative “emotional” impulses as the least costly route to making a decision.<sup>154</sup> As discussed more fully below, modern consumer research suggests that such impulses often arise in response to second-order sources of information, such as stimulus familiarity, as a proxy for consideration of the persuasive or functional benefits of a product.

A growing body of literature suggests that emotions do not operate in opposition to reason, but are in fact critical to any form of decision-making.<sup>155</sup> Emotions are part of the nervous system, arising from evolutionarily old parts of the mammalian brain that propel behavior in histori-

---

152. E.g., W. Bentley MacLeod, *Complexity, Bounded Rationality and Heuristic Search*, CONTRIBUTIONS TO ECON. ANALYSIS & POL'Y art. 8, 2-3 (Sept. 30, 2002), <http://www.bepress.com/bejeap/contributions/vol1/iss1/art8>; Herbert A. Simon, *A Behavioral Model of Rational Choice*, 69 Q. J. ECON. 99, 100 (1955).

153. SUSAN T. FISKE & SHELLEY E. TAYLOR, *SOCIAL COGNITION* (1984); Ellen C. Garbarino & Julie A. Edell, *Cognitive Effort, Affect, and Choice*, 24 J. CONSUMER RES. 147, 148 (1997).

154. See *infra* text accompanying notes 171-213.

155. E.g., Richard P. Bagozzi et al., *The Role of Emotions in Marketing*, 27 J. ACAD. MARKETING SCI. 184 (1999); Antoine Bechara & Antonio R. Damasio, *The Somatic Marker Hypothesis: A Neural Theory of Economic Decision*, 52 GAMES & ECON. BEHAV. 336 (2005); Yaniv Hanoch, “Neither an Angel nor an Ant”: *Emotion as an Aid to Bounded Rationality*, 23 J. ECON. PSYCHOL. 1 (2002); Macleod, *supra* note 152, at 42; Paul Slovic et al., *Rational Actors or Rational Fools: Implications of the Affect Heuristic for Behavioral Economics*, 31 J. SOCIO-ECON. 329 (2002) [hereinafter Slovic et al., *Rational Actors*].

cally advantageous ways.<sup>156</sup> Emotions generally identify goals and desires, leading individuals to pursue those desires in conscious and subconscious ways.<sup>157</sup>

Recent neuroscience research suggests that the ability to experience certain outcomes as relatively more positive than others drives decision-making. Antonio Damasio, a cognitive neurologist, pioneered this theory using work with patients with frontal lobe damage. This damage caused them to become “emotionally flat,” while retaining all of their other cognitive abilities. Such patients, although showing no diminishment in memory, attention, or logical reasoning, were profoundly damaged in their decision-making capabilities.<sup>158</sup> For example, one patient could reason endlessly between one of two possible dates for an appointment, but was unable to reach a decision as to either one, or even to decide to stop deliberating.<sup>159</sup> Damasio concluded that the impaired ability to experience emotion was responsible for the man’s inability to arrive at a decision.<sup>160</sup>

Such research suggests that before individuals can effectively deliberate between options, they need a mechanism to focus on the possible consequences of their decisions and to measure the relative desirability of different outcomes. Scientists have posited the existence of “somatic markers” that lead us to classify stimuli as “good” or “bad” as we experience them.<sup>161</sup> We retrieve these feelings again upon encountering or remembering a known object or situation.<sup>162</sup>

Without such somatic tones we might be paralyzed by inaction.<sup>163</sup> The pleasant or unpleasant sensation attached to an image leads the body to react instinctively, much in the same way it reacts unconsciously to hunger, pain, fatigue, or other internal stimuli. Such reactions happen automatically and cause an instant reaction without conscious thought.<sup>164</sup>

---

156. STEVEN PINKER, *HOW THE MIND WORKS* 370-74 (1997); Owen D. Jones & Timothy H. Goldsmith, *Law and Behavioral Biology*, 105 COLUM. L. REV. 405, 438-39 (2005).

157. Jones & Goldsmith, *supra* note 156, at 438-39.

158. ANTONIO R. DAMASIO, *DESCARTES’ ERROR: EMOTION, REASON, AND THE HUMAN BRAIN* (1994).

159. *Id.* at 193-94.

160. *Id.*

161. *Id.* at 173-75; Slovic et al., *Rational Actors*, *supra* note 155, at 332.

162. DAMASIO, *supra* note 158, at 174; Bechara & Damasio, *supra* note 155, at 340; Slovic et al., *Rational Actors*, *supra* note 155, at 332.

163. DAMASIO, *supra* note 158, at 173, 193-94.

164. PINKER, *supra* note 156, at 384-86; Matthias Siemer & Rainer Reisenzein, *The Process of Emotion Inference*, 7 EMOTION 1, 6 (2007) (emotional judgments require less time than cognitive judgments and often precede them); Slovic, *supra* note 2, at 973-974; Pham et al., *supra* note 5, at 175; Zajonc, *Feeling and Thinking*, *supra* note 5, at 157.

Scientists theorize that such mechanisms enabled primitive humans to run when they encountered dangerous situations without first having to pause to plan how to react.<sup>165</sup> In the same way, modern humans rely on somatic tones in weighing abstract outcomes. We are instinctively drawn in or repulsed by the affective markers our experiences have assigned to each outcome.<sup>166</sup> Thus, emotion is what gives us the impetus to make decisions.<sup>167</sup>

Because individuals are not always aware of the degree of attraction or aversion (the “emotional valence”) they may harbor toward a specific object or event, emotion can act as an unconscious shortcut or heuristic. Individuals are drawn to one option, ignoring or dismissing the rest, for reasons they cannot consciously describe.<sup>168</sup> In some cases, the strength of emotional impulse leads people to overreact to the negative, as is the case with the commonly-held preference for driving over flying, though flying is statistically safer.<sup>169</sup> In other cases, such as aesthetic appreciation of art objects, snap affective judgments often lead to more satisfying results than sustained analyses.<sup>170</sup>

## 2. *Emotion as a Heuristic in Consumer Decision-Making*

Emotion is especially important in what social psychologists have termed “low-involvement” processing situations. Recent studies on consumer purchasing behavior suggest that consumers use one of two processes in evaluating advertising and making purchasing decisions.<sup>171</sup>

---

165. PINKER, *supra* note 156, at 386; Slovic, *supra* note 2, at 973.

166. DAMASIO, *supra* note 158, at 174.

167. *See id.* at 190-95 (feelings of good and bad act as “stop,” “go,” and “turn” signals necessary for some aspects of decision-making); *id.* at 198-99 (somatic markers, or good/bad feelings are one of three prerequisites for decision-making, along with attention and working memory).

168. *See id.* at 190-95.

169. Eric A. Posner, *Law and the Emotions*, 89 GEO. L.J. 1977, 1984-85, 2005 (2001).

170. In one study, students were asked to pick a favorite poster from a set of posters. The students who were required to give reasons for their choice ended up less satisfied with the poster than those allowed to choose without explanation. Camerer et al., *supra* note 40, at 29; *see also* Macleod, *supra* note 152, at 41 (surveying the economic literature on bounded rationality and noting studies in which consumers with less information performed objectively better in making decisions than those with more information).

171. Camerer et al., *supra* note 40, at 15-54 (discussing the application of dual-process theories to economic problem-solving generally); Slovic, *supra* note 2, at 972 (2004); Slovic et al., *Rational Actors*, *supra* note 155, at 330-32; Paul Slovic et al., *The Affect Heuristic*, 177 EUR. J. OPERATIONAL RES. 1333, 1334-36 (2007) [hereinafter Slovic et al., *The Affect Heuristic*]; *see generally* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY (Shelly Chaiken & Yaacov Trope eds., 1999); Steven A. Sloman, *The Empiri-*

Where individual motivation and interest is high (for example, the purchase is expensive, risky, or otherwise significant to the consumer), the consumer is likely to use what has been called “high-involvement,” elaborative, or “System II” processing.<sup>172</sup> Consumers using this method behave much the same way that classic trademark theory imagines: they evaluate the attributes of different products and weigh the pros and cons of each choice. In System II processing, individuals will attempt to correct for automatic emotional biases through careful consideration of the implications of different choices.<sup>173</sup> Advertising and branding materials are important sources of information, but purchasers likely evaluate their claims against the purchaser’s own experience, the testimony of other credible sources, and any other available diagnostic information.<sup>174</sup>

The second form of processing, which has been termed “low-involvement” or “System I” processing, is more beneficial for advertisers.<sup>175</sup> This processing strategy is relied on by consumers with few cognitive resources available for purchasing decisions due to distractions, time constraints, information constraints, or lack of motivation.<sup>176</sup> Consumers

---

*cal Case for Two Systems of Reasoning*, 119 PSYCHOL. BULL. 3 (1996); see also Kahan, *supra* note 2, at 107.

172. See Daniel Kahneman, *Maps of Bounded Rationality: Psychology for Behavioral Economics*, 93 AM. ECON. REV. 1449, 1451 (2003).

173. *Id.*

174. See Jong-Won Park & Manoj Hastak, *Memory-Based Product Judgments: Effects of Involvement at Encoding and Retrieval*, 21 J. CONSUMER RES. 534, 544-45 (1994) (hypothesizing that high-involvement consumers make a more extensive search of memory for fact-based information about a product at purchase than low-involvement consumers).

175. Kahneman, *supra* note 172, at 1451.

176. See Norbert Schwarz, *Feelings as Information: Mood Influence Judgments and Processing Strategies*, in *Heuristics & Biases: The Psychology of Intuitive Judgment* 534, 539 (Thomas Gilovich et al. eds., 2002) (summarizing research showing that reliance on feelings as a basis of judgment is more likely when the subject has little other information or when relatively few cognitive resources are available for the decision); Baba Shiv & Alexander Fedorikhin, *Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making*, 26 J. CONSUMER RES. 278 (1999) (demonstrating more reliance on affect-based decision-making under conditions of cognitive strain); Matthias Siemer & Rainer Reisenzein, *Effects of Mood on Evaluative Judgments: Influence of Reduced Processing Capacity and Mood Salience*, 12 COGNITION AND EMOTION 783, 785, 799 (1998) (hypothesizing that use of feelings as a basis for judgment is enhanced in conditions of low motivation and demonstrating that reliance on feelings increases where time constraints and competing task demands limit attentional resources); see also Gita Venkataramani Johar, et al., *MAPping the Frontiers: Theoretical Advances in Consumer Research on Memory, Affect, and Persuasion*, 33 J. CONSUMER RES. 139, 140-41 (2006) (describing strategies that consumers use to process persuasive communications).

using System I processing primarily rely on heuristics, such as emotional impulse, to make purchasing decisions.

In these low-involvement situations, the somatic markers hypothesized by Damasio, the “good/bad” classification that we automatically assign to stimuli, play a dominant role.<sup>177</sup> In a low-involvement strategy, the consumer’s primary goal is to conserve taxed cognitive resources.<sup>178</sup> Because affective classification happens automatically without the consumer having to exert any effort at all, reliance on automatic affective markers will provide the least costly route to a decision.<sup>179</sup>

The more taxed or limited the cognitive resources, the more likely the individual will rely primarily on automatic positive or negative emotional cues to arrive at a choice.<sup>180</sup> Unlike System II decision-making, the consumer will make little attempt to elaborate on or refine the action tendencies suggested by emotion.

To know whether such reliance is utility-maximizing or utility-diminishing, one has to know more about the origin of the somatic marker. Rational maximizers would argue that individuals collapse overall product experience and word of mouth into one global brand attitude reflecting the sum of these parts.<sup>181</sup> In this case, the overall brand attitude is a rough but accurate proxy for the consumer’s overall preference. Irrational weighers would posit that the positive valence of a given brand stimulus may be an objectively poor reflection of rational consumer preferences because con-

---

177. Slovic et al., *The Affect Heuristic*, *supra* note 171, at 1335.

178. Cf. Frederick, *Cognitive Reflection*, *supra* note 176, at 26.

179. Pham et al., *supra* note 5, at 175; Siemer & Reizenzein, *supra* note 164, at 6 (emotional judgments require less time than cognitive judgments and often precede them); Slovic, *supra* note 2, at 973-974; *see generally* Zajonc, *Feeling and Thinking*, *supra* note 5.

180. *See generally* Tim Ambler et al., *supra* note 4; Hanson & Kysar, *supra* note 4, at 732-33 (“[O]ur affective responses to products more often than not determine the purchasing decision, regardless of whether we experience the decision as having resulted from ‘reasons.’”); Shiv & Fedorikhin, *supra* note 176, at 288 (1999); Slovic et al., *The Affect Heuristic*, *supra* note 171, at 1336.

181. Some models of consumer perception and choice embrace this theory. *See, e.g.*, Richard P. Bagozzi et al., *supra* note 155, at 185-88 (describing appraisal theories of emotion in which emotions are thought to result from specific evaluative judgments or cognitions). *But see* R.B. Zajonc, *On the Primacy of Affect*, 39 AM. PSYCHOLOGIST 117, 118-22 (1984) [hereinafter Zajonc, *On the Primacy of Affect*] (detailing evidence against appraisal theories and in favor of independence of affective and cognitive processes); Robert B. Zajonc & Hazel Marcus, *Affective and Cognitive Factors in Preferences*, 9 J. CONSUMER RES. 123, 126-27 (1982) (stating that emotion operates independently of cognition and often precedes cognitive appraisal on which emotions are thought to be based).

sumers mindlessly internalize the pleasant suggestions made in advertising.

The cognitive miser model suggests a third option. According to some studies, a large component of the affect generated from contemplating a known mark stems from the familiarity of the mark itself.<sup>182</sup> Familiarity operates on many levels to increase certainty and reduce effort.<sup>183</sup> To a certain extent, the positive feelings generated by viewing a familiar mark are independent of any larger associative network of stored factual information about the brand.<sup>184</sup> In low involvement situations, consumers do not want to take the time to consult branches of memory for specific brand attributes; they want to know immediately whether they “like” the choice.<sup>185</sup>

Familiarity induces preference for many reasons. First, consumers tend to rely on decision heuristics when operating in conditions of uncertainty.<sup>186</sup> Faced with a nearly endless number of possible choices and decision inputs for many common purchasing decisions, people commonly reduce uncertainty by constructing a decision set of only a few options, and disregard the rest.<sup>187</sup> This set typically includes a subset of those brands about which the consumer is already aware.<sup>188</sup> If the consumer

182. See *infra* text accompanying notes 186-207.

183. *Id.*

184. E.g. Shane Frederick, *Automated Choice Heuristics*, in HEURISTICS & BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 548, 551 (Thomas Gilovich et al. eds., 2002) [hereinafter Frederick, *Automated Choice Heuristic*] (noting the vulnerability of automated choice heuristics, such as affective impressions, to the outsized influence of familiarity); MacLeod, *supra* note 152, at 23 (quality of decision-making using an affect heuristic is affected by the quality of the heuristic but not the underlying beliefs about choices); cf. Punj & Hillyer, *supra* note 4, at 125 (“[C]onsumers rely heavily on [the global attitude toward a brand] in decision making, instead of attempting to recall and process specific brand associations.”); see also *infra* note 200 for sources cited therein.

185. See Zajonc, *Feeling and Thinking*, *supra* note 5, at 151.

186. MacLeod, *supra* note 152, at 40; Simon, *supra* note 152, at 104, 114; Thomas Gilovich & Dale Griffin, *Introduction—Heuristics and Biases: Then and Now*, in HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 1-5 (Thomas Gilovich et al. eds., 2002).

187. See James R. Bettman et al., *Constructive Consumer Choice Processes*, 25 J. CONSUMER RES. 187, 190 (1998); Sarah Coates, Laurie T. Butler & Dianne C. Berry, *Implicit Memory: A Prime Example for Brand Consideration and Choice*, 18 APPLIED COGNITIVE PSYCHOL. 1195, 1196 (2004).

188. Coates et al., *supra* note 187, at 1196; Andreas Strebinger et al., *Brand Equity & Consumer Information Processing: A Proposed Model* 17 (July 1998) (unpublished manuscript, on file with the Vienna University of Economics and Business Administration) (presented at the American Marketing Association’s Marketing Exchange Colloquium in Vienna, 1998). The question of remembering which brands are typical of the kind of

lacks the knowledge or motivation necessary to conclusively distinguish between even these options, she may rely on further heuristics to help make the decision.<sup>189</sup> She may, for example, choose based on price.<sup>190</sup> Matt Groenig's Homer Simpson, for example, advocated a decision strategy of always choosing the second least expensive bottle of wine on a menu. Along with price, the most common deciding factor is brand familiarity.<sup>191</sup> The use of a decision rule of thumb allows consumers to modularize purchasing and off-load much of the cognitive strain of considered decision-making.<sup>192</sup> Familiarity suggests a lack of risk in such circumstances. Therefore, familiarity is a useful proxy when a satisfactory (rather than optimal) decision is all that is required.<sup>193</sup>

Second, familiar brands require less effort to evaluate, and so produce positive feelings all by themselves. Significantly, familiarity enhances positive affect independently of the characteristics of the underlying products, mark owner, or even of the mark itself.<sup>194</sup> Rather, the process of evaluating the mark itself may stimulate significant positive emotions.<sup>195</sup> Experiments have confirmed that repeated exposure to a stimulus alone increases positive feelings toward the stimulus.<sup>196</sup> This is known as the

---

product sought is relevant only in those conditions where consumers are not already directly confronting an array of labeled choices as they would in a supermarket. When brand information is present and obvious, familiarity will still induce liking, however, because of the ease of identifying a familiar brand as a known quantity. *Cf.* ROB ROB WALKER, *BUYING IN: THE SECRET DIALOGUE BETWEEN WHAT WE BUY AND WHO WE ARE* 58 (2008) (consumers must be familiar with a brand to desire it).

189. Frederick, *Automated Choice Heuristic*, *supra* note 184, at 554; Strebinger et al., *supra* note 188, at 20.

190. *See* Strebinger et al., *supra* note 188, at 14, 16 (noting the tendency of consumers to use price as a heuristic).

191. *See* Frederick, *Automated Choice Heuristic*, *supra* note 184, at 554; Strebinger et al., *supra* note 188, at 15-16.

192. Frederick, *Automated Choice Heuristic*, *supra* note 184, at 554.

193. Strebinger et al., *supra* note 188, at 20 (reporting the common heuristic that "[w]ith a well-known brand, nothing much can go wrong").

194. Chris Janiszewski, *Preattentive Mere Exposure Effects*, 20 *J. CONSUMER RES.* 376 (1993) (concluding that mere exposure to a product results in an increased preference for that product); *see generally* Garbarino & Edell, *supra* note 153, at 147, 156 (finding that ease or difficulty of processing affected consumer choice in some contexts independently of choice attributes); Gita Venkataramani Johar, et al *supra* note 176 at 142-43; Zajonc & Marcus, *supra* note 181, at 125.

195. WALKER, *supra* note 188, at 58; Garbarino & Edell, *supra* note 153, at 147.

196. John A. Bargh, *Conditional Automaticity: Varieties of Automatic Influences in Social Perception and Cognition*, in *UNINTENDED THOUGHT* 3 (James S. Uleman & John A. Bargh eds., 1989); Frederick, *Automated Choice Heuristic*, *supra* note 184, at 553-54; Slovic et al., *The Affect Heuristic*, *supra* note 171, at 1336.

“mere exposure” effect.<sup>197</sup> The dominant explanation for the “mere exposure” effect is that the positive reaction to familiar stimuli is a function of ease of recall rather than a conscious appraisal of prior experience with the stimulus.<sup>198</sup> For example, in experiments using nonsense Chinese ideographs, participants experienced an enhanced affect for an ideograph after repeated exposure whether or not they recognized it as something they had seen before and regardless of the absence of any objective meaning for the stimulus.<sup>199</sup> In the context of advertising and brands, this research suggests that consumers will be more positively disposed toward a brand name, logo, or design after repeated exposure to it, regardless of that consumer’s considered evaluation of the underlying product or the actual brand.<sup>200</sup> Studies suggest that consumers are often unaware of the true

---

197. Baker, *supra* note 32, at 32; Robert F. Bornstein, *Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987*, 106 PSYCHOL. BULL. 265, 278 (1989) (“The first twenty years of research on . . . [the] mere exposure effect leaves little doubt that the exposure-affect relationship is a robust, reliable phenomenon.”).

198. Piotr Winkielman & John T. Cacioppo, *Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Elicits Positive Affect*, 81 J. PERSONALITY AND SOC. PSYCHOL. 989, 994 (2001); *see also* Piotr Winkielman et al., *The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment*, in THE PSYCHOLOGY OF EVALUATION: AFFECTIVE PROCESSES IN COGNITION AND EMOTION 189, 197, 203-04 (Jochen Musch & Karl C. Klauer eds., 2003) (feeling of processing fluency gives rise to hedonically positive feelings); *cf.* Robert F. Bornstein & Paul R. D’Agostino, *The Attribution and Discounting of Perceptual Fluency: Preliminary Tests of a Perceptual Fluency/Attributional Model of the Mere Exposure Effect*, 12 SOC. COGNITION 103, 105-07 (1994) (hypothesizing that subjects try to explain the fluency experience by attributing it “to liking (or, for that matter, to any of a variety of stimulus properties that the subject is asked to rate)”).

199. Baker, *supra* note 32; Bornstein, *supra* note 197 at 280-81.

200. Baker, *supra* note 32, at 44 (noting that consumers may chose familiar brands over those with superior attributes); Garbarino & Edell, *supra* note 153, at 156; Norbert Schwarz & Leigh Ann Vaughn, *The Availability Heuristic Revisited: Ease of Recall as Distinct Sources of Information*, in HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 103, 111 (Thomas Gilovich et al. eds., 2002) (arguing that using ease of recall as a judgment heuristic may lead to the opposite conclusion than considered evaluation of recalled content); Piotr Winkielman et al., *supra* note 198, at 207, 211; *cf.* Kenneth Heilman, *Emotional Experience: A Neurological Model*, in COGNITIVE NEUROSCIENCE OF EMOTION 328, 360 (Richard Lane & Lynn Nadel eds., 2000) (Automatic emotional responses are modular and cognitively impenetrable. These responses become the substrate for further cognitive elaboration, which is highly penetrable.); Janiszewski, *supra* note 194, at 3377 (finding that mere exposure to a product results in an increased preference for that product); Prakash Nedungadi, *Recall and Consumer Consideration Sets: Influencing Choice without Altering Brand Evaluations*, 17 J. CONSUMER RES. 263 (1990) (finding that increasing a brand’s accessibility in memory via priming can influence the probability of the brand being chosen even if its evaluations are unchanged); Rob Walker, *The Brand-ness of Strangers*, NY TIMES MAGAZINE, Nov. 16, 2008 at 26

source of the positive affect toward the brand, and misattribute it to the brand or product's likeability, credibility, or suitability.<sup>201</sup>

Third, initial familiarity with a brand influences how the consumer processes subsequent information. To avoid incongruence, people tend to interpret information in line with initial expectations. Once a consumer has formed a positive impression of a familiar brand, she will feel more positively towards subsequent exposures, such as advertisements.<sup>202</sup> The information from advertising also indirectly impacts purchasing decisions.<sup>203</sup> Thus, familiarity with a brand can lead to an endlessly reinforcing positive feedback loop with advertising increasing familiarity and positive associations, and familiarity and positive associations increasing receptiveness to advertising claims.<sup>204</sup> Accordingly, brand familiarity has emerged as one of the most powerful predictors of consumer choice.<sup>205</sup>

This reliance on familiarity can be seen as either irrational or adaptive. Under a theory where consumers are irrationally emotional, advertisers manipulate consumers into feeling more positively toward their brands by barraging them with irrelevant commercial messages designed to condition them to the stimulus. Under an adaptive theory, consumers rational-

---

(describing a study in which subjects who were subconsciously primed with exposure to a brand showed a marked increase in preference for that brand over subjects who were not primed).

201. Camerer et al., *supra* note 40, at 17; Chris Janiszewski & Tom Meyvis, *Effects of Brand Logo Complexity, Repetition, and Spacing on Processing Fluency and Judgment*, 28 J. CONSUMER RES. 18, 18 (2001); Timothy D. Wilson et al., *A Model of Dual Attitudes*, 107 PSYCHOL. REV. 101; Zajonc & Markus, *supra* note 181, at 126-27.

202. E.g., Scott A. Hawkins & Stephen J. Hoch, *Low-Involvement Learning: Memory without Evaluation*, 19 J. CONSUMER RES. 212, 215-16, 223 (1992) (familiarity with a claim increases credibility, especially in low-involvement settings); Catherine W. M. Yeung & Robert S. Wyer, Jr., *Affect, Appraisal, and Consumer Judgment*, 31 J. CONSUMER RES. 412 (2004).

203. Rajeev Batra & Michael Ray, *How Advertising Works at Contact*, in PSYCHOLOGICAL PROCESSES AND ADVERTISING EFFECTS: THEORY, RESEARCH AND APPLICATION 13, 37, 39 (Linda F. Alwitt & Andrew A. Mitchell eds., 1985).

204. *Id.* at 39 (charting the multiple pathways by which advertising exposure, brand familiarity and attitudes toward specific advertising influence each other and impact consumer intentions); see Strebinger et al., *supra* note 188, at 32.

205. Ambler et al., *supra* note 4, at 253 (finding familiarity with a brand to be a good predictor of choice); Emma K. Macdonald & Byron M. Sharp, *Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product: A Replication*, 48 J. BUS. RES. 5, 5-15 (2000) (finding brand awareness to be a powerful influence on brand choice); Batra & Ray, *supra* note 203, at 36 (finding that familiarity is the major determinant of purchase intentions in low-involvement conditions); Strebinger et al., *supra* note 188, at 32 (listing all the pathways through which brand awareness influences choice).

ly prefer more familiar objects because they have relatively more experience with these objects. Assuming nothing bad happened during the individual's past encounters with the brand, the branded object poses less risk than a completely novel option.<sup>206</sup> Furthermore, their very familiarity suggests a sufficient level of investment by an underlying firm to guarantee a minimum quality level. Where a consumer needs only a satisfactory product, and is not inclined to bear the costs of further search, use of familiarity as a proxy may be a utility-maximizing decision strategy.<sup>207</sup>

To the extent that consumers rely on familiarity as a proxy, however, they are not relying on a complex associative memory network of product information to make a choice.<sup>208</sup> Some brain imaging studies done on purchasers while making shopping decisions support this theory.<sup>209</sup> Such studies, although preliminary, reveal divergent levels of brain activity when consumers confront familiar versus unfamiliar brands, but the differences do not involve semantic memory, as classic trademark theory would suggest.<sup>210</sup> One such study, which measured brain activity levels of purchas-

---

206. See, e.g., Bornstein, *supra* note 197, at 282 (hypothesizing that preference for the familiar is an adaptive trait that evolved in humans over many generations; the effects protect an organism from interaction with unfamiliar substances or creatures until there is some evidence that they are not dangerous).

207. Hanoch, *supra* note 155, at sec. 5.1.

208. See, e.g., Punj & Hillyer, *supra* note 4, at 125-26, 130 (finding that strength of preference is largely independent of brand associations and depends on the choice of heuristic used by the consumer); Piotr Winkielman et. al, *supra* note 198, at 204, 211; cf. John G. Lynch, Jr. et al., *Choices From Sets Including Remembered Brands: Use of Recalled Attributes and Prior Overall Evaluations*, 15 J. CONSUMER RES. 169, 171 (1988) (positing that consumers rely to a greater extent on an overall brand evaluation rather than specific recalled attributes in making product decisions). *But see id.* at 177-78 (stating that consumers will rely on recalled attributes if overall evaluations are not diagnostic for the task at hand, such as when the consumer has formed relatively equal overall evaluations of two product choices).

209. Ambler et al., *supra* note 4, at 256; Michael Deppe et al., *Nonlinear Responses Within the Medial Prefrontal Cortex Reveal When Specific Implicit Information Influences Economic Decision Making*, 15 J. NEUROIMAGING 171 (2005).

210. Semantic memory is the area of long-term memory that stores words, meanings and general facts. Alex Martin and Linda L. Chao, *Semantic Memory and the Brain: Structure and Processes*, 11 CURRENT OPINION NEUROBIOLOGY 194, 194 (2001). Semantic memory includes generalized knowledge that does not involve memory of a specific event. See Endel Tulving & Daniel L. Schacter, *Priming and Human Memory Systems*, 247 SCIENCE 301, 301 (1990). It is thought to be distinct from implicit memory, which involves emotional conditioning, unconscious reflexes, and procedural skills. *Id.* at 301, 306; see also Larry R. Squire, *Memory and the Hippocampus: A Synthesis From Findings with Rats, Monkeys, and Humans*, 99 PSYCHOLOGICAL REV. 195, 210 (1992) (noting independence of declarative (semantic) memory from skill- and emotion-based implicit memory).

ers as they made purchase decisions in a supermarket setting, showed little difference in the level of activity in the memory recall centers of the brain for choices involving familiar versus unfamiliar brands.<sup>211</sup> Another study found that choices involving familiar, favorite brands tended to stimulate less activity in working memory and reasoning centers of the brain than choices between unfamiliar brands.<sup>212</sup> These results are inconsistent with what one would expect if consumers were accessing an information-rich store of associations in choosing familiar brands. Instead, such studies reveal that the primary activity difference for familiar or favorite brands relates to brain areas thought to be associated with positive emotions, arousal and focused attention.<sup>213</sup> This research, although very preliminary, sug-

---

211. Ambler et al., *supra* note 4, at 256 (finding no difference in activity of semantic memory based on familiarity of a brand).

212. Deppe et al., *supra* note 209, at 178 (finding reduced activity in neural networks associated with reasoning-based decision for choices involving favored brands). *But see* Samuel M. McClure et al., *Neural Correlates of Behavioral Preference for Culturally Familiar Drinks*, 44 NEURON 379, 385 (2004). McClure found that presence of brand information for Coke in soda taste tests instigated greater activity in semantic memory areas of the brain versus choice between blind options. Interestingly, knowledge of brand information for Pepsi did not produce the same effect. *Id.*

213. *See* Ambler et al., *supra* note 4, at 256-57 (finding higher activity rates for familiar vs. unfamiliar brands in the right parietal lobe—the brain area that is associated with spatial attention and working memory); Deppe et al., *supra* note 209, at 180; *see also* Heilman, *supra* note 200, at 335 (noting that the firing of attentional neurons in the parietal lobe appears to be associated with the significance of the stimulus to the monkey, such that relevant stimulus are associated with higher firing rates than unimportant stimuli).

One famous study that looked only at a comparison of two familiar brands, Coke and Pepsi, did find that differential levels of activity in areas of the brain associated with semantic memory accounted for differences in preferences between two familiar brands. Specifically, the study found that subjects equally preferred Coke and Pepsi in blind taste tests, but more chose Coca-Cola when they had brand information. McClure, *supra* note 212, at 379. Neuroimaging results from that study indicated that access to positive memories about the brand changed the subject's functional preference. *Id.* at 385. A more recent reconstruction of that study using patients with impaired emotional functioning but no memory deficits showed no change in preference when brand information was present. Michael Koenigs & Daniel Tranel, *Prefrontal Cortex Damage Abolishes Brand-Cued Changes in Cola Preference*, 3 SCAN 1, 1-6 (2008). These two studies together show that emotional processing was crucial to the subjects' ability to form a brand-related preference, but suggest that at least in the Coke-Pepsi context, emotional impulse was preceded by a search of brand associations in memory. But these two studies do not indicate that affective reactions always depend on retrieved memories about a stimulus and cannot arise independent of such information. Affective judgments have often been revealed to be faster than factual deliberation, which suggests the two can happen independently. *See, e.g.*, Ambler et al., *supra* note 4, at 254; Pham et al., *supra* note 5, at 175. A better explanation for the difference in results between the McClure and Ambler/Deppe

gests that familiar brands act primarily through an affect heuristic that, in low-involvement conditions, is largely independent of a larger associative information network.

### 3. *Familiarity as Trademark Selling Power*

This research calls into question exactly what dilution law aims to protect by shoring up a mark's "selling power." Trademark owners and many marketing experts have equated this concept with a mark's ability to stimulate certain positive associations in the minds of consumers.<sup>214</sup> The ability of a brand name to stimulate positive brand associations in memory is the leading definition of "brand equity" or overall brand value in marketing literature.<sup>215</sup> But because consumer research suggests that consumers often do not reexamine in detail their brand knowledge each time they encounter a mark, this "brand equity" concept is not a convincing explanation of "selling power." Consumers in low-involvement cases would never reach this level of concern about specific product or producer qualities. If such consumers are concerned primarily with whether they like the mark, and this determination is greatly informed by how familiar the mark is, then "selling power" probably stands for little more than the consumer's general familiarity with a famous mark.<sup>216</sup>

This leads to a second important question: why is anyone is worried about diluting or blurring "selling power" if it is nothing more than the awareness of a mark? Recall that much of the positive reaction to familiar marks stems from their effectiveness at reducing cognitive effort. It is unsurprising that newcomers might want to trade on the positive affect gen-

---

studies would be that when confronted with two brands of roughly equal familiarity, such as Coke and Pepsi, consumers drop back to more detailed examination of relevant affective associations. *E.g.*, Lynch et al., *supra* note 213, at 177-78 (stating that consumers rely on recalled attributes when overall evaluations of two product choices are relatively equal); Baker, *supra* note 32, at 1 (finding that familiarity with a brand did not provide a competitive advantage against equally well-known, established competitors). In such circumstances consumers would be forced to abandon pure System I processing and fall back on at least some System II recall of attribute information. *Cf.* McClure et al., *supra* note 212, at 385 (hypothesizing the independence of systems measuring sensory preference and systems measuring informational preference).

214. *E.g.*, Kevin Lane Keller, *Conceptualizing, Measuring, and Managing Consumer-Based Brand Equity*, 57 J. MARKETING 1, 2 (1993); Punj & Hillyer, *supra* note 4, at 125.

215. *See, e.g.*, David A. Aaker, *Measuring Brand Equity Across Products and Markets*, 38 CAL. MGMT. REV. 102, 102-20 (1996); Keller, *supra* note 214, at 1-22; Punj & Hillyer, *supra* note 4, at 125.

216. *Cf.* Brown, *supra* note 7, at 1194 (noting the advertising maxim that "Repetition is reputation").

erated by famous marks. But presumably, trademark owners should want newcomers to use the brand and so make it even more familiar.

However, the marketing literature does suggest two ways that follow-on uses of a mark could increase the brand's cognitive costs for consumers, and so reduce its effectiveness as a sales tool.<sup>217</sup> These are "inconsistency" and "wearout" effects.

First, studies suggest that inconsistent uses of familiar brands can cause a loss of credibility and diminished confidence in the brand signal. I refer to this as the "inconsistency effect." Familiar trademarks are reassuring because we think we know what they are.<sup>218</sup> If we start to encounter a mark in incongruent settings, even if we don't consciously evaluate the new context or its relationship to the familiar brand, our confidence in our understanding of the brand starts to wane.<sup>219</sup> Inconsistency in this context could refer to an aesthetic conflict, such as if a familiar mark is encountered in an unexpected color, size, or style.<sup>220</sup> Inconsistency can also refer to a conceptual conflict. This would occur when a mark associated with one type of goods is associated with another conflicting type, such as when a brand associated with "mild" products is used on something known to be harsh and abrasive.<sup>221</sup> Interestingly, incongruity may not

---

217. See Loken & John, *supra* note 108, at 71 (noting two ways in which brands can be diluted: wearout and lowering of brand equity); cf. Tülin Erdem & Joffre Swait, *Brand Credibility, Brand Consideration, and Choice*, 31 J. CONSUMER RES. 191, 192 (2004) (noting that high levels of cognitive effort may induce a negative affect and decrease the likelihood of a brand being considered and chosen by a consumer); Garbarino & Edell, *supra* note 153, at 156 (same).

218. Cf. Gregory S. Carpenter, et al., *Market-Driving Strategies: Toward a New Concept of Competitive Advantage*, in *KELLOGG ON MARKETING* (Dawn Iabucci ed., 2000) ("Brands with distinctive personalities are competitively unique, easier to remember and are more positively viewed. Advantage remains so long as the personality remains distinctive.").

219. Dacin & Smith, *supra* note 142, at 232 ("[P]eople have greater confidence in their judgments when they perceive the sample to be homogenous . . . than when they perceive it to be heterogeneous."); Erdem & Swait, *supra* note 27, at 137-38 (1998) (concluding that clarity of brand signals, defined by consistency of the marketing mix, is an important component of the brand's credibility).

220. Incongruity here could be purely a function of decreased "processing fluency" such as when a mark doesn't correspond with its prototype in memory. See Piotr Winkielman et al., *Prototypes Are Attractive Because They Are Easy on the Mind*, 17 PSYCHOL. SCI. 799 (2006) (finding that the prototypicality of a stimulus, its "averageness," is associated with increased processing fluency and positive affect); Winkielman et. al, *supra* note 198, at 193-96 (same).

221. See Loken & John, *supra* note 108, at 73; see also Winkielman et. al, *supra* note 198, at 206 (noting that "conceptual fluency," defined as the semantic congruency between items, also is associated with positive affect).

cause our conscious beliefs about the brand and the original product to change. These beliefs are well-rehearsed and resistant to alteration.<sup>222</sup> What might change is our level of confidence in these beliefs; our willingness to rely on the mark as a decision heuristic therefore would decline.<sup>223</sup> We may find that the brand requires more effort to evaluate and a negative tone associated with risk could appear.<sup>224</sup> At the extreme, this could cause an involuntary switch from an overall positive to an overall negative affect toward the brand.<sup>225</sup>

Psychologists have observed evidence of inconsistency effects when examining reactions to proposed extensions of brands to new product lines. When observers perceive a lack of fit between the physical or conceptual attributes of a second product labeled with a well-known brand, areas of the brain correlated with conflict and negative emotion light up.<sup>226</sup> Similarly, if unrelated third parties repurpose brands on products with a poor “fit” to the consumer’s existing impressions, consumers may lose confidence in their assessment of the brand’s signaling value.<sup>227</sup> At worst, the negative emotions they experience as a result of the poor fit could be translated into a negative reaction to the senior mark.<sup>228</sup> Because

---

222. Deborah Roedder John et al., *The Negative Impact of Extensions: Can Flagship Products Be Diluted?*, 62 J. MARKETING 19, 20 (1998); see also Stephen J. Hoch, *Product Experience Is Seductive*, 29 J. CONSUMER RES. 448, 451 (2002) (stating that existing brand associations block the learning of new attribute associations).

223. Erdem & Swait, *supra* note 27, at 137-38.

224. See Qinggou Ma et al., *Event-Related Potential N270 Correlates of Brand Extension*, 18 NEUROREPORT 1031, 1031-34 (2007) (measuring activity in areas of the brain associated with conflict and negative emotion after viewing inconsistent brand extensions).

225. Dacin & Smith, *supra* note 142, at 233 (“product judgments are negatively related to consumers’ uncertainty in their beliefs about a product.”); Ma et al., *supra* note 224, at 1031-34; Swann, *supra* note 15, at 613-14 (“[U]ltimately, consumers would subconsciously ask ‘why should I try to remember a term that may not lead to a known, but to a variety of experiences.’”). *But see* Kathryn A. Braun-LaTour et al., *Mood, Information Congruency, and Overload*, 60 J. BUS. RES. 1109, 1115 (2007) (finding that consumers may respond positively or negatively to conceptually incongruent messages depending on the mood and type of the processing being relied on).

226. Ma et al., *supra* note 224, at 1031-34.

227. Dacin & Smith, *supra* note 142, at 233; Erdem & Swait, *supra* note 27, at 137-38.

228. Initial studies measuring the impact of the negative assessments of brand extensions found no impact on the overall attitudes to the parent brand. *E.g.* David A. Aaker & Kevin Lane Keller, *Consumer Evaluations of Brand Extensions*, 54 J. MARKETING 27 (1990). However, these studies relied on hypothetical, instead of real, brands and measured attitudes through focused survey questioning. Later studies that focused on real world purchase data found that a negative opinion of an actual extension translated into decreased sales of the parent brand. Erdem, *supra* note 40, at 347 (evaluating scanner

affective reactions arise automatically and faster than cognitive appraisals,<sup>229</sup> such effects occur whether or not the consumer is consciously aware that the new use of a mark is unauthorized.<sup>230</sup>

---

data for toothbrush and toothpaste purchases); Swaminathan et al., *The Impact of Brand Extension Introduction*, *supra* note 40, at 12-14 (evaluating household purchase data). Other studies found evidence of negative spillover to brand attribute beliefs from extensions to incongruent products. *E.g.*, Loken & John, *supra* note 108, at 71-84. Another study found that after seeing an incongruent brand use, participants viewed subsequent brand extensions less favorably. Kevin Lane Keller & David A. Aaker, *The Effects of Sequential Introduction of Brand Extensions*, 29 J. MARKETING RES. 35 (1992); *see also* Swaminathan, *Sequential Brand Extensions*, *supra* note 40, at 440-41 (finding that negative evaluations of brand extensions lowered the likelihood of trying subsequent extensions but did not effect the purchases of the parent brand).

229. *See* Piotr Winkielman et al., *supra* note 198, at 195, 209; sources cited *supra* note 179.

230. Some survey-based studies suggest that use of a sub-brand or other device to signal a distance between the parent and the extension insulates the parent brand from negative spillover effects. *See, e.g.*, Amna Kirmani et al., *The Ownership Effect in Consumer Responses to Brand Line Stretches*, 63 J. MARKETING 88, 89 (1999) (finding that signaling a difference between the main brand and the extension prevents dilutive effects on the main brand); Sandra J. Milberg et al., *Managing Negative Feedback Effects Associated with Brand Extensions: The Impact of Alternative Branding Strategies*, 6 J. CONSUMER PSYCHOL. 119, 136-37 (1997) (finding that sub-branding may prevent negatively evaluated brand extensions from harming the parent brand); Tushnet, *supra* note 40, at 543, 543 n.17. This might suggest that negative spillovers are unlikely to occur in the case of trademark blurring because the consumer will differentiate between authorized and unauthorized uses. Tushnet, *supra* note 40, at 543. However, one has to be cautious both ways with behavioral research. These studies only examined changes in the subject's conscious beliefs about a brand after a limited exposure to a potentially dilutive alliance. Real spillover effects would take time to develop. *See* Swaminathan et al., *The Impact of Brand Extension Introduction* *supra* note 40, at 2 (noting the inability of survey studies to measure spillover effects over time). Furthermore, survey-based measures of conscious brand beliefs do not effectively capture the impacts on automatic affective responses. Because some studies reveal that behavior or affect can change without a corresponding change in conscious beliefs, it may be that the two evaluation systems can work independently, at least under some conditions. *See, e.g.*, Zajonc, *On the Primacy of Affect*, *supra* note 181, at 117-23 (1984) (detailing studies suggesting that affect precedes conscious cognition and appraisal, and that changes in affect and cognition are not always correlated). Finally, some trademark "blurring" may not provide the quick and easy debiasing information necessary to avoid dilutive effects. Trademark dilution is usually considered non-confusing because of the larger context in which it occurs. The oft-cited examples of Kodak pianos and Buick shoes are supposedly not confusing because the product line is so dissimilar we wouldn't automatically assume a connection with the famous brand, not because the new marks offer immediately apparent signals of difference from the famous ones. Because changes in affect are linked to demands on cognitive resources, the harder consumers have to work to distinguish uses, the more likely affect changes are to appear. Thus, such studies support denying liability where the allegedly dilutive mark is obviously different from the famous mark, the balance I suggest in Part

Second, the marketing literature suggests that although familiarity leads to a positive affect, certain kinds of over-familiarity lead to boredom and a negative affect. Although repeated exposure to a stimulus initially leads to liking, unvaried exposure eventually causes boredom.<sup>231</sup> If we see the same things too often, we resent having to use effort to evaluate them repeatedly.<sup>232</sup> Brand owners manage this advertising wearout by varying advertising campaigns and redesigning logos from time to time.<sup>233</sup> The research is mixed as to whether wearout even exists, and if so, to what extent it impacts familiar brands.<sup>234</sup> Nevertheless, real-world brand managers express concern about it.<sup>235</sup> A few studies support the notion that blurring could lead to wearout because market newcomers are more likely to choose promotional strategies that are known to produce wearout effects.<sup>236</sup> Established brands tend to rely on evocative advertising strategies meant not to challenge the consumer, but merely to remind him in a background way about the mark.<sup>237</sup> Advertising for new products typically tries to grab the consumer's attention and challenge them to change their

---

IV.D. They do not really speak to the harder case of identical marks used in contextually different ways.

231. See, e.g., Bobby J. Calder & Brian Sternthal, *Television Commercial Wearout: An Information Processing View*, 17 J. MARKETING RES. 173, 185 (1980) (finding that evaluations of television commercials and advertised products became more negative after multiple repetitions); cf. Walker, *supra* note 200, at 26 (describing a study that found that some exposure to a brand increased preference for it, but that too much exposure caused fewer subjects to choose that brand).

232. Donald E. Berlyne, *Novelty, Complexity and Hedonic Value*, 8 PERCEPTION PSYCHOPHYSICS 279 (1970) (proposing that affective consequences of exposure are a function of learning and satiation, with learning leading to positive affect and satiation creating boredom that leads to negative affect); Calder & Sternthal, *supra* note 231, at 185; Margaret C. Campbell & Kevin Lane Keller, *Brand Familiarity and Advertising Repetition Effects*, 30 J. CONSUMER RES. 292 (2003).

233. See Christie L. Nordheim, *The Influence of Level of Processing on Advertising Repetition Effects*, 29 J. CONSUMER RES. 371, 371 (2002).

234. Douglas R. Scott & Debbie Solomon, *What Is Wearout Anyway?*, 38 J. ADVERTISING RES. 19 (1998); cf. Swaminathan, *Sequential Brand Extensions* *supra* note 40, at 441 (hypothesizing that multiple product extensions increase consumer interest and knowledge across different segments of society).

235. E.g., Bill Britt, *Disney's Global Goals*, MARKETING, May 17, 1990, at 26; Richard Gibson, *The End of the Line? Overkill on Extensions*, WALL ST. J., June 18, 1990, at B1.

236. See, e.g., Nordheim, *supra* note 233, at 372-82 (measuring the effect of levels of processing on wearout of affective response to advertising); cf. Campbell, *supra* note 232 (finding that ads for familiar brands wearout more slowly than the same ads attributed to unfamiliar brands).

237. Scott & Solomon, *supra* note 234, at 22-23.

behavior.<sup>238</sup> This second kind of ad stimulates more elaborative processing and is more likely to lead to wearout.<sup>239</sup> These studies suggest that advertising strategies adopted by newcomers could hasten wearout for familiar brands if they cause audience members to think more deeply about the brand.<sup>240</sup>

It is difficult to measure the strength of inconsistency and wearout effects. Much of the existing research in this area comes from laboratory studies that imperfectly reflect the crowded conditions of the marketplace. It is not clear that student reactions to carefully studied blank “tombstone” ads in an experimental setting bears any relationship to consumer reactions to incidental encounters with the hundreds of ads and trademarks that consumers typically see in a day.<sup>241</sup> Furthermore, some studies measure changes in people’s affect or beliefs through direct focused questioning designed to stimulate thoughtful responses.<sup>242</sup> Few of us pause long enough to consider our reactions to brands and logos in this way, and engaged consideration of our reactions may inadvertently change them.<sup>243</sup>

Furthermore, famous brands may be impervious to the effects of inconsistency and wearout. One study designed to measure the delay in memory recall of brands after exposure to diluting uses found no significant effects for familiar marks such as Continental Airlines and Avon.<sup>244</sup> It may be that our positive impressions of these kinds of brands are so well-established that it would take negative reinforcement of cataclysmic

---

238. *Id.*

239. *See* Nordheim, *supra* note 233, at 373, 380 (demonstrating that wearout is more likely where consumers think more extensively about the content or presentation of advertising).

240. *Id.* The research on wearout is far from settled. Some have suggested that multiple uses of similar brand names will enhance consumer interest and learning about a brand, thus enhancing affect toward the famous mark. *See* Tushnet, *supra* note 40, at 536-39. Because so many free-riders attempt witty word play with famous marks, consumer appreciation of and memory for the original mark might be enhanced by these new opportunities to learn. On the other hand, the follow-on activity might enhance interest in the new brands while contributing to a sense of tedium about the original. *Cf.* Calder & Sternthal, *supra* note 231, at 185 (finding that variations of ad executions forestalled wearout for the ads but not the advertised product).

241. Tushnet, *supra* note 40, at 542.

242. *See, e.g.,* Aaker & Keller, *supra* note 228; Loken & John, *supra* note 108 at 71-84.

243. Johar et al., *supra* note 194; Demetrios Vakratsas & Tim Ambler, *How Advertising Works: What Do We Really Know?*, 63 J. MAREKTING 26, 36 (1999).

244. Morrin & Jacoby, *supra* note 40, at 14.

proportions to unseat them.<sup>245</sup> Still, history is not without examples of once-famous brands that lost their footing and their customer base through poor image management. Gucci, for example, came close to bankruptcy in 1994 after the brand was licensed on everything from toilet paper to tote bags.<sup>246</sup> The tipping point may be high for well-known brands, but there is no reason to believe it doesn't exist.<sup>247</sup> Further, even if flagship products were completely impervious to dilution, blurring still might negatively impact the ability of the famous mark owner to launch new products and extensions because of the effort required to distinguish authorized from unauthorized new uses.<sup>248</sup>

Current behavioral research cannot establish that cognitive or affective changes are certain or even likely to result from conduct defined as "blurring" under current law. However, it can provide a theoretical model that better explains the concerns expressed by trademark owners. Under the cognitive miser theory of consumer decision-making, the harm of dilution is a feeling of dislike or anger that may result from the extra cognitive effort required to evaluate a brand. Because emotional valence is so important to purchasing decisions in low-involvement situations, that negative response might cause an involuntary and, ultimately inefficient change in purchasing behavior.<sup>249</sup>

---

245. John et al., *supra* note 222, at 19, 20 (stating that consumer's network of beliefs linked to flagship products tends to extreme, strongly held, and resistant to change). *But see* Swaminathan et al., *The Impact of Brand Extension Introduction supra* note 40, at 2 ("The beliefs associated with the core parent brand are likely to be of varying strength across different segments of consumers.").

246. *See, e.g.*, John Carreyrou & Cecilie Rohwedder, *Style & Substance: In Again at Gucci: Licensing—PPR Says It May Start to Sell Franchises, Product Rights For Smaller Designer Brands*, WALL ST. J., Mar. 5, 2004, at A9; *see also* Suzy Menkes, *A License to Kill; Fashion Houses Move to Tighten Brand Control*, INT'L HERALD TRIB., July 4, 2000, at 9 (discussing a trend in fashion to not license famous brands to outside manufacturers because of risks of declining image, poor quality control, and poor valuation by potential investors); Gibson, *supra* note 235, at B1.

247. *E.g.*, Camerer et al., *supra* note 40, at 25 (stating that people tend to disregard information that conflicts with well-defined beliefs until an accumulation of evidence leads to an abrupt recategorization).

248. *See* Swaminathan, *Sequential Brand Extensions supra* note 40 at 440-41; Aker & Keller, *supra* note 228.

249. A final reason for trademark owners to be concerned by uncontrolled use by third parties, whatever its cognitive effects for consumers, is that the law assumes that such use harms the distinctiveness of the senior mark. The presence of many variations of a famous mark in the marketplace will decrease the first mark's effective scope of protection. For example, trademark owners who fail to object to similar uses of their mark on unrelated goods risk losing the ability to object to any new use because adjudicators may assume that any additional use is unlikely to create confusion in what is already a

## IV. IMPLICATIONS

### A. Blurring is a Form of Tarnishment

One benefit of this research is that it offers a theoretical model for the harm that dilution causes that is intuitively easier to understand. Dilution harms trademark owners not because it causes consumers to “think for a minute,” but because that pause may cause the consumer to become frustrated with the mark and punish the brand.

The introduction of emotion to the equation answers many of the critic’s concerns about the elusive harm underlying the blurring doctrine. In particular, the emotional explanation of dilution answers the question of why a consumers’ behavior toward a senior mark might change even if she is not consciously confused about whether the mark’s owner has authorized a subsequent “blurring” use. Affective reactions to marks are automatic and can precede cognitive appraisal.<sup>250</sup> Therefore, assuming there is sufficient similarity between two marks, a consumer is likely to imbue a second product with the halo of a famous mark’s reputation even if further cognitive elaboration reveals the mistake.<sup>251</sup> She may react less positively toward the senior mark after a sufficient exposure to blurring not because she believes the blurring uses were authorized, but because she now reflexively finds the senior mark less credible and therefore more taxing.<sup>252</sup> Essentially, the term dilution refers to an alternative form of “confusion” that is pre-conscious and automatic. Those who blur a famous mark may receive the benefits of its automatic positive response even if they are not entitled to it, and too many such users may cause a change in a consumer’s

---

“crowded field.” *Cf.* *Nautilus Group, Inc. v. Savvier, Inc.*, 427 F. Supp. 2d 990, 995 (W.D. Wash. 2006) (arguing that protection of even strong marks is weakened by presence of a “crowded field”); *Moose Creek, Inc. v. Abercrombie & Fitch Co.*, 331 F. Supp. 2d 1214, 1225 (C.D. Cal. 2004) (same). Similarly, the new Trademark Dilution Revision Act of 2006 instructs courts to consider the extent to which a senior mark owner is making “exclusive use” of the mark as a factor in gauging the likelihood of blurring. Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, § 2, 120 Stat. 1730, 1730 (to be codified at 15 U.S.C. § 1125(c)). The more uses of similar marks a defendant can point to, even on unrelated goods, the less chance any new use will be found to “blur” the distinctiveness of the senior mark beyond what has already occurred. In this respect it pays for trademark owners to object to even trivial third party use to prevent the weight of those sorts of uses from undercutting the law’s recognition of the mark’s distinctiveness: a legal “death by a thousand cuts.” Thus whatever the state of the research in support of the harm of dilution by blurring, trademark law already, perhaps inadvertently, internalizes some of its core assumptions.

250. *See supra* note 179 for sources cited therein.

251. *See supra* text accompanying notes 82-83.

252. *See supra* text accompanying notes 82-83.

automatic response to the mark even without conscious confusion about the source of the mark.

This model reverses the common understanding of blurring and tarnishment. Tarnishment, defined as uses of a mark on unrelated goods that cause harm to the senior user's reputation, is sometimes referred to as a "subset" of blurring.<sup>253</sup> The common assumption is that all prominent uses of similar marks on unrelated goods cause blurring by disassociating the mark from the senior owner. However, only some uses with particular kinds of products are thought to negatively influence the consumer's opinion of the original mark. In fact, it's probably more accurate to say that blurring is a subset of the ways in which a mark may be tarnished. Through wearout or inconsistency, blurring may slowly cause the same kind of negative change in affective valence that occurs immediately with tarnishment. Both tarnishment and blurring eventually cause the consumer to feel worse about the senior brand.

### **B. Information Quality and Property Rights in Mental Processes**

That dilution has the potential to be harmful does not answer the question of whether we should forbid it. This section examines the likely harms flowing from dilution by blurring and considers the extent to which these harms implicate any larger societal interests, as opposed to private harms to individual producers. One interesting aspect of dilution theory is that the harm at issue, an unconscious and involuntary response to advertising that frustrates an individual's more considered preferences, is similar to the original critiques of persuasive advertising. But proponents of dilution law cannot have it both ways. If "involuntary" alteration of consumer preferences is wrong, then the positive feelings generated by persuasive advertising are at least as problematic as any negative ones stimulated through dilutive conduct. If persuasive advertising can be valuable even if consumers don't understand how they use it, then perhaps the same might be said for dilution by blurring. Because the law has no inherent interest in product preference per se, dilution regulations must stand or fall on whether they implicate societal interests in promoting competition by allowing effective communication about products and services.<sup>254</sup>

Recall that our primary goals with regard to trademark protection are to lower search costs and to increase the overall quality of information ex-

---

253. Long, *supra* note 21, at 1059; Swann, *supra* note 15, at 622.

254. See Kratzke, *supra* note 16, at 214-29 (arguing that trademark law's paramount concern should be the interests of consumers as served by competition among brands, and that trademarks serve this interest by efficiently providing information about available choices).

changed in the marketplace in the service of inter-brand competition.<sup>255</sup> Protection against dilution by blurring is relatively easy to square with trademark law's first goal. Protecting the automatic allure of familiar marks certainly helps to lower search costs. Reliance on preconscious impulse serves a consumer's interest in acquiring goods with the lowest possible expenditure of cognitive effort.

The second goal—to increase the quality of the information exchange to stimulate competition—poses a harder case. The argument in favor of excluding arguably non-confusing blurring such as the coffee shop Federal Espresso or the medical equipment importer Nikepal from the marketplace relies on one of two assumptions about consumer behavior.<sup>256</sup> Consumers will either (1) inadvertently favor lower quality products because of their automatic positive response to a well-known mark, or they may (2) avoid high-quality products sold under the original mark because they perceive increased risks to the brand that are not real. These arguments do not necessarily hold up in light of the current understanding of how automatic emotional processes guide purchasing decisions.

Let's consider the low-quality, dilutive product first. If people automatically judge products associated with well-known marks as more credible and less risky, then they may embrace a good that uses a famous mark but has no relationship to it. Thus, they will buy more risk than they intended. A consumer, for example, might be marginally more inclined to purchase "Mercedes Climbing Ropes" because the familiar mark subconsciously will suggest a lack of risk. Because the rope producer lacks the car company's incentives to ensure quality, it might operate cheaply, using low quality materials that will fail under stress. In this case, the consumer could suffer a grave harm. The law forbids the use of confusing trademarks to prevent just this kind of mistaken purchase. However, if the consumer can easily ascertain from context that the rope comes from a source unrelated to the high-quality car, she will most likely correct for the attribution bias herself because she has an interest in not being misled.<sup>257</sup> It is unlikely that consumers buy important goods solely because of a brand signal they consciously know to be false. If the use of the signal alerts the

---

255. See *supra* text accompanying note 15.

256. *Fed. Express Corp. v. Fed. Espresso, Inc.*, 201 F.3d 168 (2d Cir. 2000); *Nike Inc. v. Nikepal Int'l Inc.*, 84 U.S.P.Q.2d (BNA) 1820 (E.D. Cal. 2007).

257. Cf. *Park & Hastak*, *supra* note 174, at 544 (arguing that consumers will rely less on automatic evaluations where they are highly motivated, such as when a purchase is important to them); Posner, *supra* note 169, at 1985 (theorizing that people are generally knowledgeable about their emotional dispositions and can recognize when emotions are likely to lead them astray as well as take steps to modify their reliance on emotions).

consumer to an option that, on further reflection, she would like to purchase even though it is unrelated to the well-known good, it is hard to see much immediate harm to the consumer or the larger market.

The longer-term concern is that the inconsistent use of the senior brand symbol will lower the credibility of the senior brand.<sup>258</sup> The brand's ability to generate an automatic positive response may be compromised. In this way, free-riders might cause people to involuntarily change their preferences for the senior brand, thus lowering the producer's incentives to invest in quality brands generally. This will also prevent communication of "real" preferences to producers, interfering with the efficient allocation of resources to production of the goods that are most desirable to consumers.<sup>259</sup>

This argument engages in an interesting paternalism about preferences. It's not quite the approach imagined by Ralph Brown: that only deliberative, conscious consideration of product features is rational and worth protection. It's also not quite the approach envisioned by the rational maximizers: that the law should protect the consumer's ability to focus without distraction on the positive attributes of the mark and product. Consumers in many circumstances do not bother to retrieve brand associations with this level of detail and semantic meaning. Instead they use the familiarity and consistency of a stimulus as a rough proxy for a more detailed semantic evaluation. The preference expressed by the consumer in these cases is for an easy, safe choice without much effort. It is worth asking whether protecting convenience, divorced from more specific informational content, draws the correct balance between lowering search costs and promoting beneficial competition.

1. *Lowering Brand Credibility does not necessarily frustrate Consumer Preferences*

First, it is not clear that the lowering of one brand's credibility through blurring will actually frustrate consumer preferences. Recall that consumers have a choice of strategies when making a purchasing decision.<sup>260</sup> This choice may dictate how much the "emotional" valence of a given trademark will influence their behavior. As previously discussed, consumers have a choice of relying on deliberative, high-cost System II decision-making, and automatic, low-cost System I decision-making. It follows that consumers will rely on deliberative methods when they deem the choices of sufficient importance and they have sufficient information

---

258. See *supra* text accompanying notes 217-230.

259. E.g., Sheff, *supra* note 102, at 358-61.

260. See *supra* text accompanying note 150.

to make an informed choice.<sup>261</sup> Consumers themselves are probably at least partially aware of their susceptibility to familiarity heuristics, and will take steps to counter these influences when they judge it important to do so.<sup>262</sup>

If the consumer would prefer to rely on System I decision-making, she has implicitly determined that she does not mind being slightly misled by automatic judgment heuristics. Consumers rely on System I processes when they prefer to conserve cognitive resources.<sup>263</sup> They are willing to accept the risk of a less than optimal choice in exchange for lower decision making costs.<sup>264</sup> In this sense, consumers can be seen as complicit in allowing automatic affective responses to guide decision-making. Although their expressed preference might differ from the choice they would have made if they had unlimited time and resources to choose, it accurately reflects their combined preference for the product and the amount of time spent searching for it.<sup>265</sup> Thus, blurring is likely to have the greatest impact on product choice in situations where consumers lack clearly defined functional preferences. If one choice has been “blurred,” the consumer can substitute another familiar brand that also meets her taste for risk-aversion and cognitive cost-avoidance. Such substitution harms the senior producer, undoubtedly. Social welfare, on the other hand, has declined little if the consumer finds a substitute that adequately meets her needs.

---

261. Indeed, research suggests that consumers rely to a greater extent on licensing and certification information than on seller experience claims in the case of high-cost, infrequently purchased credence goods and services. See Robert B. Ekelund, Jr. et al., *Advertising and Information: An Empirical Study of Search, Experience and Credence Goods*, 22 J. ECON. STUD. 33, 41 (1995).

262. Posner, *supra* note 169, at 1985 (2001) (theorizing that people are generally knowledgeable about their emotional dispositions and can recognize when emotions are likely to lead them astray as well as take steps to modify their reliance on emotions).

263. Frederick, *Cognitive Reflection*, *supra* note 176, at 26.

264. See *supra* text accompanying notes 170-188.

265. Cf. Frederick, *Automated Choice Heuristic*, *supra* note 184, at 557-58 (noting theories of adaptive decision-making that trade off effort and accuracy by tailoring use of heuristics to importance of the decision); Gerd Gigerenzer et al., *How Good are Fast and Frugal Heuristics?*, in *HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT* 559, 580 (Thomas Gilovich et al., eds., 2002) (noting that consumers probably trade off simplifying versus complex decision strategies against the time and effort they are willing to spend).

2. *Property Rights in Positive Feelings is Beyond the Scope of IP Law*

Second, granting proprietary rights in reflexive positive feelings also seems to be a departure from the conventions of intellectual property law. Even in areas of law designed to reward producers, such as patent and copyright, we refuse to grant rights in pure “mental processes” divorced from any specific tangible result.<sup>266</sup> We do this out of a concern for competition; it is not desirable for any one entity to own the ability of all people to use their wits in a certain way.<sup>267</sup> Awarding property rights for the ability to signal general familiarity and consistency may raise similar monopoly concerns. No competitor can ever compete effectively with a category leader on this basis. The established brand will always be more familiar.<sup>268</sup> Constraints on consumer’s time and motivation to search will always favor familiar brands because they will always seem less risky.<sup>269</sup>

Market leaders benefit from such effects even without dilution law. For example, one study comparing twenty-two leading brands in 1925 found that in all but three of the product classes, the same brand was still the leader in 1985.<sup>270</sup> Consumers may have good reasons for continuing to

266. Copyright law explicitly excludes ideas and processes. 17 U.S.C. § 102(b) (2000). Patent law also excludes protection for “mental processes” that are divorced from any technological embodiment or tangible effect. *Diamond v. Diehr*, 450 U.S. 175, 185 (1981); *In re Comiskey*, No. 2006-1268, 2009 WL 162408 (Fed. Cir. Jan. 26, 2009).

267. *Comiskey*, 2009 WL 162408 at \*6 (noting that patent’s subject matter restrictions are designed to weed out monopolies that will not enrich the public); *id.* at \*7 (stating that mental processes provide the basic blocks of scientific inquiry and therefore should not be the subject of exclusive rights) (quoting *Gottschalk v. Benson*, 407 U.S. 63, 67 (1972)).

268. *Baker*, *supra* note 32, at 44 (finding that brand familiarity superiority inoculates established brands from competitor attempts to use familiarity to influence choice; the dominant brand is always more familiar).

269. Anusree Mitra & John G. Lynch, Jr., *Toward a Reconciliation of Market Power and Information Theories of Advertising Effects on Price Elasticity*, 21 J. CONSUMER RES. 644, 645 (1995) (stating that consumers tend to simplify their purchasing decisions by making their selection from smaller subsets of all possible available brands); Erdem & Swait, *supra* note 27, at 140 n.2 (arguing that uncertainty and information costs influence which brands are included in consideration sets); Lynch et al., *supra* note 213, at 171 (theorizing that consumers consult the most accessible inputs first in making decisions; if the first input is diagnostic, the search terminates); *cf.* John J. Siegfried & Laurie Beth Evans, *Empirical Studies of Entry and Exit: A Survey of the Evidence*, 9 REV. INDUS. ORG. 121, 139 (1994) (hypothesizing that consumer’s risk-aversion favors established players).

270. Steve Hartman, *Brand Equity Impairment—The Meaning of Dilution*, 87 TRADEMARK REP. 418, 429-30 (1997). Federal dilution protection was not passed until 1998, although some states offered dilution protection as early as 1926.

patronize familiar sellers, but the law insulates such tendencies by proper-tizing brand familiarity. Protection against dilution may be tantamount to giving property rights in market leadership.

3. *Enforcing Different Advertising Standards for Established Producers and Market Newcomers Promotes Neither Accuracy nor Competition*

Dilution regulation in practice treats advertising designed to appeal to affective biases differently depending on whether such communications come from an established producer or a market newcomer. Although other areas of the law offer precedent for excluding affectively biasing information in the interest of accurate decision-making, such rules are usually applied consistently to every party engaging in persuasion. For example, the Federal Rules of Evidence require judges to exclude otherwise relevant evidence from jurors if the probative value of that information is substantially outweighed by the danger of unfair prejudice.<sup>271</sup> “Unfair prejudice” means an undue tendency to influence a decision on an improper basis; this is commonly, though not necessarily, an emotional one.<sup>272</sup> Rule 403 assumes that certain kinds of information and suggestion may so captivate the listener’s attention and imagination that she is not capable of disregarding it.<sup>273</sup> In such situations, considerations of economy and fairness mitigate in favor of excluding the information at the outset.

In the trademark context, owners essentially argue that commercial free-riding on famous trademarks offers little relevant information and has a tendency to elicit emotional responses that cause objectively poor decisions (essentially arguing that the informational value provided by free-riding is substantially outweighed by its prejudicial value). This analogy can explain why dilution law does not target parody, news reporting and general idiosyncratic personal experience; not because such uses do not blur the senior mark, but that their relevance to the senior producer outweighs any likely harm. We believe that consumers benefit by considering

---

271. Fed. R. Evid. 403.

272. *United States v. Puckett*, 405 F.3d 589, 598 (7th Cir. 2005). Typically the kinds of evidence excluded under this rule fall into one of three categories: (1) evidence which seeks to affect the jurors’ perception of a party, either favorably or unfavorably, because of evidence of past crimes or bad acts, bad habits, or a party’s past good acts; (2) evidence damaging the position of a party because of the party’s association with certain groups, as by showing that a party is insured or associated with an unpopular political group; and (3) evidence which will incite the jury’s rage or desire for revenge against a defendant, the most successful method being the introduction of an inflammatory picture. *Id.*

273. *Id.*

relevant criticism and commentary in making purchasing decisions. By contrast, dilution law proponents argue that free-riding is entirely opportunistic,<sup>274</sup> and in such cases the balance tips the other way.

Courts have found these kinds of concerns most persuasive in the tarnishment context. Judges are willing to forbid uses of marks that are likely to incite disgust or fear out of concern that such emotions, once stimulated, will be difficult to put aside even when the consumer knows the use is unauthorized.<sup>275</sup> These cases typically concern the commercial connection of a famous mark to products connected with sex, drugs, or violence.<sup>276</sup> If blurring is a form of tarnishment, as has been argued above, perhaps it should comply with a similar balancing test.

It would be troubling, however, to apply a rule requiring relevance selectively only to some commercial speakers. At trial, the rules of evidence apply equally to each adversarial party. Neither side is permitted to introduce evidence with low probativeness that has a tendency to bias the preferences of the factfinder. In trademark law, we allow and even encourage trademark owners regularly to provide information of dubious relevance and a large propensity to appeal to affective biases. We ask trademark consumers to do the work of parsing relevant from irrelevant information in this context. If preserving the quality of information availa-

---

274. See, e.g., *Ty Inc. v. Perryman*, 306 F.3d 509, 512 (7th Cir. 2002) (Posner, J.) (noting an example of a “Tiffany’s” restaurant in Kuala Lumpur as an example of trademark free-riding that might dilute; the restaurant owner selects the name solely to benefit from the fame of the senior Tiffany’s mark); David J. Franklyn, *Debunking the Dilution Doctrine: Toward a Coherent Theory of the Anti-Free-Rider Principle in American Trademark Law*, 56 HASTINGS L.J. 117, 141 (2004) (arguing that trademark free-riders select their marks to capitalize on fame they did not create, that they select their marks solely due to similarity with well-known marks, and that allowing such use grants no discernable benefit to society); cf. *Nike Inc. v. Nikepal Int’l Inc.*, 84 U.S.P.Q.2d (BNA) 1820, 1823 (E.D. Cal. 2007) (noting that defendant argued that he had picked the name “Nikepal” from randomly pointing in the dictionary).

275. See *Ty Inc.*, 306 F.3d at 511 (arguing that because of the inveterate tendency of the human mind to proceed by association, people will connect the jeweler Tiffany’s to the strip joint Tiffany’s even if they know the two are unrelated); *Cynthia Grey v. Campbell Soup Co.*, 650 F.Supp. 1166, 1175 (C.D. Cal. 1986) (“Grey’s use of DOGIVA and CATIVA . . . injures Campbell’s business reputation because of the association which the public makes between DOGIVA and CATIVA treats for animals and GODIVA premium quality food products which are intended for human consumption.”); *Dallas Cowboys Cheerleaders, Inc. v. Pussycat Cinema, Ltd.*, 467 F. Supp. 366 (S.D.N.Y. 1979) (enjoining a pornographic film that prominently featured plaintiff’s trademarks because of tarnishment of reputation of a famous cheerleading squad); *Coca-Cola Co. v. Gemini Rising, Inc.* 346 F. Supp. 1183 (E.D.N.Y. 1972) (enjoining a poster that connected the Coke beverage to the slogan “Enjoy Cocaine”).

276. Long, *supra* note 21, at 1057.

ble to the consumer is the goal of trademark dilution law, then it is problematic to have two different standards for the trademark owner and for follow-on speakers. If we acknowledge that much of the brand information that owners seek to protect can also be called “more prejudicial than probative,” it’s unclear that we should protect the integrity of this information against free-riding. Holding owners of established marks to materially different informational standards than market newcomers seems unlikely to increase competition or the quality of information in the marketplace.

### C. Free-Riding as Information

An alternative approach, more in line with the “advertising as information” school, is to acknowledge the value that simple, familiar brands offer to consumers in a crowded marketplace, and to recognize similar, competing value offered by others who free-ride on those symbols for their own purposes. Most critiques of dilution law focus on the law’s propensity to be used against traditionally high-value First Amendment speakers such as satirists and expressive critics.<sup>277</sup> Few commentators offer any defense of the pure commercial “free-rider,” typically a small, upstart that uses a famous trademark in a hapless way to grab attention.<sup>278</sup> Some examples from the case law include a medical equipment importer called “Nikepal,”<sup>279</sup> a perfume and beauty products site called “Perfumebay,”<sup>280</sup> an erotic novelty shop called “Victor’s Little Secret,”<sup>281</sup> a state tourism board that promotes its ski industry as “The Greatest Snow on Earth,”<sup>282</sup> a Syracuse coffee shop called “Federal Espresso,”<sup>283</sup> a coffee bean variety known as “Charbucks”<sup>284</sup> and a moving company’s whimsical depiction on its vans of a brown sofa emerging from a candy bar wrap-

---

277. *E.g.*, Dogan *supra* note 16 at 106 (arguing that the focus on free-riding in dilution law is misguided because even artists, critics and competitors looking to draw functional comparisons “free-ride” on famous brands in making their points); Tushnet, *supra* note 40 (arguing that dilution law violates First Amendment commercial speech doctrine rules because it fails to distinguish harm caused blurring from similar harms caused by protected critical and parodic speech).

278. *E.g.* Franklyn, *supra* note 275 (arguing for replacing dilution with a cause of action that explicitly targets free-riding on famous marks); Tushnet, *supra* note 40 (advocating a limited dilution cause of action aimed at deterring unfair competition through misappropriation of brand value).

279. *Nike Inc.*, 84 U.S.P.Q.2d (BNA) at 1820.

280. *Perfumebay.com Inc. v. eBay, Inc.*, 506 F.3d 1165 (9th Cir. 2007).

281. *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418 (2003).

282. *Ringling Bros.-Barnum & Bailey Combined Shows, Inc. v. Utah Div. of Travel Dev.*, 170 F.3d 449 (4th Cir. 1999).

283. *Fed. Express Corp. v. Fed. Espresso, Inc.*, 201 F.3d 168 (2d Cir. 2000).

284. *Starbucks Corp. v. Wolfe's Borough Coffee, Inc.*, 477 F.3d 765 (2d Cir. 2007).

per evocative of the Hershey chocolate bar's trade dress.<sup>285</sup> Each represents a blatant attempt to harness someone else's trademark for commercial value. Nonetheless, for the same reasons that persuasive advertising might offer useful cues about quality even if consumers don't consciously understand those cues, blurring may offer useful information to consumers. Blurring refers to a long-standing strategy used by newcomers to ease entry in crowded marketplaces. Consumers can benefit through blurring by learning about new product choices. Consumers also benefit by learning additional information about a brand owner and its products. Overregulation of the use of these symbols may be as socially harmful as under-regulation.

### 1. *Blurring Eases Barriers to Entry*

Dilution of well-known symbols for commercial gain has a long and distinguished pedigree. The use of a familiar symbol can attract consumer attention to an unfamiliar product. Such references inevitably free-ride on the famous marks, but few other strategies are as effective at commanding attention in oversaturated marketplaces.

A few historical examples can illustrate the point. Recall that the paradigmatic case of blurring involves the simultaneous use of the mark "Tiffany's" on a jewelry store and an unrelated restaurant.<sup>286</sup> The free-riding restaurant introduces irrelevant associations with the prestigious Tiffany's mark and so degrades its salience for consumers.<sup>287</sup> But of course, the Tiffany's brand did not always have the magnetism it does now. In 1837, Tiffany & Young (later Tiffany, Young & Ellis) opened as one of dozens of novelty and jewelry stores in lower Manhattan.<sup>288</sup> Perhaps in recognition of the lack of inherent distinctiveness in a personal name such as "Tiffany," the store chose a different branding mechanism to set itself apart. Tiffany's first newspaper advertisement was tinted with a singular shade of blue.<sup>289</sup> This shade of blue continues to play an important role in Tiffany's marketing efforts and was registered as a federal trademark in the United States in 2000.<sup>290</sup> Even in 1841, the blue was like-

---

285. *The Hershey Co. v. Art Van Furniture*, No. CV-14463, 2008 WL 4724756 at \*15 (E.D. Mich. Oct. 24, 2008) (finding a reasonable likelihood of success that Hershey would succeed at trial on its claim that the "couch bar" blurred its trade dress).

286. *See supra* text accompanying note 76.

287. *Id.*

288. Katie Sweeney, *Charles Tiffany Found All that Glittered in Jewelry*, INVESTOR'S BUS. DAILY, Oct. 26, 2007.

289. FirstMention.com, Tiffany's, <http://firstmention.com/tiffanys.aspx> (last visited Sept. 29, 2008).

290. U.S. Trademark No. 2,359,351 (filed Aug. 24, 1998) (issued June 20, 2000).

ly to be eye-catching but not only for aesthetic reasons. The color would already have been familiar to wealthy, well-educated consumers. A French painter Jean-Marc Nattier had pioneered its use in portraits of female members of the royal court at Versailles in the 1700s.<sup>291</sup> For French gentry of the time, the ability to commission a Nattier portrait, complete with his trademark combinations of silver and blue, was a token of relative affluence and influence.<sup>292</sup> Subsequently, at roughly the same time that Tiffany's adopted the color, the Empress Eugenie of France chose the hue as her signature color for dresses and upholstery.<sup>293</sup> The blue was thus a useful shorthand to a certain kind of consumer for European refinement and elite connections. Under the limited unfair competition laws then available, Nattier's estate would have had no conceivable cause of action against Tiffany's for use of the color. Nowadays, however, broader conceptions of what can operate as a "mark"<sup>294</sup> and extension of rights across national boundaries for well-known source-identifiers<sup>295</sup> might make Tiffany's conduct actionable as dilution. Tiffany's arguably made its name by blurring Nattier's trademark shade of blue.

---

291. Philippe Bordes, *Jean-Marc Nattier. Versailles*, 142 BURLINGTON MAG. 183, 184 (2000); see Neil Steinberg, *Happiness in a Box*, CHI. SUN TIMES, Nov. 7, 1999, at 13. Of course, even before Nattier, robins had pioneered use of the color on their eggs.

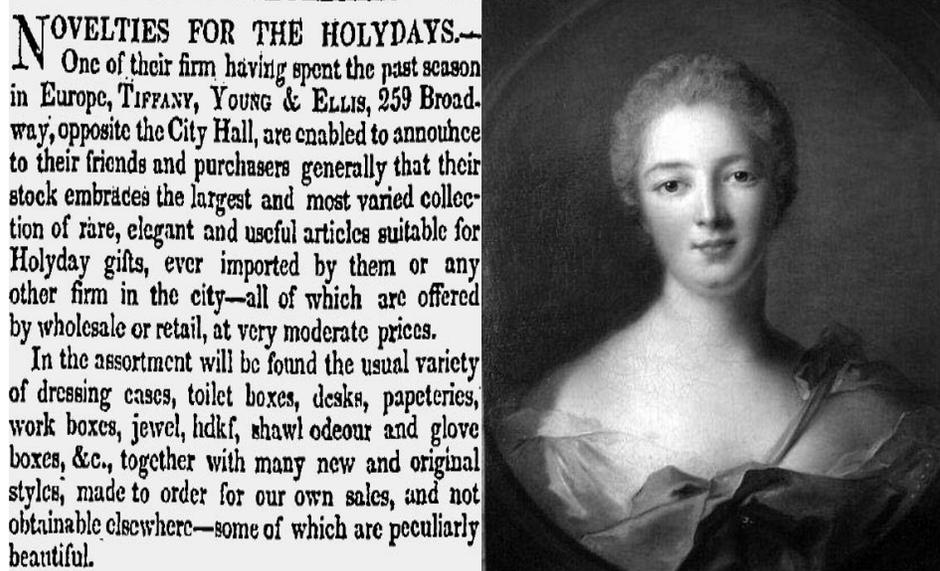
292. Bordes, *supra* note 291, at 184.

293. Connie Glaser, *Tips from Tiffany's*, BIZJOURNALS, Sept. 17, 2007, [http://www.bizjournals.com/extraedge/consultants/winning\\_at\\_work/2007/09/17/column144.html](http://www.bizjournals.com/extraedge/consultants/winning_at_work/2007/09/17/column144.html).

294. See *Romm Art Creations, Ltd. v. Simcha Int'l, Inc.*, 786 F. Supp. 1126 (E.D.N.Y. 1992) (painting style can operate as a trademark).

295. *E.g.*, *Grupo Gigante SA De CV v. Dallo & Co.*, 391 F.3d 1088 (9th Cir. 2004) (recognizing rights of a foreign producer even without commercial use of the mark in the territorial United States because the mark was famous among the relevant class of producers). This rule has not been universally followed. See, *e.g.*, *ITC Ltd. v. Punchgini, Inc.*, 482 F.3d 135, 163-65 (2d Cir. 2007) (finding no famous mark exception to the use requirement in federal trademark law).

Figure 1: Tiffany's first newspaper advertisement in 1841 vs. J.M. Nattier's 1748 portrait of Madame de Pompadour



Nor is Tiffany's the only owner of a famous trademark to engage in a little blurring on its way up. McDonald's is another trademark owner who has been aggressive about policing free-riders who capitalize on the informational salience of its marks.<sup>296</sup> But McDonalds itself was once a scrappy upstart trying to capture attention. As many newcomers do, its proprietor, Ray Kroc, turned to a symbol that was already well-known and popular among his targeted market. He hired Willard Scott, the actor who had just played the popular Bozo the Clown character on a popular children's TV show, to create a purposefully similar character to promote his own brand.<sup>297</sup> The resulting "Ronald McDonald" would not necessarily have caused confusion with Bozo, but the presence of the same actor with

296. See, e.g., *Quality Inns Int'l, Inc. v. McDonald's Corp.*, 695 F. Supp. 198 (D. Md. 1988).

297. See WILLARD SCOTT, *THE JOY OF LIVING* 132-33 (1982). Willard Scott stated: At the time, Bozo was the hottest children's show on the air. You could probably have sent Pluto the Dog or Dumbo the Elephant over and it would have been equally as successful. But I was there, and I was Bozo . . . There was something about the combination of hamburgers and Bozo that was irresistible to kids. That's why when Bozo went off the air a few years later, the local McDonald's people asked me to come up with a new character to take Bozo's place. So, I sat down and created Ronald McDonald.

*Id.*

the same voice in a similar costume surely would have sustained a blurring claim had the cause of action been available at the time.<sup>298</sup> Whether the inability to free-ride on Bozo's familiarity would have prevented McDonald's from becoming the "famous brand" it is today is unknowable.

A third example is Coca-Cola, who in the 1930s used illustrations of Santa Claus to persuade consumers to drink its beverage even in cold weather.<sup>299</sup> Not only was Coca Cola "free-riding" off of St Nicholas' positive emotional associations, it was not even the first company to do it. White Rock Ginger Ale, the brand leader for mineral water and ginger ale, had already used Santa Claus in its advertising for 15 years.<sup>300</sup> Coke's advertising campaign may have been an attempt to create associations with a more successful beverage company.<sup>301</sup>

Modern free-riders similarly use the "atmospherics" surrounding well-known brands to communicate information about their own products.<sup>302</sup> Rochelle Dreyfuss has written that modern brands have replaced classic literature as the basis for rhetorical and literary allusion: "Betty Crocker has replaced Hestia in the public consciousness. Accordingly, it is not surprising that speakers and writers are drawn to those devices that are, by dint of heavy advertising, universally familiar."<sup>303</sup> Professor Dreyfuss was referring mostly to noncommercial writers and speakers, but commercial speakers are also interested in communicating forceful messages with economy. For example, Natural Answers, Inc., the makers of an herbal mood uplifter, chose the mark "Herbrozac" to communicate suc-

---

298. *Cf.* Nabisco, Inc. v. PF Brands, Inc., 191 F.3d 208 (2d Cir. 1999) (finding goldfish cracker character too similar and likely to blur Nabisco's famous goldfish-shaped snack).

299. *See, e.g.*, Snopes.com, Coca-Cola Invents Santa Claus? (Dec. 25, 2007), <http://www.snopes.com/cokelore/santa.asp>.

300. *Coca-Cola's Santa Claus, Not the Real Thing!*, PR NEWSWIRE, Dec. 15, 2006, available at WestLaw, 12/15/06 PR Newswire 16:56:00; Wikipedia, Santa Claus, [http://en.wikipedia.org/wiki/Santa\\_Claus](http://en.wikipedia.org/wiki/Santa_Claus) (last modified Sept. 19, 2008); *see also* White Rock Collectors Association, Santa Claus and White Rock from 1915, <http://www.whiterocking.org/santa.html> (last visited Sept. 26, 2008).

301. Coca-Cola would later sue a competitor for similar conduct involving the use of polar bears in advertising. Coca-Cola later dropped the claim that use of even a non-similar polar bear by another drink company would cause trademark dilution. *See* Polar Corp. v. Coca-Cola Co., 871 F. Supp. 1520 (D. Mass. 1994).

302. *Cf.* Litman, *supra* note 132, at 1735 (arguing that if we think that atmospherics underlying famous brands offer something valuable to consumers, then we should prefer to let different producers compete to offer that value instead of assigning monopoly rights to one party).

303. Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 424 (1990).

cinctly that the product was an herbal alternative to Prozac, the leading antidepressant drug made by Eli Lilly.<sup>304</sup> The Seventh Circuit found the use dilutive of Lilly's trademark.<sup>305</sup> The Court suggested that Natural Answers could have achieved the same informational effect legally with a name such as "Natural Answers' Herbal Mood Elevator."<sup>306</sup> The same argument could be made about any clever slogan: Nike could have called its shoe Speedy Shoe instead of referring to the Greek goddess of victory, but most people would agree that something is lost in the translation. Furthermore, if we agree that Natural Answers can legally say "just like Prozac" in its advertising, it seems formalistic to deny use of the phrase as a slogan or even a mark.

If we are unwilling to say *ex post* that we would be better off if Bozo and not Ronald were still on the air and White Rock and not Coke still ruled the shelves, then we should be cautious about limiting the ability *ex ante* of newcomers to adopt similar free-riding strategies. Each of the historical examples above could be offered as an argument in support of strong dilution regulation: Nattier, Bozo and White Rock are not household names anymore. One could argue that the blurring of their trade symbols quickened their decline. It seems more likely, however, that free-riding brought better products to customer attention and the older brands collapsed under their own weight. The companies behind Herbrozac, Charbucks and Nikepal may seem of little concern now, but perhaps they will be tomorrow's Tiffany's, Coca-Cola and McDonalds.<sup>307</sup>

## 2. *Blurring Can Increase Useful Product Information Available to Consumers*

Dilution by blurring may also produce beneficial effects by correcting for persistent failures of information supply in product markets. Current intellectual property regimes reward brand advertising over product-related, informational advertising.<sup>308</sup> By lowering the incentives to advertise based on brand, blurring could increase the level of useful product information in the marketplace. This might lead to gains in social welfare that outweigh the costs of lost convenience.

---

304. *Eli Lilly & Co. v. Natural Answers, Inc.*, 233 F.3d 456, 465 (7th Cir. 2000).

305. *Id.* at 469.

306. *Id.* at 465.

307. *Cf. Siegfried & Evans, supra* note 269, at 123 (noting that two stage entry is a common pattern in which a small firm first enters into a niche and then later expands into the mainstream of the industry).

308. *See infra* text accompanying notes 319-320.

Consumer product markets suffer from information asymmetries because sellers have better information than buyers about the true value of goods and services.<sup>309</sup> Many products require use and experience for proper evaluation. Some products, such as vitamins or education, resist effective evaluation by users and users must take it on faith that sellers have given them what they need.<sup>310</sup> Sellers know the value of such goods, but buyers do not (at least before purchase). Buyers must depend on signals from the sellers or other credible parties to form judgments *ex ante* about the quality of different choices. Advertising is one form of signaling. Trademarks are another. As discussed above, sellers with well-known trademarks have an incentive to provide truthful advertising and to police the quality of goods sold under the mark.

However, sellers may under-provide product information and over-provide brand information. Sellers have an incentive to provide information to consumers that will cause them to prefer the seller's goods or services, but only to extent that the marginal benefits of providing that information equal the marginal costs.<sup>311</sup> All else being equal, sellers may under-provide general product information because other sellers of the same product can free-ride off of that expenditure. Unless the seller has a patent or other ability to exclude competitors from acquiring beneficial features of a product, it has little incentive to make these attributes the center of their advertising campaigns. Thus, while providing general information about products provides a social benefit, it will be undersupplied to the extent that sellers cannot capture that benefit as additional revenue to themselves.<sup>312</sup> By contrast, sellers can capture the benefits of brand-specific advertising because they can exclude others from using their brands. Brand advertising diverts customers from competitors, but may do little to increase social welfare overall.<sup>313</sup> Therefore, sellers have an incentive to oversupply brand-specific information and under-supply product-specific information.<sup>314</sup>

---

309. See Nelson, *Information and Consumer Behavior* *supra* note 148 at 311-12; Stigler, *supra* note 27.

310. Ekelund et al., *supra* note 261, at 34.

311. Beales et al., *supra* note 29, at 508.

312. *Id.*

313. Advertising serves a beneficial function by giving consumers information about product choices and so increases price elasticity. However many studies suggest that informational advertising does a better job at this than persuasive brand-oriented advertising. *E.g.* William Boulding et al., *The Long-Term Differentiation Value of Marketing Communication Actions* (Mktg. Scis. Inst., Working Paper No. 92-133, 1992) (informational advertising increases elasticity but persuasive advertising decreases it).

314. Beales et al., *supra* note 29, at 509.

Comparative advertising is one way to correct this asymmetry. By allowing competitors to factually compare their products to more established brands by name, the law encourages newcomers to provide informational advertising about product features. This increases the amount of useful product-related information available to the market.<sup>315</sup> It also allows competitors to legally free-ride off the recognition of the senior owner's mark.<sup>316</sup>

However, direct competitors to the mark owner may undersupply comparative factual information for reasons similar to the original sellers. Competitors have some incentive to provide critical information about product and brand flaws of dominant sellers, but information that discourages consumers from patronizing one seller will equally benefit all other sellers in that product class.<sup>317</sup> In non-oligopolistic markets, competitors may undersupply information about dominant sellers because they cannot capture all of the gains to themselves. Furthermore, the relatively narrow comparative advertising exception still favors the owners of famous marks because they can rely exclusively on affect-laden advertising using the familiar mark.<sup>318</sup>

Those who blur famous trademarks can exploit these asymmetries and may even help correct for them. Users of a mark in unrelated markets have a greater incentive to exploit critical or alternative understandings of the senior mark.<sup>319</sup> Consumer interest in this information will bring atten-

315. I am indebted to Bruce Kobayashi for this point.

316. Dogan, *supra* note 16, at 106 (noting that the Pepsi challenge was a form of free-riding on Coca-Cola's brand).

317. Beales et al., *supra* note 29, at 508; Xavier Gabaix & David Laibson, *Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets* (Mass. Inst. of Tech. Dep't of Econ., Working Paper No. 05-18, 2005), available at <http://ssrn.com/abstract=728545> (noting that sellers may not inform consumers about hidden add-on mark-ups in competing products if the add-ons have close substitutes because the information will not benefit the discloser).

318. *E.g.* *New Kids on the Block v. News Am. Publ'g, Inc.*, 971 F.2d 302, 307-308 (9th Cir. 1992) (stating that second users that attempt to "appropriate the cachet" of well-known marks to themselves are not eligible for nominative fair use defense).

319. Companies that adopt slogans or symbols that blur more famous trademarks are guided by the same rational self-interest that governs the senior mark owner. Economists assume that because advertising raises expectations, it will only be efficient for those companies that can deliver on the heightened expectations of consumers. Adoption of a well-known mark involves a variation of the same cost-benefit analysis. The use ensures attention from consumers with relatively little advertising cost. In this respect the newcomers are misappropriating the advertising expenditures of the bigger players. However, free-riding will only be efficient for those who will benefit in a sustained manner from the increased attention. Blurring requires less expenditure, but any kind of promotion requires costs. Use of a well-known mark will raise consumer expectations about the un-

tion to the new product. The free-rider will benefit in a way that cannot effectively be shared by its competitors,<sup>320</sup> And the new advertising can offer indirect information about the senior mark.

For example, many blurring defendants use the senior mark in a way that reveals whimsical or critical opinions of the brand owner and its products. Haute Diggity Dog, the maker of a “Chewy Vuitton” luggage dog chew toy, presumably chose to mock “Louis Vuitton” because some people see the brand as pretentious.<sup>321</sup> Similarly, the sale of “Charbucks” coffee by a small retailer in New Hampshire publicizes a common perception that Starbucks coffee tastes burnt. Even the Syracuse coffee joint Federal Espresso mocks the over-caffeinated pace which services such as Federal Express make possible. Because in each case the defendant uses the famous mark in its own “source-identifier,” none of these uses are protected by statutory exclusions for parody or criticism.<sup>322</sup> Such uses are profitable for blurrers because consumers may appreciate seeing the famous mark unmasked and, as a result, may reward the junior user with attention. The blurring information may eventually alter a consumer’s prefe-

---

derlying product. Only those companies that have the capacity to meet consumer expectations will benefit from the use. Those that fall short will receive a negative response from consumers. Such companies may receive a short-term bump from the exposure, but the same would be true of an initial choice to advertise. Free-riders who fail to meet expectations will either switch to promotion strategies that highlight their true competitive advantages (e.g. lower cost) or will cease advertising altogether. In this latter case, blurring will cease to be an issue for the senior mark owner and its consumers. For the same reason that the choice to advertise can signal quality, the choice to anchor advertising messages in a famous symbol usually signals some kind of relevance to the message conveyed by the familiar symbol.

320. Blurring is likely to reveal new information about the senior brand because uses of famous brands that do not reveal new information will be confusing. Classic trademark law already forbids any use of a mark that is likely to confuse consumers. Any attempt by a newcomer, or an established player, to masquerade behind the senior producer’s mark will be enjoined as pure trademark infringement. Laws against blurring thus chiefly regulate uses of marks that alter the context of the mark sufficiently to preclude confusion. Furthermore, although trademark dilution articles all invoke the specters of “Kodak pianos,” “Buick shoes,” etc., it is difficult to see why anyone would adopt such a mark. Because famous brands are invariably associated with the senior mark owner, it will be difficult for the junior user to differentiate his own product or create his own goodwill in the second mark. Perhaps for this reason, very few of the reported cases actually concern exact use of a well-known mark by a free-rider in a new market.

321. *Louis Vuitton Malletier S.A. v. Haute Diggity Dog, LLC*, 507 F.3d 252 (4th Cir. 2007) (dismissing trademark dilution case against maker of dog toy luggage).

322. However, courts may still find them not likely to dilute on the basis of the statutory blurring factors. *See, e.g., Starbucks Corp. v. Wolfe’s Borough Coffee, Inc.*, 559 F. Supp. 2d 472, 477-79 (S.D.N.Y. 2008); *Haute Diggity Dog*, 507 F.3d. 252.

rence for associating herself with the original brand, but that may be a welfare-enhancing choice.

Even if the new message offers no new information, allowing others to benefit from the familiarity effects of famous marks would lower incentives to invest in persuasive brand advertising. Advertising signals quality because of the perceived investment in the product by the underlying firm. Both product-related (informative) advertising and brand-related (persuasive) advertising can send this signal.<sup>323</sup> Consumers thus have no functional reason to prefer persuasive advertising over informational ads.<sup>324</sup> This suggests that sellers over-supply persuasive advertising, and that consumers would benefit from a shift in advertising from persuasive to informational strategies.<sup>325</sup> Decreasing the monopoly profits available for familiarity divorced from function might help to instigate such a shift.

One must be careful here, though. If blurring reached a level at which it lowered the incentives to invest in familiar brands and in product quality generally, that could harm social welfare. This would harm consumers because without the signaling effect of strong brands, they would have to work harder to make judgments about risk. Without quality control, they would face a greater risk of poor quality choices. However, there are two reasons to think that the effects of increased blurring on seller advertising would be minimal. First, because consumers sometimes use System II decision-making, sellers will still have incentives to make quality products and to provide information about quality that might influence purchase decisions. If consumers rely less on System I processing, they may spend more time comparing actual product features. Sellers may then choose to provide more product-related advertising and less brand-related, emotionally charged information. For the reasons discussed above, this would probably increase social welfare. Second, the research suggests that inconsistency effects must be significant before attitudes toward a familiar brand will change.<sup>326</sup> The ability to police against confusing uses likely will exclude a lot of potential free-riding. Because the impact of most free-riding is likely to be small, dilution laws should be targeted in scope.

---

323. Brown, *supra* note 7.

324. Furthermore, some empirical research suggests that persuasive advertising can lower price elasticity for certain goods, while informational advertising can raise it. *E.g.*, William Boulding et al., *The Long-Term Differentiation Value of Marketing Communication Actions* (Mktg. Scis. Inst., Working Paper No. 92-133, 1992).

325. *Cf.* Vakratas & Ambler *supra* note 232 (suggesting that sellers oversupply advertising relative to consumer preferences).

326. *See supra* text accompanying notes 228-230.

### 3. *Other Alternatives to Dilution Regulation Exist*

In the absence of dilution regulation, private efforts to help consumers reduce search costs would still exist. Regulation of blurring can be compared to government attempts to “debias” the presentation of information to assist consumers in making better choices.<sup>327</sup> The risk with such debiasing efforts is that they are prone to regulatory capture and displace more targeted private attempts to address information asymmetries.<sup>328</sup> In the case of dilution, regulatory debiasing overcorrects for the problem by distorting market information in favor of strong brands. It removes a beneficial source of information about flaws in established choices from the market. It also forbids competition over the meaning of familiar symbols, and so creates incentives for wasteful advertising expenditures on brand atmospherics instead of more useful advertising on product attributes.

Dilution law also represents a preference for brand strength as the primary heuristic through which consumers increase convenience and lower their “internal search costs.” Strong brands may not represent the optimal social or industrial policy for achieving this goal, however. If widespread blurring lowered the ability to rely on the risk-reduction signals of super-brands, consumers would most likely adopt alternative strategies to conserve time and effort. Markets will always suffer from information asymmetry because sellers have better information about the quality of their goods than buyers do. However, technology is providing new ways beyond advertising and branding to correct this asymmetry. For example, in online markets made up of relatively unknown sellers (such as eBay and Etsy, a hand-made crafts site), users rely on aggregated satisfaction ratings to lower the risks of dealing with new sellers. As the costs of aggregating buyers’ reactions go down, the need to rely exclusively on quality signals generated by sellers also declines. In other markets, buyers can rely on the credibility of third-party taste-makers to lower the costs of search.<sup>329</sup> Many consumers go to specialized markets (such as Whole Foods) despite a comparative lack of familiar brands on its shelves because they trust the store to cater to their tastes for premium and naturally-produced items. As markets everywhere become more segmented, the need for universal “famous” brands to lower risk declines.

---

327. See, e.g., Richard A. Epstein, *Behavioral Economics: Human Errors and Market Corrections*, 73 U. CHI. L. REV. 111 (2006) (criticizing government attempts to debias consumer information markets as ill-advised and prone to capture and over-correction).

328. *Id.*

329. See, e.g., Wujin Chu & Woosik Chu, *Signaling Quality by Selling through a Reputable Retailer: An Example of Renting the Reputation of Another Agent*, 13 MARKETING SCI. 177 (1994).

The choice of whether to protect against blurring thus depends on a complicated balancing between consumer interests in convenience and equally legitimate interests in competition. Famous brands can provide valuable quality signals to overburdened consumers, but blurring can ease market entry for newcomers. Both sides of the dilution debate attempt to make an easy case out of a situation that is in fact quite nuanced.

#### **D. Limiting Dilution to Use of the Exact Mark**

One way to balance the competing concerns underlying trademark dilution would be to continue to prohibit commercial uses of identical or nearly identical marks, but to allow greater latitude for more than trivial variations.<sup>330</sup> Uses of exact replicas of famous marks are likely to cause confusion no matter what segment of the market they target.<sup>331</sup> Dilution protection can lower enforcement costs by removing the need to meet trademark's more onerous multi-factored confusion test for this special class of marks.<sup>332</sup> A nontrivial variation, however, such as Charbucks, Herbrozac, Chewy Vuitton, Federal Espresso, McDental, "Greatest Snow on Earth," etc., usually signifies either harmless word-play or implicit comparison. Revising the federal dilution law to require identity between marks would conserve consumer confidence in the brand signal while allowing newcomers to use it to draw attention to their own offerings.

#### **V. BEYOND DILUTION: DECOUPLING CONSUMER SURVEY RESULTS AND THE EFFICIENT MARKET**

Whether regulation of trademark dilution is a game worth the candle is not as easy as dilution law's critics contend. The ability of consumers to rely on the credibility of established brands offers real benefits. On the other hand, the ability of new producers to reference stronger marks eases barriers to entry. Arguments exist on both sides as to which policy (protecting against blurring or allowing it to occur) is the best policy.

What *cannot* decide the question is naked data from consumer surveys. In dilution cases, plaintiffs offer as evidence that people associate a

---

330. Some circuits embrace this kind of standard, but individual decisions vary widely as to what will be found "nearly identical" with a famous mark. *See, e.g., Nike Inc. v. Nikepal Int'l Inc.*, 84 U.S.P.Q.2d (BNA) 1820 (E.D. Cal. 2007) (articulating a near identity standard but finding that "Nikepal" is nearly identical with Nike).

331. *See, e.g., Klieger, supra* note 123.

332. *See Bone, Enforcement Costs, supra* note 38 (arguing that dilution prohibitions lower the costs of enforcing famous marks in situations where confusion is likely).

junior mark with a famous one,<sup>333</sup> evidence of an increase in the time required to connect a famous mark with its original owner,<sup>334</sup> or direct evidence of an increase in negative feelings about the senior mark after exposure to blurring.<sup>335</sup> Evidence of association, without proof of confusion, does not indicate any diminishment in the effectiveness of the senior mark. An increase in time or effort to evaluate a mark may generate negative feelings towards the mark, just as the convenience of a familiar brand automatically generates positive feelings. These feelings are not diagnostic by themselves. They must be situated within a larger network of preferences about price, product attributes, optimal number of sellers, optimal attributes of sellers, types of retailers, and larger social policy questions. Momentary irritation tells us nothing about what balance of interests the consumer ultimately prefers. Only the individual consumer herself can integrate these competing motivations. Perhaps she ought to be given more chance to do so.

---

333. Nike Inc., 84 U.S.P.Q.2d (BNA) at 1820; Jacoby, *Dilution in Light of Victoria's Secret*, *supra* note 15, at 10.

334. Morrin & Jacoby, *supra* note 40, at 271.

335. Julie Manning Magid et al., *Quantifying Brand Image: Empirical Evidence of Trademark Dilution*, 43 AM. BUS. L.J. 1, 34, 38 (2006).

# RECENT DEVELOPMENTS AFFECTING THE ENFORCEMENT, PROCUREMENT, AND LICENSING OF RESEARCH TOOL PATENTS

By Joshua D. Sarnoff<sup>††</sup> & Christopher M. Holman<sup>††</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1300
II. DEFINITION OF RESEARCH TOOLS.....	1302
III. BRIEF HISTORY OF THE EXPERIMENTAL-USE AND REGULATORY-APPROVAL EXCEPTIONS .....	1303
A. ORIGINS AND EARLY INTERPRETATIONS OF THE EXPERIMENTAL-USE EXCEPTION .....	1304
B. THE 1984 <i>BOLAR</i> DECISION AND LEGISLATIVE ADOPTION OF THE REGULATORY-APPROVAL EXCEPTION .....	1306
C. SUBSEQUENT FEDERAL CIRCUIT INTERPRETATIONS NARROWLY CONSTRUING THE EXPERIMENTAL-USE EXCEPTION.....	1308
D. PROPOSALS FOR LEGISLATION TO CODIFY A BROADER EXPERIMENTAL-USE EXCEPTION .....	1310
E. <i>MERCK V. INTEGRA</i> : INITIAL FEDERAL CIRCUIT AND SUBSEQUENT SUPREME COURT INTERPRETATIONS OF THE REGULATORY-APPROVAL EXCEPTION .....	1314
F. CASES INTERPRETING THE REGULATORY-APPROVAL EXCEPTION SINCE THE 2005 SUPREME COURT DECISION IN <i>MERCK</i> .....	1316
IV. RECENT STUDIES OF SCIENTIFIC RESEARCHER AND PATENT HOLDER PRACTICES.....	1320
A. THE WALSH, ARORA, AND COHEN STUDY (2003).....	1322
B. THE WALSH, CHO, COHEN STUDY (2005) .....	1325
C. THE AAAS STUDY (2006-2007).....	1326

---

© 2008 Joshua D. Sarnoff & Christopher M. Holman. This article was originally prepared for inclusion by IFTECH in an internal report for the Japanese Patent Office on the worldwide state of research tool patents.

† Practitioner-in-Residence & Assistant Director, Glushko-Samuelson Intellectual Property Law Clinic, Washington College of Law, American University.

†† Associate Professor of Law, University of Missouri, Kansas City School of Law.

D.	THE HOLMAN HUMAN GENE PATENT LITIGATION STUDY (2007).....	1327
V.	<b>RECENT CHANGES TO PATENTING AND LICENSING POLICIES AND PRACTICES</b> .....	1331
VI.	<b>RECENT AND PROPOSED CHANGES TO THE PATENT SYSTEM THAT MAY AFFECT RESEARCH TOOL PATENTS AND USE OF RESEARCH TOOLS</b> .....	1337
A.	THE UTILITY REQUIREMENT OF SECTION 101 .....	1339
B.	THE PATENTABLE SUBJECT MATTER REQUIREMENT OF SECTION 101 .....	1340
C.	THE NONOBVIOUSNESS REQUIREMENT .....	1343
D.	INJUNCTIVE RELIEF UNDER SECTION 283 .....	1346
E.	POTENTIAL LEGISLATION AFFECTING DAMAGES REMEDIES UNDER SECTION 284.....	1349
VII.	<b>ALTERNATIVES TO EXPERIMENTAL-USE AND REGULATORY-APPROVAL EXCEPTIONS TO INFRINGEMENT</b> .....	1350
A.	COMPULSORY LICENSING .....	1351
B.	GOVERNMENT RIGHTS TO INVENTIONS PATENTED UNDER THE BAYH-DOLE ACT .....	1356
C.	REACH THROUGH LICENSING AGREEMENTS.....	1358
D.	PATENT POOLS.....	1359
E.	ANTITRUST APPROACHES.....	1361
F.	OFF-SHORING RESEARCH.....	1364
VIII.	<b>CONCLUSIONS</b> .....	1365

## I. INTRODUCTION

This Article summarizes recent developments under U.S. patent laws and provides insights into the practices of various academic sciences, industries, and government agencies regarding the treatment of so-called “research tool” inventions. For many years a vigorous discussion has existed about the need to provide exclusive patent rights as incentives to invent and to disclose research tools, whether such exclusive rights should apply to all uses and users of patented research tools, and whether exclusive rights to prohibit all uses of research tools would unduly discourage sequential invention. Concerns over the proper scope of patent rights in regard to subsequent research uses of inventions have a long history, but have received increased scrutiny in light of judicial decisions since the

turn of the century. New studies of uses of research tools and efforts to assert research tool patents have been performed in light of the decision of the U.S. Court of Appeals for the Federal Circuit (Federal Circuit) in 2002 to provide a restrictive interpretation of the experimental-use exception to patent infringement in *Madey v. Duke University*,<sup>2</sup> and the decision of the U.S. Supreme Court in 2005 to provide an expansive interpretation of the codified regulatory-approval exception<sup>3</sup> in *Merck KGaA v. Integra LifeSciences I Ltd.*<sup>4</sup> This Article seeks to describe the broad parameters of these developments.

In general terms, the law regarding patents and research tool inventions has become clearer since 2000. The Federal Circuit's 2002 *Madey* decision has increasingly been recognized as expressing the state of the law regarding the experimental-use exception, particularly as neither the U.S. Supreme Court nor the U.S. Congress have chosen to intervene to revise the Federal Circuit's approach. Thus, uses of patented research tools in almost all contexts, even for university-based basic research, must for now be considered an actionable infringement of exclusive patent rights. Only in the context of the regulatory-approval exception does significant uncertainty remain regarding whether uses of patented research tools constitute actionable infringements. In that context, the broad language of the Supreme Court's interpretation in *Merck*, the subsequent decision of the Federal Circuit on remand,<sup>5</sup> and other recent cases suggest that the exception may apply to at least some research tool uses of inventions closely related to the target of the regulatory-approval decision.

At the same time, social practices have become more complex. Recent studies demonstrate that both academic and commercial researchers ignore the actual state of the law and routinely use patented inventions without the authorization of patent holders. This approach appears justified in light of other studies that demonstrate that many research tool patent holders will not assert their patents to restrict research. However, fears of potential liability may nevertheless unduly restrict research, and routine disregard of legal rights (even if unlikely to be asserted) may not be a stable position. In some contexts, such as diagnostic and stem-cell inventions, aggressive assertion of research tool patents has led to public criticism, and new academic and government guidelines have developed to assure broad licensing of research tools on reasonable terms.

---

2. See 307 F.3d 1351 (Fed. Cir. 2002).

3. 35 U.S.C. § 271(e)(1) (2000).

4. 545 U.S. 193 (2005).

5. *Integra Lifesciences I, Ltd. v. Merck KGaA*, 496 F.3d 1334 (Fed. Cir. 2007).

This Article provides basic definitions, briefly describes the history of the experimental-use and regulatory-approval exceptions and their application to research tools, and then summarizes recent developments in the case law, studies of recent practices of researchers and patent holders, and recent changes to licensing policies in regard to research tools. It also provides a brief discussion of alternatives to a broad experimental-use exception and throughout contains references to relevant academic articles.

## II. DEFINITION OF RESEARCH TOOLS

This Part provides a basic definition of research tools addressed in this Article. A broad definition is adopted because the focus of this Article is on the effects of potential patent liability on scientific research. Narrower definitions are more applicable in the discussions about incentivizing the development of technologies intended for research.

“Research tools” may have many definitions and may include a very wide range of technologies. For example, patented inventions covering the following are all sometimes referred to as research tools: cell lines, genetic sequences, assay methods, software, and instruments such as microscopes and lasers. Research tools are often defined as inventions whose patent application discloses that their intended use is solely or principally for scientific research.<sup>6</sup> However, this definition is problematic because technologies are commonly used for research that the patent holder did not contemplate, and the right to exclude others from using patented inventions is not limited in the United States to the disclosed and claimed uses.<sup>7</sup> Another approach defines research tools as patented technologies that are only used to produce products that do not incorporate the technology.<sup>8</sup> This approach focuses only on liability for the research market because there is no infringement by the products produced with the research tool. For purposes of discussing the full scope of potential liability, one must consider a broader definition of research tools than inventions that are pa-

---

6. See, e.g., Philippe Ducor, *Research Tool Patents and the Experimental Use Exception—a No-Win Situation?*, 17 NATURE BIOTECHNOLOGY 1027, 1027-28 (1999); Thomas D. Mays, *Sharing Biomedical Research Resources: Principles and Guidelines for Recipients of NIH Research Grants and Contracts: Race Horse or Trojan Horse?*, 3 BIO-SCI. L. REV. 56, 61 (1999-2000).

7. See 35 U.S.C. § 271(a) (2000).

8. See, e.g., Janice M. Mueller, *No “Dilettante Affair”:* Rethinking the Experimental Use Exception to Patent Infringement for Biomedical Research Tools, 76 WASH. L. REV. 1, 14-15 (2001); Esther Pfaff, *“Bolar” Exemptions—A Threat to the Research Tool Industry in the U.S. and the EU?*, 38 INT’L REV. OF INTELL. PROP. & COMPETITION L. 258, 262-63 (2007).

tented with a disclosed research purpose or inventions that are used to produce, but are not incorporated into, a commercialized product. Nevertheless, such patents are often the subject of greatest concern regarding the need for patent protection, given that the anticipated market for any commercial value for the patent is for research.

More expansive definitions of “research tools” focus on the possible uses of patented inventions. A recent Federal Circuit case defined research tools as “tools that scientists use in the laboratory, including cell lines, monoclonal antibodies, reagents, animal models, growth factors, combinatorial chemistry and DNA libraries, clones and cloning tools (such as PCR), methods, laboratory equipment and machines.”<sup>9</sup> Similarly, analysts and academics have defined research tools broadly as “any . . . input into the process of discovering” products<sup>10</sup> and as “the technological developments that enable particular lines of research to be pursued.”<sup>11</sup> We rely on these more expansive definitions below, so for purposes of this Article “research tools” are patented technologies used in conducting research that are not themselves the object of the research inquiry at that time. It bears repeating, however, that this expansive definition applies to many types of patented technologies having different intended markets than the research at issue.

### III. BRIEF HISTORY OF THE EXPERIMENTAL-USE AND REGULATORY-APPROVAL EXCEPTIONS

The following Part summarizes the origins and history of judicial interpretations of the experimental-use and regulatory-approval exceptions in the United States.<sup>12</sup> The summary identifies significant changes over

---

9. *Integra LifeSciences I Ltd. v. Merck KGaA*, 331 F.3d 860, 872 n.4 (Fed. Cir. 2003) (Newman, J., dissenting) (quoting *Sharing Biomedical Research Resources: Principles and Guidelines for Recipients of NIH Research Grants and Contracts*, 64 Fed. Reg. 72,090, 72,092 n.1 (Dec. 23, 1999)).

10. John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in *PATENTS IN THE KNOWLEDGE-BASED ECONOMY* 287 (Wesley M. Cohen & Stephen A. Merrill eds. 2003).

11. Dianne Nicol, *Cooperative Intellectual Property in Biotechnology*, 4 *SCRIPT-ED* 136, 137 (2007), available at <http://www.law.ed.ac.uk/ahrc/script-ed/vol4-1/nicol.asp>.

12. The summary is largely based on an article co-authored by one of the authors of this Article (Sarnoff), which provides additional details and a comparison to European law. See Henrik Holzapfel & Joshua D. Sarnoff, *A Cross-Atlantic Dialog on Experimental Use and Research Tools*, 48 *IDEA* 123 (2008). For another comparative perspective, see Sean O'Connor, *Enabling Research or Unfair Competition? De Jure and De Facto Research Use Exceptions in Major Technological Countries*, in *COMPARATIVE PATENT*

time to the scope of the experimental-use exception, as well as unresolved questions regarding its basic nature and regarding the nature, scope, and application of the regulatory-approval exception.

### A. Origins and Early Interpretations of the Experimental-Use Exception

Supreme Court Justice Joseph Story first articulated the experimental-use exception in the United States in the early Nineteenth Century. As Justice Story stated in 1813 in *Whittemore v. Cutter*,<sup>13</sup> “it could never have been the intention of the legislature to punish a man, who constructed such a machine merely for philosophical experiments, or for the purpose of ascertaining the sufficiency of the machine to produce its described effects.”<sup>14</sup> The patent statute at the time provided liability for any person who shall “make, devise, use, or sell” the patented invention, and the statutory language had been amended earlier to make clear that making without use constituted an infringement of the exclusive right.<sup>15</sup> Thus, the language of the *Whittemore* decision may be understood in one of two ways—either as a statutory interpretation of the limits of the specific rights granted by a patent, or as a judicially imposed exception to the rights granted, consistent with the more extensive judicial common law-making powers of the time. The distinction is significant, both substantively and procedurally, as the first approach would define the limits of property initially granted and the second approach would impose restrictions (in the nature of an affirmative defense to liability) on the use of that property.<sup>16</sup> The dispute over which approach is correct has not yet been settled, but the experimental-use exception is most frequently referred to as a “common law” exemption from liability.<sup>17</sup>

The *Whittemore* decision articulated two different grounds for an experimental-use exception to patent infringement. The first was for “philosophical experiments,”<sup>18</sup> and the second was to “ascertain . . . sufficien-

---

LAW: A HANDBOOK OF CONTEMPORARY RESEARCH (Toshiko Takenaka & Rainer Mounfang, eds., forthcoming 2008).

13. 29 F. Cas. 1120 (C.C.D. Mass. 1813) (No. 17,600).

14. *Id.* at 1121.

15. Act of Apr. 17, 1800, ch. 25, § 3, 2 Stat. 37, 38 (1800) (current version at 35 U.S.C. § 271(a) (2000)); see Act of Feb. 21, 1793, ch. 11, § 5, 1 Stat. 318, 321 (1793).

16. See O’Conner, *supra* note 12, at 3, 7 (discussing alternative treatment of the doctrine as an exception or as an exemption).

17. See, e.g., *Integra LifeSciences I, Ltd. v. Merck KGaA*, 331 F.3d 860, 863 n.2 (Fed. Cir. 2003).

18. At the time, “philosophical experiments” was understood to mean scientific research, particularly in physics. See, e.g., THE COMPACT EDITION OF THE OXFORD ENG-

cy”<sup>19</sup> of the patented invention for the disclosed uses. The scope of these two prongs of the exception has been the subject of extensive dispute and numerous cases over the course of the next two centuries.

In *Sawin v. Guild*,<sup>20</sup> Justice Story sought to clarify further the scope of the exception as follows:

[T]he making of a patented machine to be an offence within the purview of it, must be the making with an intent to use for profit, and not for the mere purpose of philosophical experiment, or to ascertain the verity and exactness of the specification. In other words, that the making must be with an intent to infringe the patent-right, and deprive the owner of the lawful rewards of his discovery.<sup>21</sup>

Unfortunately, the decision did not clearly define what constituted “lawful rewards,” although it seemed to suggest that a patentee’s monopoly does not extend to cover noncommercial experiments and validity-testing experiments. In 1852, Justice Curtis explained the basis for this distinction in *Byam v. Bullard*,<sup>22</sup> where he noted that scientific research and competitive evaluation do not cause injury to the exclusive patent right and are not performed “with [an] intent to deprive the patentees of some lawful profit.”<sup>23</sup>

Numerous cases were decided between 1852 and 1950 that explored the limits of the “intent to deprive . . . of some lawful profit” standard. Commentators differ regarding the nature of the standard that the courts actually applied, but generally agree that a finding of infringement required the user of the patented technology either to have a commercial intent to make a profit through the use of the patented invention or to derive some actual commercial benefit from the use of the invention (such as sales or reduced costs of production).<sup>24</sup> During this period, only one case,

---

LISH DICTIONARY 180 (Oxford Univ. Press 1971) (defining philosophical experiments as “[p]ertaining to, or used in the study of, natural philosophy, or some branch of physical science”).

19. *Whittemore*, 29 F. Cas. At 1121.

20. 21 F. Cas. 554 (C.C.D. Mass. 1813) (No. 12,391).

21. *Id.* at 555 (citation omitted).

22. 4 F. Cas. 934 (C.C.D. Mass. 1852) (No. 2,262).

23. *Id.* at 935.

24. See, e.g., Ronald D. Hantman, Letter to the Editor, *Re: The Experimental Use Defense*, 87 J. PAT. & TRADEMARK OFF. SOC’Y 348, 348-49 (2005) (noting that the historic case law for the exception required both experimentation and the absence of an intent to use for profit, i.e., where “the infringer makes or attempts to make a monetary profit while infringing the patent”); Ronald D. Hantman, *Experimental Use as an Exception to Patent Infringement*, 67 J. PAT. & TRADEMARK OFF. SOC’Y 617, 625 (1985) [he-

*Ruth v. Stearns-Roger Manufacturing Co.*,<sup>25</sup> addressed scientific research in a university setting.<sup>26</sup> In *Ruth*, the court found that the defendant was not liable for contributing to infringement by supplying replacement parts used at a mining school, given that the patented machines were used only experimentally in a laboratory and subsequently were cut up and changed.<sup>27</sup>

In 1950, Congress proposed legislation that would have explicitly codified the experimental-use exception, excluding from infringement “making or using of a patented invention solely for the purpose of research or experiment” and not for sale.<sup>28</sup> But in 1952, Congress enacted a revised patent law that did not provide an express exception for experimental use. Instead it merely codified in section 271(a) the exclusive rights to make, use, and sell and the existing judicial standards for infringement.<sup>29</sup>

### **B. The 1984 *Bolar* Decision and Legislative Adoption of the Regulatory-Approval Exception**

Between 1952 and 1984, relatively few experimental-use cases were decided, and none involved scientific research.<sup>30</sup> In 1984, however, the Federal Circuit decided *Roche Products Inc. v. Bolar Pharmaceuticals*

---

reinafter Hantman, *Experimental Use as an Exception*] (distinguishing “use for profit” from cases in which “the experimenter neither made money nor tried to make money while infringing the patented invention”); N. Scott Pierce, *A New Day Yesterday: Benefit as the Foundation and Limit of Exclusive Rights in Patent Law*, 6 J. MARSHALL REV. INTELL. PROP. L. 373, 384-412 (2007) (discussing cases finding infringement that focused on the benefit of the invention gained by use, rather than profits obtained, and later cases focusing on commercial intent); Andrew S. Baluch, Note, *Relating the Two Experimental Uses in Patent Law: Inventor’s Negation and Infringer’s Defense*, 87 B.U. L. REV. 213, 250-53 (2007) (discussing factors to distinguish experimental from commercial use derived from experimental-use cases relating to the public use bar of 35 U.S.C. § 102(b)); cf. Richard E. Bee, *Experimental Use as an Act of Patent Infringement*, 39 J. PAT. OFF. SOC’Y 357, 367-68 (1957) (noting the failure of courts to impose reasonably royalty damages for noncommercial uses and arguing that courts used to treat the experimental-use exception very narrowly and found it to apply only when the experiment was performed to gratify a philosophical taste, curiosity, or for amusement).

25. 13 F. Supp. 697 (D. Colo. 1935), *rev’d on other grounds*, 87 F.2d 35 (10th Cir. 1936).

26. *Id.*

27. *Id.* at 703, 713.

28. STAFF OF H. COMM. ON THE JUDICIARY, 81ST CONG., PROPOSED REVISION AND AMENDMENT OF THE PATENT LAWS 59 (Comm. Print 1950) (proposed Section 73).

29. See 35 U.S.C. § 271(a) (2000); S. REP. NO. 82-1979 (1952), *reprinted in* 1952 U.S.C.C.A.N. 2394, 2402 (noting that proposed section 271(a) was merely declaratory of what constitutes infringement).

30. See Hantman, *Experimental Use as an Exception*, *supra* note 24, at 630-39 (discussing the cases).

Co.<sup>31</sup> In *Bolar*, the court held that the experimental-use exception did not apply to scientific tests using a patented pharmaceutical compound for the purpose of obtaining generic product marketing approval from the Food and Drug Administration (FDA).<sup>32</sup> Specifically, the court construed the experimental-use exception to be narrow (limited to “amusement, to satisfy idle curiosity, or for strictly philosophical inquiry”) and held that the defendant’s tests were “solely for business reasons.”<sup>33</sup>

Congress responded to *Bolar* by codifying a separate regulatory-approval exception to patent infringement as part of broader legislation balancing the rights of pioneering and generic pharmaceutical manufacturers.<sup>34</sup> The principal concerns expressed by Congress when adopting this exception were that *Bolar* had been wrongly decided and that patent holders should not be able to dominate research and development during the patent term in a manner that would result in the improper effective extension of the right to exclude beyond the patent term (due to the need to obtain regulatory approval).<sup>35</sup> Specifically, Congress created new section 271(e)(1), which excepted from infringement any making, using, or selling of a patented invention “solely for uses reasonably related to the development and submission of information under a Federal law which regulates . . . drugs.”<sup>36</sup>

In section 271(e)(1), Congress codified broad categorical language that implicitly rejected the narrow *Bolar* construction of the experimental-use exception in the particular context of human drug development.<sup>37</sup> Section 271(e)(1) protects experimenters and their suppliers by excepting from infringement sales for the specified experimental uses. Further, and relevant to the research tool question, the language of section 271(e)(1) encompasses experiments performed by commercial entities with the intent

---

31. 733 F.2d 858 (Fed. Cir. 1984).

32. *Id.* at 862-63.

33. *Id.* at 863.

34. See 35 U.S.C. §§ 155, 155A, 156, 271(e) (2000 & Supp. II 2002).

35. See, e.g., H.R. REP. NO. 98-857, pt. 2, at 60-61 (1984); H.R. REP. NO. 98-857, pt. 1, at 46 (1984).

36. Drug Price Competition and Patent Term Restoration Act of 1984, Pub. L. No. 98-417, § 202, 98 Stat. 1585, 1603 (1984) (codified as amended at 35 U.S.C. § 271(e)(1)). Congress later extended the exception to offers to sell and imports and to approval of veterinary biological products. See Generic Animal Drug and Patent Term Restoration Act, Pub. L. No. 100-670, § 201, 102 Stat. 3971, 3988 (1988) (codified as amended at 35 U.S.C. §§ 156, 271 (2000 & Supp. III 2003); Uruguay Round Agreements Act, Pub. L. No. 103-465, § 533, 108 Stat. 4809, 4988 (1994) (codified as amended in scattered sections of 35 U.S.C.).

37. See *Eli Lilly & Co. v. Medtronic, Inc.*, 872 F.2d 402, 406 (Fed. Cir. 1989) (citing 35 U.S.C. § 271(a), (e)(1)).

subsequently to market products. Moreover, section 271(e)(1) does not on its face distinguish among types of patented inventions or among their roles in regard to experiments designed to obtain regulatory approval. Thus, not only does the exception cover pharmaceuticals being tested for regulatory approval, but it might also arguably be interpreted to cover the use of a patented syringe to draw blood in order to perform that testing. In *Eli Lilly & Co. v. Medtronic, Inc.*,<sup>38</sup> the Supreme Court subsequently interpreted section 271(e)(1) to apply not only to patented inventions used in human drug development, but also to inventions used in developing information for all products requiring pre-market approval by the FDA and subject to the patent term extension provisions of the U.S. patent law.<sup>39</sup>

### C. Subsequent Federal Circuit Interpretations Narrowly Construing the Experimental-Use Exception

Since Congress revised section 271 in 1984, the Federal Circuit has narrowly construed the experimental-use exception. In 2000, in *Embrex, Inc. v. Service Engineering Corp.*,<sup>40</sup> the Federal Circuit reiterated language from *Bolar* that the experimental-use exception does not apply when the experiments have “definite, cognizable, and not insubstantial commercial purposes.”<sup>41</sup> The court upheld a jury verdict of infringement of a patent for a method of injecting eggs based on injection tests performed by scientists, who were employed by a company that unsuccessfully sought to demonstrate a commercial vaccination machine for use as an alternative to the patented method. Specifically, the court held that the tests did not qualify as *de minimis* infringement or as experimental use, given that the tests were performed “expressly for commercial purposes.”<sup>42</sup>

In 2002, in *Madey v. Duke University*,<sup>43</sup> the Federal Circuit held for the first time that the experimental-use exception may not apply to university-based scientific research. The District Court had granted summary judgment of noninfringement to Duke for constructing and using (in ways that allegedly were not authorized under a federal government contract, as Duke would have no liability if the use was so authorized)<sup>44</sup> certain free electron lasers and microwave guns that embodied the claims of two pa-

---

38. 496 U.S. 661 (1990).

39. *See id.* at 669-78; 35 U.S.C. § 156 (2000 & Supp. II 2002).

40. 216 F.3d 1343 (Fed. Cir. 2000).

41. *Id.* at 1349 (quoting *Roche Prods. Inc. v. Bolar Pharms. Co.*, 733 F.2d 858, 863 (Fed. Cir. 1984)).

42. *Id.*

43. 307 F.3d 1351 (Fed. Cir. 2002).

44. *See* 28 U.S.C. § 1498 (2000).

tents.<sup>45</sup> The Federal Circuit reversed, holding that its precedents obligated it to recognize a “judicially created experimental use defense, however, in a very limited form.”<sup>46</sup> That exception “does not immunize use that is in any way commercial in nature . . . [or] that is in keeping with the alleged infringer’s legitimate business, regardless of commercial implications.”<sup>47</sup>

With regard to universities, the court in *Madey* noted that scientific research “projects unmistakably further the institution’s legitimate business objectives, including educating and enlightening students and faculty participating in these projects. These projects also serve, for example, to increase the status of the institution and lure lucrative research grants, students, and faculty.”<sup>48</sup> The court thus remanded for further evaluation of “the legitimate business Duke is involved in and whether or not the use was solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry.”<sup>49</sup> On remand, the District Court found that Duke had presented no evidence to suggest that its experiments were not “in keeping with its legitimate business as an educational institution” and thus denied Duke’s motion for summary judgment but left the issue open for proof at trial.<sup>50</sup>

Few reported district court cases have addressed the experimental-use exception since the *Madey* decision. Those cases that do so either have reiterated the narrow scope of the exception or have simply referred to *Madey* as binding precedent.<sup>51</sup> In *Integra Lifesciences I Ltd. v. Merck KGaA*,<sup>52</sup> the Federal Circuit suggested in dicta that the experimental-use exception is not only narrow but also based on the concept of *de minimis* damages rather than the lack of infringement.<sup>53</sup> In contrast, a dissenting opinion suggested that the exception is significantly broader—i.e., that the

---

45. *Madey*, 307 F.3d at 1352.

46. *Id.* at 1360; *see also id.* at 1361 (recognizing the exception exists “in the very narrow form articulated by this court” in *Embrex and Bolar*).

47. *Id.* at 1362.

48. *Id.*

49. *Id.* at 1363.

50. *Madey v. Duke Univ.*, 336 F. Supp. 2d 583, 591-92 (M.D.N.C. 2004).

51. *See, e.g., Eli Lilly & Co. v. Emisphere Techs., Inc.*, 408 F. Supp. 2d 668, 678 n.2 (S.D. Ind. 2006) (noting pleading of experimental-use defense to infringement counterclaim, and citing to *Madey* for a discussion of the doctrine, but refusing to address the issue as premature); *Third Wave Techs., Inc. v. Stratagene Corp.*, 381 F. Supp. 2d 891, 911-12 (W.D. Wis. 2005) (rejecting arguments that testing of products for cleaving nucleic acids that might infringe patented cleaving methods allegedly to obtain FDA approval for the products would not qualify as experimental use, given the narrow scope of the exception in *Madey* and the commercial motivation to market the products).

52. 331 F.3d 860 (Fed. Cir. 2003).

53. *Id.* at 863-64 n.2.

“subject matter of patents may be studied in order to understand it, or to improve upon it, or to find a new use for it, or to modify or ‘design around’ it.”<sup>54</sup>

#### **D. Proposals for Legislation to Codify a Broader Experimental-Use Exception**

Since *Bolar*, Congress has, on a few occasions, proposed legislation to codify a broader experimental-use exception. But these efforts have not resulted in adoption of a change to the law. For example, in 1990, Congress introduced a bill that would have excepted from infringement any making and use for “research or experimentation purposes,” unless the primary purpose of the patented invention was for research (i.e., intended for use as a research tool), in which case it would not be an act of infringement to study the invention or use it to develop new inventions outside the scope of the patent.<sup>55</sup> Similarly, in 2002, Congress introduced a bill that would have excepted from infringement any patented genetic sequences “for purposes of research,” which would not apply to commercial manufactures and sales.<sup>56</sup> In contrast, in 2007, Congress introduced a bill that would prospectively ban the patenting of any “nucleotide sequence, or its functions or correlations, or the naturally occurring products it specifies.”<sup>57</sup> The bill thus would preclude a particular category of patents (not only gene patents but all patents on polynucleotides), which may be used as research tools. However, the biotechnology industry and others have expressed significant opposition to the bill,<sup>58</sup> and it currently appears unlikely to be enacted into law.

In 2002, the Federal Trade Commission (FTC) heard testimony on the effects of research tool patents on third-party research and innovation.<sup>59</sup>

---

54. *Id.* at 875 (Newman, J., concurring in part and dissenting in part).

55. Patent Competitiveness and Technological Innovation Act of 1990, H.R. 5598, 101st Cong. § 402 (1990) (proposed 35 U.S.C. § 271(j)).

56. Genomic Research and Diagnostic Accessibility Act of 2002, H.R. 3967, 107th Cong. § 2 (2002) (proposed 35 U.S.C. § 271(j)(1)).

57. Genomic Research and Accessibility Act, H.R. 977, 110th Cong. § 2 (2007) (proposed 35 U.S.C. § 106).

58. *See, e.g., Stifling or Stimulating—The Role of Gene Patents in Research and Genetic Testing: Hearing Before the Comm. on The Judiciary Subcomm. on Courts, the Internet and Intellectual Property*, 110th Cong., at 73-77 (2007) [hereinafter *Hearings*] (statements of Jeffrey Kushan on behalf of the Biotechnology Industry Organization and E. Jonathan Soderstrom, Managing Director, Office of Cooperative Research, Yale University).

59. FTC, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY, executive summary, at 3-4, ch. 3, at 18-20, ch. 4, at 34-36

The hearing, which addressed many issues, “involved more than 300 panelists, including business representatives from large and small firms, and the independent inventor community; leading patent and antitrust organizations; leading antitrust and patent practitioners; and leading scholars in economics and antitrust and patent law.”<sup>60</sup> Various panelists voiced general approval for codifying a broader experimental-use exemption that would apply to research directed at understanding if and how a patented invention works (recall the statement in *Whittemore* regarding sufficiency of the machine to produce its desired effects).<sup>61</sup> They were more divided on the question of whether an exception should apply to research directed at improvement or follow-on innovation resulting from use of patented research tools. They generally rejected the idea of providing an exemption for use of a research tool to develop another product.<sup>62</sup> A report based on the hearings concluded that developers of research tools “need an income stream from those who use their inventions” and that the “hearing record provides no basis for exempting such tools from patent protection.”<sup>63</sup>

Proposals to explicitly codify an experimental-use exception have also come from the private sector. A 2004 report sponsored by the National Academy of Sciences (NAS) recommended codification of an experimental-use exception in light of *Madey*, given that “there should be some level of protection for noncommercial uses of patented inventions.”<sup>64</sup> The report also recommended taking various administrative actions to ensure access, given that legislative enactment might not occur.<sup>65</sup> Some members of the committee consulted in the preparation of the report expressed the opinion that any codified experimental-use exception should be conditioned upon the researcher agreeing to refrain from patenting the results of the protected research, “the results of the research not undermining a patentee’s commercial markets, a covenant not to use the research results for commercial purposes, and provision for terminating the exemption if the protected research yields patents that are asserted against another party lacking the exemption.”<sup>66</sup>

---

(2003), available at <http://www.ftc.gov/os/2003/10/innovationrpt.pdf> [hereinafter FTC, TO PROMOTE INNOVATION].

60. *Id.*

61. *Id.* ch. 4, at 34-36.

62. *Id.* ch. 4, at 36.

63. *Id.*

64. See NAT’L RESEARCH COUNCIL, A PATENT SYSTEM FOR THE 21ST CENTURY 82, 109 (Stephen A. Merrill et al. eds., 2004) [hereinafter PATENT SYSTEM].

65. See *id.* at 108-17.

66. See *id.* at 115.

Later in 2004, the American Intellectual Property Lawyer's Association (AIPLA) endorsed the NAS recommendation and proposed language for a broader codified experimental-use exception.<sup>67</sup> Specifically, the AIPLA proposal would have excepted from infringement the acts of:

- (1) evaluating the validity of the patent and the scope of protection afforded under the patent;
- (2) understanding features, properties, inherent characteristics or advantages of the patented subject matter;
- (3) finding other methods of making or using the patented subject matter; and
- (4) finding alternatives to the patented subject matter, improvements thereto or substitutes therefor.<sup>68</sup>

In 2006 the NAS published a report focused more specifically on the impact of patents on genomic and proteomic research.<sup>69</sup> This report recommended:

Congress should consider exempting research "on" inventions from patent infringement liability. The exemption should state that making or using a patented invention should not be considered infringement if done to discern or to discover:

- a. the validity of the patent and scope of afforded protection;
- b. the features, properties, or inherent characteristics or advantages of the invention;
- c. novel methods of making or using the patented invention; or
- d. novel alternatives, improvements, or substitutes.

Further making or using the invention in activities incidental to preparation for commercialization of noninfringing alternatives also should be considered noninfringing. Nevertheless, a statutory research exemption should be limited to these circumstances and not be unbounded. In particular, it should not extend to unauthorized use of research tools for their intended purpose, in other words, to research "with" patented inventions. According-

---

67. See AM. INTELLECTUAL PROP. LAW ASS'N, AIPLA RESPONSE TO THE NATIONAL ACADEMIES REPORT ENTITLED "A Patent System for the 21st Century" 23-27 (2004), available at [http://www.aipla.org/Content/ContentGroups/Issues\\_and\\_Advocacy/Comments2/Patent\\_and\\_Trademark\\_Office/2004/NAS092304.pdf](http://www.aipla.org/Content/ContentGroups/Issues_and_Advocacy/Comments2/Patent_and_Trademark_Office/2004/NAS092304.pdf) [hereinafter AIPLA RESPONSE].

68. *Id.* at 25.

69. COMM. ON INTELLECTUAL PROPERTY RIGHTS IN GENOMIC & PROTEIN RESEARCH & INNOVATION, NAT'L RESEARCH COUNCIL, REAPING THE BENEFITS OF GENOMIC AND PROTEOMIC RESEARCH: INTELLECTUAL PROPERTY RIGHTS, INNOVATION, AND PUBLIC HEALTH (2006) [hereinafter REAPING THE BENEFITS].

ly, our recommendation would not address the circumstances of the *Madey* case, which clearly entailed research “with” the patented laser; but it would shield some types of biomedical research involving patented subject matter.<sup>70</sup>

Scholars also have debated for many years the need for the U.S. to implement an expanded experimental-use exception.<sup>71</sup> Some have advocated the creation of broad exemptions for use of patented technologies by university and nonprofit researchers,<sup>72</sup> while others have pointed out a host of practical difficulties that might arise if such plans were implemented.<sup>73</sup> Some worry that a broad experimental-use exception would remove incentives for the development of new research tools.<sup>74</sup> Others question whether patent protection is needed to develop research tools, although often recognizing that patents can play a useful role when investment is needed to make the technology practically available.<sup>75</sup> Some commentators have proposed application of the doctrine of fair use to promote access to re-

---

70. *Id.* at 145.

71. *See, e.g.*, HAROLD C. WEGNER, *PATENT LAW IN BIOTECHNOLOGY, CHEMICALS & PHARMACEUTICALS* 460 *passim* (2d ed. 1994); Lauren C. Bruzzone, *The Research Exception: A Proposal*, 21 *AIPLA Q.J.* 52 *passim* (1993); Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 *U. CHI. L. REV.* 1017 *passim* (1989); Irving N. Feit, *Biotechnology Research and the Experimental Use Exception to Patent Infringement*, 71 *J. PAT. & TRADEMARK OFF. SOC'Y* 819, 832 (1989); Steven J. Grossman, *Experimental Use or Fair Use as a Defense to Patent Infringement*, 30 *IDEA* 243, 247 (1990); Ned A. Israelsen, *Making, Using, and Selling Without Infringing: An Examination of 35 U.S.C. Section 271(e) and the Experimental Use Exception to Patent Infringement*, 16 *AIPLA Q.J.* 457, 458, 472, 474 (1988-1989); Suzanne T. Michel, Comment, *The Experimental Use Exception to Infringement Applied to Federally Funded Inventions*, 7 *HIGH TECH. L.J.* 369, 376, 389 (1992); Patricia M. Thayer & Richard A. De Liberty, *The Research Exception to Patent Infringement: The Time Has Come for Legislation*, 4 *J. BIOLAW & BUS.* 15 *passim* (2000); Jordan P. Karp, Note, *Experimental Use as Patent Infringement: The Impropriety of a Broad Exception*, 100 *YALE L.J.* 2169, 2170 (1991).

72. *See, e.g.*, Eyal H. Barash, Comment, *Experimental Uses, Patents, and Scientific Progress*, 91 *NW. U. L. REV.* 667, 699-700 (1997); Kevin Sandstrom, Note, *How Much Do We Value Research and Development?: Broadening the Experimental Use Exemption to Patent Infringement in Light of Integra Lifesciences I, Ltd. v. Merck KGaA*, 331 *F.3d* 860 (Fed. Cir. 2003), 30 *WM. MITCHELL L. REV.* 1059, 1111 (2004).

73. *See, e.g.*, Elizabeth A. Rowe, *The Experimental Use Exception to Patent Infringement: Do Universities Deserve Special Treatment?*, 59 *ME. L. REV.* 283, 308-10 (2007).

74. *See, e.g.*, Mueller, *supra* note 8, at 39-40.

75. *See, e.g.*, Michael S. Mireles, *States as Innovation System Laboratories: California, Patents, and Stem Cell Technology*, 28 *CARDOZO L. REV.* 1133, 1152 (2006).

search tools,<sup>76</sup> but others have criticized this approach.<sup>77</sup> A number of scholars have proposed hybrid systems combining limited experimental-use exceptions with compulsory licensing or other alternative approaches, some of which are discussed in more detail below in Part VII.<sup>78</sup>

**E. *Merck v. Integra*: Initial Federal Circuit and Subsequent Supreme Court Interpretations of the Regulatory-Approval Exception**

In *Integra Lifesciences I Ltd. v. Merck KGaA*,<sup>79</sup> the Federal Circuit assessed the boundary between the experimental-use exception and the regulatory-approval exception of section 271(e)(1) in the context of the drug development and regulatory-approval process. The district court held that the use of patented materials in early-stage experiments assessing the materials' potential applicability to cancer treatment<sup>80</sup> qualified for the experimental-use exception.<sup>81</sup> In contrast, the district court held that later experiments conducted before performing human clinical trials (some of which, as discussed below, may have been research tool uses) did not qualify for the regulatory-approval exception.<sup>82</sup> *Integra* did not appeal the experimental-use exception holding for the early stage experiments, even though they likely had been performed with ultimate commercial pharmaceutical applications in mind. *Merck* appealed the holding that section 271(e)(1) did not apply to the subsequent preclinical experiments, and the Federal Circuit affirmed.<sup>83</sup> Specifically, the Federal Circuit noted that these experiments were not "solely for uses reasonably related to the development and submission of information," because they did not "reasonably relate to the development and submission of information for FDA's safety and effectiveness approval processes."<sup>84</sup> In contrast, the dissent would have held that the regulatory-approval and experimental-use exceptions should be read co-extensively to avoid any gap in coverage, and thus

---

76. See, e.g., Maureen A. O'Rourke, *Toward a Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177, 1249-50 (2000).

77. See, e.g., Rowe, *supra* note 73, at 309-10.

78. See, e.g., Katherine J. Strandburg, *What does the Public Get? Experimental Use and the Patent Bargain*, 2004 WIS. L. REV. 81, 143-144.

79. 331 F.3d 860 (Fed. Cir. 2003).

80. The experiments assessed the patented materials' potential to block certain receptors and thereby inhibit blood vessel proliferation. *Id.* at 863.

81. See *Merck KGaA v. Integra LifeSciences I, Ltd.*, 545 U.S. 193, 200 (2005),

82. See *id.* at 201.

83. See *id.*

84. *Integra LifeSciences I, Ltd.*, 331 F.3d at 866 (quoting 35 U.S.C. § 271(e)(1)).

that “the statutory immunity of § 271(e) takes effect wherever the research exemption ends.”<sup>85</sup>

In 2005, the Supreme Court in *Merck, KGaA v. Integra LifeSciences I Ltd.*<sup>86</sup> reversed the Federal Circuit’s narrow construction of the regulatory-approval exception of section 271(e)(1), but did not resolve whether a gap exists between the two exceptions and expressly refused to address whether section 271(e)(1) applies to patented inventions used as research tools.<sup>87</sup> The Court held that the exception was not limited to tests that generate safety and effectiveness data; rather, it included any tests (including basic research on biological mechanisms) that might generate data submitted to the FDA:

At least where a drugmaker has a reasonable basis for believing that a patented compound may work, through a particular biological process, to produce a particular physiological effect, and uses the compound in research that, if successful, would be appropriate to include in a submission to the FDA, that use is “reasonably related” to the “development and submission of information under . . . Federal law.”<sup>88</sup>

Following the *Merck* Supreme Court decision, commentators have noted that research tools were involved in at least some of the allegedly infringing experiments at issue on appeal, notwithstanding the parties’ arguments to the contrary. For example, the patented materials at issue were used as positive controls to measure the effectiveness of other materials.<sup>89</sup> Use as an experimental control is a research tool use because the materials are not the object of the experiment but rather supply the means to conduct it. Commentators also have raised concerns that application of section 271(e)(1) to research tool inventions would eviscerate patent rights and incentives.<sup>90</sup> To avoid this result, they have argued that the term “pa-

---

85. *Id.* at 875-76 (Newman, J., concurring in part and dissenting in part).

86. 545 U.S. 193 (2005).

87. *Id.* at 205 n.7.

88. *Id.* at 207 (quoting 35 U.S.C. § 271(e)(1)).

89. See, e.g., Paul Wiegel, *Was the FDA Exemption to Patent Infringement, 35 U.S.C. § 271(e)(1), Intended to Exempt a Pharmaceutical Manufacturer’s Activities in the Development of New Drugs?*, 2007 B.C. INTELL. PROP. & TECH. F. 112901 (2007); Benjamin G. Jackson, Note, *Merck v. Integra: Bailing Water Without Plugging the Hole*, 20 BYU J. PUB. L. 579, 596 (2006) (noting the court’s statement that “Scripps used the RGD peptide in . . . tests as ‘positive controls’ against which to measure the efficacy of the mimetics”).

90. See, e.g., Daniel J. Ford, *Merck v. Integra: Implications for the Common Law and Statutory Exemptions*, 7 LOY. L. & TECH. ANN. 123 (2007); Vihar R. Patel, *Are Patented Research Tools Still Valuable? Use, Intent, and a Rebuttable Presumption: A Pro-*

tented inventions” within section 271(e)(1) should be interpreted to be limited to patented drug and medical device inventions that are subject to regulatory approval and term extension under section 156, which was the focus of the broader legislation enacting section 271(e)(1).<sup>91</sup>

In contrast, other commentators have suggested that the effects of the *Merck* decision on research tool inventions will be minimal, because “the sanctioned research is *into*, not *using*, patented technology and patents have a smaller impact on research tools and instruments than on drug development.”<sup>92</sup> Other commentators have suggested expanding the exception further to minimize incentives for drug companies to “outsourc[e] their early stage research from the United States,”<sup>93</sup> to jurisdictions where broader experimental-use exceptions exist or where patent rights in research tool inventions do not exist.

#### F. Cases Interpreting the Regulatory-Approval Exception Since the 2005 Supreme Court Decision in *Merck*

Cases since *Merck* have for the most part followed the trend of construing the regulatory-approval exception of section 271(e)(1) broadly, with some notable exceptions discussed later in this section. For example, on remand from the Supreme Court the Federal Circuit held in *Integra Lifesciences I Ltd. v. Merck KGaA*<sup>94</sup> that section 271(e)(1) applied to experiments with patented compounds that at the time were candidates for, but were not ultimately the subject of, the regulatory-approval application.<sup>95</sup> The experiments developed information “after the biological mechanism and physiological effect of a candidate drug have been recognized, such that if the research is successful it would appropriately be included in a

---

*posed Modification for Analyzing the Exemption from Patent Infringement Under 35 U.S.C. 271(e)(i)*, 47 IDEA 407 (2007); Tara Stuart, Comment, *Has the Supreme Court Incorrectly Expanded § 271(e)(1) to Risk a Regulatory Taking?*, 5 J. MARSHALL REV. INTELL. PROP. L. 216 (2006); Anna McMinn, Note, *Judicial Interpretation of 35 USC § 271(e)(1): An Improper Expansion Beyond the Legislative Intent*, 16 ALB. L.J. SCI. & TECH. 195 (2006).

91. See 35 U.S.C. § 156 (2000 & Supp. II 2002); Pierce, *supra* note 24. A patent holder has also made this argument in litigation. See, e.g., Oral Arguments in *Proveris Sci. Corp. v. Innovasystems, Inc.*, 536 F.3d 1256 (Fed. Cir. 2008) (No. 2007-1428), available at <http://oralarguments.cafo.uscourts.gov/mp3/2007-1428.mp3>.

92. Daniel A. Lev, *A Realist Approach to Merck KGaA v. Integra*, 5 NW. J. TECH. & INTELL. PROP. 135, 150 (2006) (emphasis added).

93. Katherine A. Helm, Note, *Outsourcing the Fire of Genius: The Effects of Patent Infringement Jurisprudence on Pharmaceutical Drug Development*, 17 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 153 (2006).

94. 496 F.3d 1334 (Fed. Cir. 2007).

95. *Id.* at 1347.

submission to the FDA.”<sup>96</sup> Significantly, the court held that whether the experiments were “reasonably related” to submission “does not depend on the success or failure of the experimentation or actual submission of the experimental results.”<sup>97</sup>

The Federal Circuit also held in 2008, in *Amgen, Inc. v. Roche Holding Ltd.*,<sup>98</sup> that section 271(e)(1) applies to patent infringement actions brought under section 337 of the Tariff Act of 1930.<sup>99</sup> Section 337 prohibits importation and sale after importation of articles that infringe a valid U.S. patent or that are produced by a process covered by a valid U.S. patent.<sup>100</sup> Amgen had argued that the regulatory-approval exception of section 271(e)(1) did not apply to section 337 actions for products of patented processes, given that the language of section 271(e)(1) excepts only a “patented invention” and that Congress had stated an intent to preserve section 337 remedies when it separately provided infringement liability for importing products of patented processes.<sup>101</sup> The Federal Circuit upheld the Commission’s interpretation that section 271(e)(1) applies to section 337 actions, based on a “broadly stated Congressional policy” in the legislative history of section 271 that importing products made by patented processes for regulatory approval should not constitute patent infringement and on Supreme Court statements that in enacting section 271(e)(1) Congress “intended that the immunity of regulatory activity not be inhibited.”<sup>102</sup> However, the Federal Circuit remanded the case to the Commission to determine whether to provide a remedy against potential infringement that might occur after a regulatory-approval decision, when section 271(e) would no longer be applicable.<sup>103</sup>

Several district court decisions have also interpreted 271(e)(1) broadly. For example, in *Classen Immunotherapies, Inc. v. Biogen IDEC*<sup>104</sup> a district court held that 271(e)(1) applies to activities aimed at generating data for post-approval submissions to FDA.<sup>105</sup> The alleged infringing activities involved examining risks associated with the administration of vaccines

---

96. *Id.* at 1339; *see id.* at 1340.

97. *Id.* at 1341.

98. 519 F.3d 1343 (Fed. Cir. 2008).

99. *Id.* at 1345.

100. 19 U.S.C. § 1337(a)(1)(B)(i)-(ii) (2000).

101. 35 U.S.C. § 271(e)(1) (2000); *see* 35 U.S.C. § 271(g) (2000); 19 U.S.C. § 2901(b)(10) (2000); *Amgen*, 519 F.3d at 1346-47.

102. *Amgen*, 519 F.3d at 1348 (citing S. Rep. No. 100-83, 48 (1987) and *Merck KGaA v. Integra LifeSciences I, Ltd.*, 545 U.S. 193, 202 (2005)).

103. *Id.* at 1350-53.

104. 381 F. Supp. 2d 452 (D. Md. 2005).

105. *Id.* at 456.

that had already secured FDA marketing approval and submitting the data under FDA regulations.<sup>106</sup> The court dismissed infringement claims against defendants for participating in the studies, given the broad construction of section 271(e)(1) in *Merck*.<sup>107</sup> Specifically, the district court rejected the argument that section 271(e)(1) applied only to data for regulatory decisions made before initial regulatory approval to market products.<sup>108</sup>

Similarly, in *Genentech, Inc. v. Insmid Inc.*,<sup>109</sup> a district court held that section 271(e)(1) applies to experiments conducted in part for commercial reasons unrelated to an FDA submission, so long as “the experiments would produce information that would be given to the FDA in order to get FDA approval.”<sup>110</sup> The court granted summary judgment to a defendant that had supplied patented insulin-like growth factor for use in the arguably commercial experiments.<sup>111</sup>

Two decisions have found section 271(e)(1) applicable to activities arguably involving the use of patented technology as a research tool. In *Classen Immunotherapies, Inc. v. King Pharmaceuticals, Inc.*,<sup>112</sup> the patents at issue addressed methods of identifying and commercializing new uses of existing drugs, and infringement was alleged based on the submission of data generated using the patented methods to the FDA.<sup>113</sup> The district court held the experiments were reasonably related to the submission of information to the FDA and found “extension of the safe harbor to cover the use of these tools warranted by the language of *Merck* and a plain reading of the statute.”<sup>114</sup>

*Integra Lifesciences I Ltd. v. Merck KGaA*<sup>115</sup> also arguably applied 271(e)(1) to the use of research tools.<sup>116</sup> Noting that the parties agreed that the patented compounds were not used as research tools, the majority opinion did not address the issue of whether section 271(e)(1) applies to patented inventions used as research tools.<sup>117</sup> However, the dissent argued that the decision did apply to research tools as some of the patents claimed

---

106. *Id.* at 454.

107. *Id.* at 455-56.

108. *Id.*

109. 436 F. Supp. 2d 1080 (N.D. Cal. 2006).

110. *Id.* at 1095.

111. *Id.* at 1094-95.

112. 466 F. Supp. 2d 621 (D. Md. 2006).

113. *Id.* at 623-24.

114. *Id.* at 625 n.2.

115. 496 F.3d 1334 (Fed. Cir. 2007).

116. *Id.* at 1348.

117. *Id.* at 1347.

methods that could not have been potential regulatory-approval drug candidates.<sup>118</sup> The dissent thus argued that the holding effectively “eliminate[s] protection for research tool inventions.”<sup>119</sup> Commentators have also concluded that research tools were involved in at least some of the allegedly infringing experiments at issue on appeal, notwithstanding the parties’ arguments to the contrary.<sup>120</sup>

Several court decisions have found substantive or procedural limitations to the application of section 271(e)(1). For example, in *Amgen, Inc. v. F. Hoffman-LaRoche Ltd.*,<sup>121</sup> a district court held that section 271(e)(1) is an affirmative defense, rather than “part of the statutory definition of infringement that [the plaintiff] must establish.”<sup>122</sup> Accordingly, the district court rejected a motion to dismiss for failure to state a claim, given that the plaintiff had sufficiently alleged infringement without pleading specific acts of infringement that fell outside the scope of section 271(e)(1).<sup>123</sup> Further, the complaint alleged importation of an allegedly infringing patented drug, which was sufficient given that the district court could not conclude as a matter of law that importation was solely for uses reasonably related to submitting information for regulatory approval.<sup>124</sup>

In *Forest Laboratories, Inc. v. Ivax Pharmaceuticals, Inc.*, the Federal Circuit upheld a prospective but limited injunction against a foreign producer of patented drug products that was supplying production information and would supply products for experiments within the scope of the regulatory-approval exception of section 271(e)(1).<sup>125</sup> The injunction prohibited the domestic experimenter from commercial exploitation following FDA approval and during the life of the patent, and the court held it was appropriate to include the foreign producer as such supply would induce infringement under section 271(b) following such approval.<sup>126</sup>

Finally, in *Proveris Scientific Corp. v. Innovasystems, Inc.*,<sup>127</sup> the Federal Circuit held that the sale of a device intended solely for use in generating data for submission to FDA, and only sold to pharmaceutical com-

---

118. *Id.* at 1350-51 (Rader, J., concurring in part and dissenting in part).

119. *Id.* at 1348.

120. See Jackson, *supra* note 89, at 595-98; Wiegand, *supra* note 89.

121. 456 F. Supp. 2d 267 (D. Mass. 2006).

122. *Id.* at 273 (quoting *Amgen, Inc. v. Hoechst Marion Roussel, Inc.*, 3 F. Supp. 2d 104, 109 (D. Mass. 1998)).

123. *Id.* at 274.

124. *Id.*

125. *Forest Labs., Inc. v. Ivax Pharms., Inc.*, 501 F.3d 1263 (Fed. Cir. 2007).

126. *Id.* at 1272.

127. 536 F.3d 1256 (Fed. Cir. 2008).

panies or the FDA, does not fall within the section 271(e)(1) safe harbor.<sup>128</sup> In reaching its decision, the court ruled that such a device is not a “patented invention” within the meaning of 271(e)(1) because it “is not subject to FDA premarket approval, and therefore faces no regulatory barriers to market entry upon patent expiration.”<sup>129</sup> For this reason, the court did not “think Congress could have intended that the safe harbor of section 271(e)(1) apply to it.”<sup>130</sup>

In summary, the Federal Circuit has narrowed and clarified the scope of the experimental-use exception, and under that scope almost no scientific research, including university-based, nonprofit basic research, will qualify for the exception. Such research is likely to be performed with commercial intent or to further the legitimate business of the experimenter’s business. In contrast, the Supreme Court has expanded the scope of the regulatory-approval exception of section 271(e)(1), which will apply to a broad range of experiments that may generate data that regulators would be interested in reviewing. However, the limits of the regulatory-approval exception remain unclear. In particular, the courts have yet to draw clear lines for determining: (1) in regard to scientific experimentation not excepted from infringement under the experimental-use exception, when protected regulatory-approval activities begin; (2) whether and under what circumstances patented research tools may be subject to the regulatory-approval exception because they are used in a manner reasonably relating to development and submission of information to the FDA; and (3) when such patented research tools should be considered made, used, or sold solely for regulatory-approval purposes. However, unless the Federal Circuit’s approach in the *Proveris* case is later revised or reversed, patented research tools will not be subject to the regulatory-approval exception unless they are themselves potentially subject to FDA premarket approval.

#### IV. RECENT STUDIES OF SCIENTIFIC RESEARCHER AND PATENT HOLDER PRACTICES

This Part discusses the existing studies of the practices of scientific researchers and patent holders regarding the researchers’ acquisition of patented technologies used as research tools, and liability for such research uses. The empirical analysis is important, not only to understand the effects of the legal developments described above, but also to discern potential trends in regard to changing social practices or the need for further

---

128. *Id.* at 1266.

129. *Id.* at 1265.

130. *Id.*

changes to the legal rules. Unfortunately, the factual results of the surveys are subject to dispute regarding what they suggest for continuation of or changes to existing patent system policies.

A number of studies have been conducted to evaluate the effects of *Madey*, particularly regarding whether patents on inventions intended to function as research tools have impeded or delayed basic scientific research. These concerns reflect earlier theoretical work regarding the potential for development of an “anticommons,” or patent thicket requiring licensing of multiple patented inputs, that would result in higher costs, delay, and potentially abandonment of important scientific research (particularly in regard to biomedical and gene-based research).<sup>131</sup> These concerns also reflect the fact that genetic inventions are fundamental, and thus patents on genetic sequences cannot be designed around.<sup>132</sup> The results of these studies demonstrate that relatively few serious problems have resulted from the expanded legal potential for patent liability for use of patented research tools, particularly in regard to use by academic researchers, but suggest that these problems are growing. This may be because patent holders have not aggressively asserted their patents against scientific researchers and because such researchers have continued to act in ways that, in light of the *Madey* decision, infringe those patents. The studies also demonstrate that there has been an increase in warning letters and internal efforts at universities to discourage patent infringement, but neither have yet had significant effects on researcher behavior. Stated differently, there is a significant gap between the law on the books and the practices to which the law applies, and the stability of the current situation remains a subject of significant concern.

---

131. See, e.g., Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 *SCIENCE* 698 *passim* (1998); Kyle Jensen & Fiona Murray, *Intellectual Property Landscape of the Human Genome*, 310 *SCIENCE* 239, 239-40 (2005); Carl Shapiro, *Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting*, in 1 *INNOVATION POLICY AND THE ECONOMY* 119, 120 (Adam Jaffe et al. eds., 2001); cf. Lori Andrews et al., *When Patents Threaten Science*, 314 *SCIENCE* 1395, 1395-96 (2006). But see TED BUCKLEY, *BIOTECHNOLOGY INDUS. ORG. [BIO], THE MYTH OF THE ANTICOMMONS* (2007) available at <http://www.bio.org/ip/domestic/TheMythoftheAnticommons.pdf>; Timothy Caulfield et al., *Evidence and Anecdotes: An Analysis of Human Gene Patenting Controversies*, 24 *NATURE BIOTECHNOLOGY* 1091, 1091-94 (2006).

132. See, e.g., John H. Barton, *Emerging Patent Issues in Genomic Diagnostics*, 24 *NATURE BIOTECHNOLOGY* 939 (2006).

### A. The Walsh, Arora, and Cohen Study (2003)

Around the time of *Madey*, various researchers studied the effects of research tool patents on biomedical innovation, as part of broader research, commissioned by the NAS Board on Science, Technology, and Economic Policy (STEP), leading to proposals for reforming the U.S. patent system.<sup>133</sup> In the Walsh, Arora and Cohen study, the researchers specifically sought to address two questions: (1) “whether an emergent anticommons is in fact impeding the development and commercialization of new drugs, diagnostics, and other therapies”; and (2) “whether restricted access to patents on upstream, foundational discoveries is blocking important follow-on research and innovation.”<sup>134</sup>

More specifically, the researchers “conducted 70 interviews with IP attorneys, business managers, and scientists from 10 pharmaceutical firms and 15 biotechnology firms, as well as university researchers and technology transfer officers from 6 universities, patent lawyers, and government and trade association personnel.”<sup>135</sup> The interviews probed whether proliferation of patents had resulted in failures to license beneficial patented technologies and whether patents on upstream discoveries had impeded subsequent research. The researchers identified the development of “defensive” patenting strategies in genomics (where patents are obtained principally as a method of discouraging litigation rather than for use in protecting the patented innovation),<sup>136</sup> and that “different parties with different agendas” owned research tool patents.<sup>137</sup> Nevertheless, the researchers found that the number of ongoing R&D projects stopped due to patent problems was “small,”<sup>138</sup> finding little evidence that patent-holding entities were refusing to license needed technologies, that the need to license multiple patents was resulting in excessive royalties, or that the increased costs of licensing individual research tool patents were unreasonable (given beliefs that the “productivity gains conferred by the licensed

---

133. See John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in PATENTS IN THE KNOWLEDGE-BASED ECONOMY, *supra* note 10, at 285. The recommendations are published in PATENT SYSTEM, *supra* note 64.

134. Wesley M. Cohen & Stephen A. Merrill, *Introduction to PATENTS IN THE KNOWLEDGE BASED ECONOMY*, *supra* note 10, at 13.

135. John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in PATENTS IN THE KNOWLEDGE-BASED ECONOMY, *supra* note 10, at 292.

136. See *id.* at 295.

137. *Id.* at 296.

138. *Id.* at 303.

research tools were thought to be worth the price”).<sup>139</sup> Patent holders also generally tolerated infringing academic uses of research tools (except for diagnostic tests used in clinical research), as such use could increase the value of the technology and as legal fees, risks of having the patent narrowed or found invalid, and bad publicity from suing universities typically outweighed the potential benefits from such lawsuits.<sup>140</sup>

The research did not find that the growth of patents on fundamental upstream discoveries and more aggressive licensing by nonprofit research institutions, small businesses, and research universities had to that time impeded the development of drugs or other therapies in a significant way. Significantly, “firms and other institutions have developed a number of ‘working solutions’ that limit the effects of the intellectual property complexities that exist,” including “fairly pervasive infringement of patents in the course of laboratory research at the pre-product stage.”<sup>141</sup> Pervasive infringement was “informally rationalized as causing no commercial harm and, in any event [was believed to be] shielded from infringement liability by the court-interpreted ‘research exception.’”<sup>142</sup> However, *Madey* clearly called these common beliefs into question, and “undermine[d] one of the working solutions that has contributed to the progress of biomedical research.”<sup>143</sup>

In contrast, “at least for licensing relationships between universities and small firms, access to relatively upstream discoveries . . . is commonly restricted.”<sup>144</sup> However, it was not clear that such restrictive (typically exclusive) licensing impeded follow-on discovery, given that it may lead to increased motivation for further development of the upstream technology by the licensee.<sup>145</sup> The study noted the potential for *Madey* to “chill” some

---

139. *Id.* at 300-301. *See id.* at 298-302.

140. *See* John P. Walsh, Ashish Arora, and Wesley M. Cohen, *Working Through the Patent Problem*, 299 *SCIENCE* 1021 (2003) [hereinafter Walsh et al., *Working Through*].

141. Wesley M. Cohen & Stephen A. Merrill, *Introduction to PATENTS IN THE KNOWLEDGE BASED ECONOMY*, *supra* note 10, at 13; *see also* John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in *PATENTS IN THE KNOWLEDGE-BASED ECONOMY*, *supra* note 10, at 322-34.

142. Wesley M. Cohen & Stephen A. Merrill, *Introduction to PATENTS IN THE KNOWLEDGE BASED ECONOMY*, *supra* note 10, at 13; *see* John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in *PATENTS IN THE KNOWLEDGE-BASED ECONOMY*, *supra* note 10, at 324-28.

143. Wesley M. Cohen & Stephen A. Merrill, *Introduction to PATENTS IN THE KNOWLEDGE BASED ECONOMY*, *supra* note 10, at 13 n.4.

144. John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in *PATENTS IN THE KNOWLEDGE-BASED ECONOMY*, *supra* note 10, at 309.

145. *See id.* at 309-10.

of the infringing biomedical research occurring in university settings<sup>146</sup> and concluded that:

Through a combination of luck and appropriate institutional response, we appear to have avoided situations where a single firm or organization using its patents has blocked research in one or more broad therapeutic areas. However, the danger remains that progress in a broad research area could be significantly impeded by a patentholder trying to reserve the area exclusively for itself.<sup>147</sup>

Further, the researchers noted significant concerns with increasing secrecy of scientists and with the ability of scientists to share or to obtain access to physical materials needed for research.<sup>148</sup> The process of negotiating material transfer agreements had become significantly longer, resulting in delays of research and in exceptional cases in abandonment of research.<sup>149</sup> Conversely, some university scientists noted that commercial licensing of reagents may result in increasing access given the difficulty of alternative methods of filling demand.<sup>150</sup> The researchers concluded that “to the degree that the patenting of biomedical discoveries may impose additional costs and delays in material transfers, it is partly because the Bayh-Dole Act<sup>151</sup> and related acts have provided universities a vested commercial interest in the disposition of intellectual property.”<sup>152</sup>

Finally, the researchers noted several institutional responses that had helped to increase access to research tools. These included the creation of public and quasi-public databases of basic research information (such as GenBank and the SNPs Consortium), and efforts of the National Institutes of Health (NIH) to negotiate greater access to research tools or to require funding recipients not to patent their research.<sup>153</sup> Further, researchers

---

146. *See id.* at 335.

147. *Id.*

148. *See id.* at 319-21.

149. *See id.* at 321.

150. *See id.* at 322.

151. *See* Act of Dec. 12, 1980, Pub. L. No. 96-517, § 6(a), 94 Stat. 3015, 3019-27 (codified in relevant part at 35 U.S.C. §§ 200-211) (commonly referred to as the Bayh-Dole Act after its legislative sponsors).

152. John Walsh, Ashish Arora & Wesley Cohen, *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, in PATENTS IN THE KNOWLEDGE-BASED ECONOMY, *supra* note 10, at 322.

153. *See id.* at 329.

avoided research tool patents by performing research outside the United States.<sup>154</sup>

### B. The Walsh, Cho, Cohen Study (2005)

Following the Wash, Cohen, Arora study, some of the same researchers sought to determine what effect *Madey* may have had on practices and on their prior conclusions.<sup>155</sup> They surveyed 414 biomedical researchers in universities, government, and nonprofit institutions to assess their patent and patented technology acquisition practices.<sup>156</sup> By the time of the study, the researchers found little evidence that *Madey* had significantly changed academic patent clearance practices—finding only five percent of respondents regularly checked for patents on knowledge inputs and only two percent had begun checking since *Madey*.<sup>157</sup> Only eight percent of respondents believed their research used information or knowledge covered by a third-party's patent, and there was little effect of such knowledge on scientific research practices—no one reported abandoning research, about one percent changed their research approach, and about one percent were delayed by more than one month. Thus, the researchers concluded that “for the time being, access to patents on knowledge inputs rarely imposes a significant burden on academic biomedical research,” noting the difference between the “law on the books” and “law in action.”<sup>158</sup>

Nevertheless, the researchers noted (compared to five years earlier) an increase from fifteen to twenty-two percent in institutional notifications to respect intellectual property rights and a slight increase from three to five percent in warning letters from patent holders.<sup>159</sup> The researchers also noted more significant concerns regarding material transfers, identifying more substantial impediments to academic research from lack of physical access. Specifically, they noted that nineteen percent of respondents had their most recent request for a material denied, and that the frequency of such denials was increasing.<sup>160</sup> However, they were unable to conclude “whether patent policy contributes to restricted access to materials, although the commercial activities fostered by patent policy do seem to re-

---

154. See Walsh et al., *Working Through*, *supra* note 140, at 1021.

155. See John P. Walsh, Charlene Cho, & Wesley M. Cohen, *View from the Bench: Patents and Material Transfers*, 309 SCIENCE 2002 (2005).

156. *Id.*

157. *Id.*

158. *Id.*

159. *Id.*

160. *Id.* at 2002-03.

strict sharing, as do the burden of producing the materials and scientific competition.”<sup>161</sup>

### C. The AAAS Study (2006-2007)

In a pilot phase of the AAAS study, the survey was administered in 2005 to 4,017 AAAS members, of which 1,111 responded.<sup>162</sup> Of the forty-six percent of respondents who reported obtaining intellectual property for their scientific discoveries or technologies since 2001, fifty-five percent reported obtaining at least one patent, and of these forty-one percent described their most important patented invention as a research tool.<sup>163</sup>

In contrast, twenty-four percent of respondents had acquired patented technology for use in their research since 2001, with rates of use and sources of acquisition varying by technology and by industrial or academic setting.<sup>164</sup> Similarly, the methods of acquiring patented technologies (including material transfer agreements (MTAs), exclusive and nonexclusive licenses, confidentiality and sponsored research agreements, and informal transfers) and the time required to do so varied significantly among respondents, with significant percentages taking more than six months.<sup>165</sup>

Unlike in the Walsh, Cho, Cohen study, which used a different methodology, the AAAS study reported that forty percent of respondents found that difficulties in obtaining patented technology since 2001 had affected their research.<sup>166</sup> Of those respondents, fifty-eight percent reported delays in research, fifty percent reported changing their research, and twenty-eight percent reported abandoning their research.<sup>167</sup>

The second phase of the AAAS study produced comparable results. The survey was administered in 2006 to scientists in the United States, the United Kingdom, Germany, and Japan.<sup>168</sup> In the United States, the survey

---

161. *Id.* at 2003.

162. STEPHEN HANSEN ET AL., THE EFFECTS OF PATENTING IN THE AAAS SCIENTIFIC COMMUNITY 5 (2006).

163. *Id.* at 7.

164. *Id.* at 14-17.

165. *Id.* at 18-20 (explaining that approximately thirty percent of respondents took more than six months in regard to acquisition by exclusive licenses).

166. *Id.* at 21; *see also id.* at 21 n.14 (noting differences in survey methodology between the studies).

167. *Id.* at 22 (reporting abandonment because of overly-complex licensing negotiations (58%), high individual royalties (49%), the patents were not licensable (40%), and licensing breakdowns (36%).

168. SCI. & INTELLECTUAL PROP. IN THE PUB. INTEREST, AM. ASS'N FOR THE ADVANCEMENT OF SCI., EFFECTS OF INTELLECTUAL PROPERTY PROTECTIONS ON THE CONDUCT OF SCIENTIFIC RESEARCH: RESULTS OF A SURVEY OF U.S. AAAS MEMBERS 2-3 (2007).

was administered to 8,000 AAAS members, of which 2,157 responded.<sup>169</sup> Fifty-two percent of those respondents answering the question had created or contributed significantly to a technology considered eligible for intellectual property protection, with the largest percentage for industry, academic, and government and others acquiring at least one patent.<sup>170</sup> Of those acquiring patents, academics principally patented research tools (forty-five percent), in contrast to industry (twenty-eight percent).

Thirty two percent of those respondents had acquired a technology protected by intellectual property for use in their research since 2002. Of these, fifty-four percent classified their last acquired technology as a research tool.<sup>171</sup> Various methods were used to acquire their last technology, but a low percentage of acquired research tools involved exclusive licenses.<sup>172</sup>

Of those who responded that they had acquired technologies protected by intellectual property, thirty-two percent reported encountering difficulties since 2002.<sup>173</sup> Most research tools were acquired within one month; in contrast, most non-research tool acquisitions took longer than six months.<sup>174</sup> The most common problem reported was overly-complex licensing negotiations, and the most common effect for academics was delay and for industry was changed research; relatively few reported abandoning projects.<sup>175</sup>

#### **D. The Holman Human Gene Patent Litigation Study (2007)**

In 2007, one of the authors of this Article (Holman) published the results of a comprehensive survey, which attempted to identify (using various databases) all lawsuits that have been filed asserting infringement of a human gene patent.<sup>176</sup> Although human gene patents represent a relatively small subset of patents covering research tools, they have raised a disproportionate level of concern both in the United States and abroad, which has led to proposals in Congress to codify a broader experimental-use exception for gene sequence patents,<sup>177</sup> to limit the enforceability and reme-

---

169. *Id.* at 2.

170. *Id.* at 3.

171. *Id.* at 2.

172. *Id.* at 2-3 (reporting 7% for academic and 13% for industrial researchers).

173. *Id.* at 3.

174. *Id.*

175. *Id.*

176. Christopher M. Holman, *The Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation*, 76 U. MO.-KAN. CITY L. REV. 295 (2007).

177. *See supra* Section III.4.

dies associated with gene patents, or to ban gene patents outright.<sup>178</sup> Although the Holman study does not directly address social practices or measure the extent to which research activities have been curtailed or modified due to the potential for patent liability, it does provide some objective insight into patent holder and research tool user behaviors. Other commentators have posited a correlation between assertion of a patent in court and patent value,<sup>179</sup> and the Holman study relies on this correlation as a useful indicator of the effects of patents on research and innovation.<sup>180</sup>

Holman identified a total of thirty-one distinct lawsuits involving human gene patents, only seven of which involved an allegation of infringement of a patented human gene in research (i.e., use as patented research tools). In sixteen of the lawsuits, the alleged infringer was a biotechnology company using a patented human gene in the manufacture of a recombinant therapeutic protein. In six of the lawsuits, the alleged infringer was a provider of genetic diagnostic testing services. The remaining two lawsuits involved patented DNA probes useful in forensic identification and paternity testing.<sup>181</sup>

None of the seven lawsuits involving patented research tools resulted in a final judicial decision. In one lawsuit, a lower court found in favor of the patent holder, but the parties settled while the case was on appeal, with the defendant reportedly paying \$718,000 for “licensing fees and other expenses.”<sup>182</sup> Five of the lawsuits settled before a final ruling by the district court.<sup>183</sup> One lawsuit, which alleged that a nonexclusive licensee had exceeded the scope of its license, was stayed pending the results of an arbitration of the underlying contract dispute.<sup>184</sup>

In five of the lawsuits involving patented research tools, the patent holder was actively using the patented technology in a commercial context at the time of the lawsuit.<sup>185</sup> Two of the lawsuits appear to have involved nonpracticing patent holders, but both of the patent holders demonstrated a willingness to license the technology on a nonexclusive basis.<sup>186</sup> In all of the seven research tool lawsuits, the infringer was alleged either to be sell-

---

178. See Holman, *supra* note 176, at 295, 359.

179. See generally John R. Allison et al., *Valuable Patents*, 92 GEO. L. J. 435 (2004).

180. See Holman, *supra* note 176, at 303-04.

181. See *id.* at 323-51.

182. *Id.* at 342 (citing to Cistron Biotechnology, Inc., Annual Report (Form 10-K), Notes to Financial Statements, at n.9 (Sept. 28, 1999)).

183. See *id.* at 341-45.

184. See *id.* at 346.

185. See *id.* at 341-45.

186. See *id.*

ing the gene (or the protein encoded by the gene) as a research tool or to be employing the gene in a commercial drug discovery effort specifically targeting the protein encoded by the gene.<sup>187</sup> In some cases, the drug discovery was part of a company's own internal research efforts, although in one case it was conducted on a contract basis.<sup>188</sup>

The Holman study identified no instance in which a lawsuit was filed to address basic, noncommercial research using gene patents. This is consistent with unpublished findings of one of the authors (Holman), who searched for but was unable to identify any instance after *Madey* in which a university researcher was sued for infringement for conducting basic research of a purely noncommercial nature. It is also consistent with the often made observation that a de facto research-use exception exists for noncommercial research.<sup>189</sup> Reasons for the lack of such lawsuits may include the desire to rely on such research to broaden markets for research tools, and the limited damages that may be obtained for such uses relative to the costs of litigation (particularly given the uncertain legal status of reach-through royalties for any products developed from the research uses).<sup>190</sup>

One of the lawsuits identified in the Holman study exemplifies the reluctance of patent holders to use their patents to block noncommercial research. The defendant was engaged in substantial commercial drug development efforts targeting the protein product of the patented gene, and the patent holder was pursuing a research program targeting the same protein. The parties settled at an early stage, before any substantive rulings by the court, with the defendant agreeing to discontinue commercial drug discovery efforts involving the patented gene. However, the settlement agreement explicitly provided that the defendant and others were free to continue using the patented gene in conjunction with basic, noncommercial research activities.<sup>191</sup>

The Holman study also found no evidence from the lawsuits of an anticommons, or patent thicket, problem in regard to gene patents and research. If a researcher were to be sued for using a gene that is only one of multiple genes being studied, this might indicate a patent thicket problem.

---

187. *See id.* at 340-42.

188. *See id.*

189. *See, e.g., Hearings, supra* note 58 (statement of Dr. Marc Grodman, Chair of the Board & CEO, Bio-Reference Laboratories, Inc.).

190. *See, e.g., Bayer AG v. Housey Pharm., Inc.*, 228 F. Supp. 2d 467, 470-71 (D. Del. 2002) (suggesting that such reach-through royalties as contractual licensing conditions could constitute patent misuse); Holzapfel & Sarnoff, *supra* note 12, at 147 & n.12; Walsh et al., *Working Through, supra* note 140, at 1021.

191. Holman, *supra* note 176, at 345.

However, all of the lawsuits identified in the study allege the use of a specifically patented human gene as a central element of a substantial commercial product or research program.<sup>192</sup> Conversely, the Holman study provided evidence of gene patents being designed around (although not in the research tool context), and of a research tool patent being circumvented by off-shoring research activities (to Taiwan) and importing the resulting data back into the United States.<sup>193</sup>

Finally, the Holman study suggests that gene patent holders have generally chosen not to assert their patents against researchers using the patented technologies, choosing instead to tolerate widespread infringement. To illustrate this point, consider the 2004 study by Kyle Jensen and Fiona Murray that identified a total of 4270 human gene patents claiming 4382 human genes (roughly 20% of human genes known at the time).<sup>194</sup> It is reasonable to assume that a significant number of these patented genes are the subject of research in the United States.<sup>195</sup> However, Holman found that these 4270 patents had resulted in only six lawsuits involving eighteen patents with claims reciting thirteen distinct human genes.<sup>196</sup> Most of these lawsuits settled early, and the only lawsuit reaching a substantive decision held that the patent had not been infringed.<sup>197</sup> Furthermore, only one of the lawsuits involved the use of a patented human gene as a research tool.<sup>198</sup> In that case, a genomics company filed a retaliatory infringement lawsuit after being sued by a research tool company for patent infringement. The parties quickly settled under terms granting the research tool company a nonexclusive license under the gene patents.<sup>199</sup>

In summary, it appears that the growing numbers of patents on research tools and the expanded liability for research uses of patented inventions resulting from *Madey* have not yet led to serious problems for the conduct of scientific research in the United States. In large part, this is because there remains a widespread practice of conducting what (in light of the *Madey* decision) can now only be considered infringing research, and

---

192. *See id.* at 340-47.

193. *See id.* at 336-37, 344.

194. Jensen & Murray, *supra* note 131, at 239-40.

195. *See id.* at 240 (noting that "heavily patented genes tended to have relevance to human health and diseases").

196. Holman, *supra* note 176, at 353-55. Most of the litigated human gene patents found by Holman were not identified in the Jensen & Murray study. *Id.* at 355.

197. *Id.* at 353.

198. *See id.* at 343. Four of the lawsuits were brought against providers of genetic diagnostic testing services, and one against a biotechnology company producing a therapeutic protein. *Id.* at 342-46.

199. *Id.*

because patent holders have continued to restrain themselves from aggressively asserting patents. Nevertheless, the studies demonstrate an increasing trend towards restriction of access and some delays in or changes to research, and the potential exists for patent holders to expand their efforts to enforce their patents (particularly if reach-through damages become available on discoveries made using their patented research tools). Thus, significant concerns remain, particularly regarding the stability of the working solutions that have been employed in the past.

## V. RECENT CHANGES TO PATENTING AND LICENSING POLICIES AND PRACTICES

This Part discusses recent licensing policies, particularly with regard to patented research tools, that have been adopted by various governmental, academic, and industrial institutions. These new policies may further affect developing scientific researcher and patent holder practices, potentially disturbing the working solutions currently in place. However, these new policies could potentially provide additional stability to the informal norms of patent infringement and forbearance of patent assertions in non-commercial contexts.

A substantial proportion of research tools patents, particularly those relating to genetics and biomedical research, arise out of government-funded and university research. Thus, one approach to addressing concerns that research tool patents might impede research and innovation is to encourage these institutions to adopt patenting and licensing practices that promote broad and nondiscriminatory access to patented research tools. Government funding agencies, including the NIH, which is the primary source of biomedical research funding in the United States, have implemented internal policies and external funding practices and have published guidelines relating to the patenting of biomedical research tools. These practices and guidelines are aimed at discouraging the patenting of certain inventions and at encouraging licensing practices that promote the dissemination of and access to biomedical research tools. Universities also have adopted patenting and licensing practices aimed at addressing concerns regarding the potential adverse effects of research tool patents.

For example, laboratories funded by the NIH and the U.S. Department of Energy (DOE) have agreed to adhere to the Bermuda Rules,<sup>200</sup> which encourage early and open access to genetic sequence information and dis-

---

200. See Bermuda Sequence Policies: Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing, [http://www.ornl.gov/sci/techresources/Human\\_Genome/research/bermuda.shtml#1](http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml#1) (last modified Oct. 29, 2003).

courage the patenting of genes by DNA sequencing laboratories.<sup>201</sup> The National Human Genome Research Institute (NHGRI), part of the NIH, has required that major genome sequencing centers (MGSCs) receiving grant funding agree to abide by the Bermuda Rules,<sup>202</sup> and NHGRI strongly encourages all of its grantees to follow these principles. The rapid public release of newly generated sequence information dictated by the Bermuda Rules serves to generate prior art that can block later patent applications. It has been suggested that prevention of DNA patenting was one factor behind the push by publicly-funded gene sequencing labs to encourage rapid entry of genetic sequence information into the public domain.<sup>203</sup> Although the Bermuda Rules are generally not binding on U.S. grant recipients (as most are not MGSCs funded by NHGRI), in practice a failure to abide by the rule would likely jeopardize the grantee's ability to secure future grant funding.<sup>204</sup>

In 1999, the NIH issued a set of principles and guidelines (the Research Tool Guidelines) that encourage grant recipients to adopt practices promoting broad access to research tools developed using NIH funds, in a manner that facilitates further biomedical research.<sup>205</sup> Although the Research Tool Guidelines are only directly applicable to recipients of NIH grant support, NIH expressed its hope that they would be adopted by the wider research community "so that all biomedical research and development can be synergistic and accelerated."<sup>206</sup> These guidelines are not regulations, and therefore are not legally enforceable. At the time they were published, the NIH expressed its view that legally enforceable regulations were not necessary, but warned that at some point in the future it might promulgate legally enforceable regulations if widespread problems continued with respect to access to NIH funded research tools.<sup>207</sup> NIH further noted that, on a case-by-case basis, the expectations set forth in the Re-

---

201. Eliot Marshall, *Bermuda Rules: Community Spirit, With Teeth*, 291 SCIENCE 1192 (2001).

202. Nat'l Human Genome Research Inst., Policy for Release and Database Deposition of Sequence Data (Dec. 21, 2000), <http://www.genome.gov/10000910>.

203. Rebecca S. Eisenberg, *Genomics in the Public Domain: Strategy and Policy*, 1 NATURE REV. GENETICS 72 (2000).

204. See Marshall, *supra* note 201, at 1192 (stating that in 1997 U.S. officials made clear that failure to comply with the rules "could be a black mark on future grant reviews.").

205. Principles and Guidelines for Recipients of NIH Research Grants and Contracts on Obtaining and Disseminating Biomedical Research Resources, 64 Fed. Reg. 72,090 (Dec. 23, 1999) (final notice) [hereinafter Research Tool Guidelines].

206. *Id.* at 72,090.

207. *Id.*

search Tool Guidelines might be imposed as specific requirements of NIH funding awards where the grant recipient has failed to demonstrate sufficient progress in implementing the Research Tool Guidelines. “Compliance with those guidelines subsequently became an explicit consideration in the award of NIH grants and contracts.”<sup>208</sup> The Research Tool Guidelines are reportedly regarded by at least some university technology transfer officers as de facto federal policy.<sup>209</sup>

The Research Tool Guidelines specifically note that “inappropriate patenting and licensing practices are likely to thwart rather than promote utilization, commercialization and public availability of research tool invention[s].”<sup>210</sup> According to the Guidelines, restrictive licensing practices are generally appropriate only in cases where “further research, development and private investment are needed to realize” the inventions’ usefulness as a research tool.<sup>211</sup> In all other cases, dissemination by “publication, deposit in an appropriate databank or repository, or widespread nonexclusive licensing” of the research tool is encouraged.<sup>212</sup> In those instances where an exclusive license is necessary to promote investment in a commercial application of a research tool, the Guidelines state that an exclusive license should ordinarily be limited to the commercial field of use, with the grant recipient retaining rights regarding the invention’s use and distribution as a research tool.<sup>213</sup>

The Research Tool Guidelines provide model language to be used in licensing agreements entered into by grant recipients, designed to promote broad dissemination of research tools. For example, the guidelines recommend that recipients reserve in their licenses the right of nonprofit institutions to use the licensed technologies internally.<sup>214</sup>

In 2005, NIH published a final notice of “Best Practices for the Licensing of Genomic Inventions” (Genomic Best Practices).<sup>215</sup> The Genomic Best Practices are generally consistent with the Research Tool Guidelines, although more explicit in clarifying that they represent rec-

---

208. Lori Pressman et al., *The Licensing of DNA Patents by US Academic Institutions: An Empirical Survey*, 24 NATURE BIOTECHNOLOGY 31, 32 (2006).

209. *Id.*

210. Research Tool Guidelines, *supra* note 205, at 72,093.

211. *Id.*

212. *Id.*

213. *Id.* at 72,095.

214. *Id.*

215. Best Practices for the Licensing of Genomic Inventions: 70 Fed. Reg. 18,413 (Apr. 11, 2005) (final notice).

ommendations of best practices, not legally binding regulations.<sup>216</sup> The Genomic Best Practices specify that:

[w]henver possible, nonexclusive licensing should be pursued as a best practice. . . . In those cases where exclusive licensing is necessary to encourage research and development by private partners, best practices dictate that exclusive licenses should be appropriately tailored to ensure expeditious development of as many aspects of the technology as possible. Specific indications, fields of use, and territories should be limited to be commensurate with the abilities and commitment of licensees to bring the technology to market expeditiously.<sup>217</sup>

The Genomic Best Practices also recommend that license agreements “be written with development milestones and benchmarks to ensure that the technology is fully developed by the licensee . . . Best practices provide for modification or termination of licenses when progress toward commercialization is inadequate.”<sup>218</sup>

In a recent survey of the nineteen U.S. academic institutions that have received the largest number of DNA patents, the researchers found that the institutions’ licensing practices were largely in agreement with NIH’s Research Tool Guidelines and Genomic Best Practices.<sup>219</sup> For example, universities prefer to enter into nonexclusive licensing arrangements with respect to most research tool DNA patents.<sup>220</sup> Some survey respondents also reported having difficulty determining whether or not an invention constituted a research tool.<sup>221</sup>

A coalition of some of the most prestigious U.S. universities have recently published a document identifying and encouraging adoption of technology licensing guidelines designed to promote broad dissemination of and access to research tool inventions. The document, entitled “In the Public Interest: Nine Points to Consider in Licensing University Technology” (Nine Points Paper)<sup>222</sup> arose out of a 2006 meeting at which representa-

---

216. *Id.* at 18,414 (“These recommendations are not intended to constitute additional regulations, guidelines, or conditions of award for any contract or grant . . .”).

217. *Id.* at 18,415.

218. *Id.*

219. Pressman, *supra* note 208, at 38-39.

220. *Id.* at 34, 38-39.

221. *Id.* at 34-35.

222. IN THE PUBLIC INTEREST: NINE POINTS TO CONSIDER IN LICENSING UNIVERSITY TECHNOLOGY (2007), available at <http://news-service.stanford.edu/news/2007/march7/>

tives of the universities gathered to discuss “societal, policy, legislative and other issues in university technology transfer.”<sup>223</sup> The licensing principles and practices identified are designed to balance the business needs of universities with their broader mandate to serve society and the public interest. The Nine Points Paper states that many of the principles were already being implemented by universities, and encourages all universities and nonprofit research entities to strive to adopt similar policies.<sup>224</sup>

In particular, the Nine Points Paper encourages universities that license patented technologies to reserve rights, in all fields of use, for themselves and for other nonprofit and government organizations to practice inventions for research and educational purposes (including research sponsored by commercial entities), even in cases where the invention is licensed exclusively to a commercial entity.<sup>225</sup> It acknowledges that in some cases the grant of an exclusive license is appropriate, perhaps even necessary, when a significant investment of time and resources in the technology is needed in order to achieve its broad implementation. However, it urges universities to strive to grant only those rights that are necessary to encourage development of the technology.<sup>226</sup>

As an overarching principle, the Nine Points Paper stresses that exclusive licenses should always be structured in a manner that encourages technology development and use.<sup>227</sup> For example, in cases where substantial investment is required to develop a research tool into a commercial product, it might be appropriate for the university to grant an exclusive license for the sale, but not the use, of such products. In doing so, the university ensures its freedom to grant other nonexclusive licenses to use the patented technology.<sup>228</sup> The Nine Points Paper notes that, absent the need for significant investment, broad nonexclusive licensing of tools such as genomic and proteomics inventions can help maximize the benefits derived from those technologies, in part by removing obstacles to further innovation.<sup>229</sup> It also emphasizes that universities are expected to make research tools as broadly available as possible.<sup>230</sup> Finally, the Nine Points Paper recommends that licensing agreements include performance miles-

---

gifs/whitepaper.pdf (white paper signed by 11 universities and the Association of American Medical Colleges).

223. *Id.* at 1.

224. *Id.*

225. *Id.* at 2.

226. *Id.*

227. *Id.*

228. *Id.*

229. *Id.* at 3.

230. *Id.* at 5.

tones to promote diligent development and broad dissemination of the licensed technology.<sup>231</sup>

The Wisconsin Alumni Research Foundation (WARF), a technology-licensing affiliate of the University of Wisconsin, has moved to improve access to its patents by researchers. WARF, although university-based, may act like a commercial entity in licensing its patented technologies. On January 23, 2007, WARF announced changes to its licensing policies that improve the terms of access for academic and nonprofit researchers.<sup>232</sup> WARF had been widely criticized for what many have characterized as overly restrictive licensing policies with respect to its broadly claimed human embryonic stem cell patents.<sup>233</sup> Under the new policies, researchers at academic and nonprofit institutions will not need a license to use WARF patented stem cells, even in private company-sponsored research.<sup>234</sup> However, this policy does not extend to any right to “develop and/or use [the human embryonic stem cells] for any therapeutic or commercial purpose, including the right to . . . perform services (including diagnostic services) for consideration, or for the production or manufacture of products for sale or distribution to third parties.”<sup>235</sup> Therapeutic or commercial users of the cells are required to seek an additional license from WARF, the terms of which do not appear in WARF’s announcement.<sup>236</sup> According to a statement by WARF Managing Director Carl E. Gulbrandsen, “WARF’s stem cell policies have evolved over the years, always in favor of increasing access and making it easier for scientists to move the technology forward. These latest changes reflect an ongoing dialogue with researchers and university administrators across the country.”<sup>237</sup>

---

231. *Id.* at 3.

232. Joyce E. Cutler, *Wisconsin Research Foundation Amends Stem Cell Policies*, 73 PAT. TRADEMARK & COPYRIGHT J. (BNA) 368 (2007) (discussing changes announced in 2007 to ease licensing requirements for academic and nonprofit researchers); Press Release, Wisconsin Alumni Research Foundation [WARF], Wisconsin Alumni Research Foundation Changes Stem Cell Policies to Encourage Greater Academic, Industry Collaboration (Jan. 23, 2007), available at [http://www.warf.org/news/news.jsp?news\\_id=209](http://www.warf.org/news/news.jsp?news_id=209).

233. See Thayer & De Liberty, *supra* note 71.

234. See Press Release, WARF, *supra* note 232.

235. WiCell Research Inst., Memorandum of Understanding—ESI Materials (July 2008), available at [http://www.wicell.org/index.php?option=com\\_content&task=blogcategory&id=124&Itemid=190](http://www.wicell.org/index.php?option=com_content&task=blogcategory&id=124&Itemid=190) (follow “ESI MOU & SLA (US and Non-US)” hyperlink) (form agreement).

236. WiCell and the National Stem Cell Bank—FAQs for Requesting Stem Cells, [http://www.wicell.org/index.php?option=com\\_content&task=blogcategory&id=124&Itemid=190&limit=1&limitstart=1](http://www.wicell.org/index.php?option=com_content&task=blogcategory&id=124&Itemid=190&limit=1&limitstart=1) (last visited Oct. 22, 2008).

237. Press Release, WARF, *supra* note 232.

Studies of industrial licensing practices in regard to patented research tools are not generally available, but are needed to provide a more complete assessment of the current licensing environment in regard to patents held by commercial entities and used as research tools. In part, such studies may be impeded by commercial desires to keep secret the terms of commercial licenses and the results of licensing negotiations.

The effects of these new patenting and licensing policies have yet to be evaluated. In particular, it remains to be seen how these policies will interact with the changes to the experimental-use and regulatory-approval exceptions and the social practices that have developed in regard thereto. Nevertheless, these policies are likely to ameliorate to some extent restrictions on access to patented technologies used in scientific research that may develop. In turn, implementation of these policies and their effectiveness in assuring access may be affected by broader changes to legal standards within the patent system.

## **VI. RECENT AND PROPOSED CHANGES TO THE PATENT SYSTEM THAT MAY AFFECT RESEARCH TOOL PATENTS AND USE OF RESEARCH TOOLS**

Since the turn of the century, government agencies, nonprofit institutions, bar associations, and academic commentators have expressed concern about the state of the U.S. patent system, and have offered various suggestions for judicial and legislative reform.<sup>238</sup> These concerns have addressed, among other things: the administrative processes and legal standards for granting patents (resulting in patents that arguably should not have been issued and that are subsequently protected by a statutory presumption of validity interpreted to impose a heightened evidentiary burden of proof)<sup>239</sup>; and expansion of patent rights and remedies (resulting in routine grants of injunctions that provide excessive negotiating leverage and excessive damage awards compared to the inventive contribution of the patented invention to the infringing product).<sup>240</sup> These concerns thus have

---

238. See, e.g., FTC, TO PROMOTE INNOVATION, *supra* note 60; PATENT SYSTEM, *supra* note 64; AIPLA RESPONSE, *supra* note 67; Patent Law Academics' Positions on Patent Law Reform Issues (June 27, 2005) (submitted to the Subcommittee on Courts, the Internet, and Intellectual Property of the House Committee on the Judiciary, on file with author); Mark A. Lemley, Douglas Lichtman & Bhaven N. Sampat, *What to Do About Bad Patents*, REGULATION, Winter 2005, at 10.

239. See 35 U.S.C. § 282 (2000 & Supp. II 2002); *Am. Hoist & Derrick Co. v. Sowa & Sons, Co.*, 725 F.2d 1350, 1359-60 (Fed. Cir. 1984).

240. See, e.g., Mark A. Lemley & Philip J. Weiser, *Should Property or Liability Rules Govern Information?*, 85 TEX. L. REV. 783 (2007); Joshua D. Sarnoff, *Bilcare*,

led to proposals for judicial or legislative reforms of existing patent law doctrines.

Recent decisions and opinions of the Supreme Court, and (to a lesser extent) of the Federal Circuit and the U.S. Patent and Trademark Office (PTO), have responded to these concerns and have significantly changed the patent law landscape in the United States.<sup>241</sup> These decisions may affect the patentability of inventions contemplated for use as research tools and have the potential to significantly reduce concerns regarding access to patented technologies for use in research. Congress also is considering comprehensive legislation to reform the patent statute, and many provisions of the current draft legislation would have similar effects.<sup>242</sup> However, these legal changes also have the potential to induce unanticipated and adverse changes to patent holders' and scientific researchers' behaviors regarding the assertion of and attention to patent rights.

This Part describes specific judicial changes and proposals for legislative reform that the authors believe are most relevant to the issues presented by research tool patents. Some of these changes have raised the standards of patentability for research tools, and others have limited or may limit applicable remedies. These changes may help to reduce concerns over the potential for research tool patents to create barriers to access.<sup>243</sup> However, these changes are quite recent, and it will take some time to determine their full impact, as courts and the PTO apply the deci-

---

KSR, *Presumptions of Validity, Preliminary Relief, and Obviousness in Patent Law*, 25 CARDOZO ARTS & ENT. L.J. 995 (2008).

241. *See, e.g.*, KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727 (2007) (altering the obviousness test); Lab. Corp. of Am. Holdings, Inc. v. Metabolite Labs., Inc., 548 U.S. 124, 126 (2006) (Breyer, J., dissenting) (arguing that dismissal was improvidently granted and discussing limits to patentable subject matter); eBay, Inc. v. MercExchange, L.L.C., 547 U.S. 388 (2006) (changing the presumption for injunctions in patent cases); *In re Bilski*, 545 F.3d 943, 954, 958-61 (Fed. Cir. 2008) (restricting patentable inventions to those meeting a "machine-or-transformation" test, finding "inadequate" the "useful, concrete, and tangible result" test, and declining to adopt a "technological arts" test); *In re Fisher*, 421 F.3d 1365, 1369-78 (Fed. Cir. 2005) (addressing the utility requirement); *Ex parte Lundgren*, 76 U.S.P.Q.2d (BNA) 1385, 1388 (B.P.A.I. 2004) (declining to recognize a "technological arts" requirement for patentable subject matter); Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility, 1300 OFF. GAZ. PAT. & TRADEMARK OFFICE 142 (Oct. 26, 2005), available at [http://www.uspto.gov/web/offices/pac/dapp/opla/preognotice/guidelines101\\_20051026.pdf](http://www.uspto.gov/web/offices/pac/dapp/opla/preognotice/guidelines101_20051026.pdf) [hereinafter Interim Guidelines].

242. Patent Reform Act of 2007, H.R. 1908, 110th Cong. (2007); Patent Reform Act of 2007, S. 1145, 110th Cong. (2007).

243. *See, e.g.*, *Hearings, supra* note 58, at 10 (statement of Lawrence M. Sung, Law School Professor & Intellectual Property Law Program Director, University of Maryland School of Law).

sions to patents claiming genes and other research tools. Additional future changes to patent law doctrines also may affect patent holders' and scientific researchers' practices in unanticipated ways. Finally, there have been and will likely continue to be changes to patent claim scope and application requirements (e.g., written description and enablement requirements and literal and doctrine of equivalents infringement doctrines) that may affect the scope of such patents and whether any particular research uses infringe issued patents.<sup>244</sup>

#### A. The Utility Requirement of Section 101

In order to be patentable under section 101, an invention must be “new and useful,” with the latter term interpreted to require some identified, practical use.<sup>245</sup> This doctrine, referred to as the utility requirement, has historically served to limit the patenting of at least some research tools, particularly those involving genetic sequences and other biomolecules. This is because the disclosed use to perform further research, which might identify more substantial uses or new materials with such uses, was not considered sufficient to warrant a patent. In order to satisfy the utility requirement, a patent application must show that an invention provides some immediate practical benefit to the public that does not require further research to identify or confirm.<sup>246</sup> The requirement is not satisfied by a showing of utility only discovered after the application was filed.<sup>247</sup>

In response to concerns that patents were being issued that claimed genetic sequences of unknown function or of unknown practical significance—e.g., the controversial patent applications for expressed sequence tags (ESTs), which essentially are fragments of expressed genes, filed by the NIH in the early 1990s—the PTO in 2001 issued revised Utility Examination Guidelines (Utility Guidelines).<sup>248</sup> The Utility Guidelines required patent applicants to articulate for their inventions a “specific and substantial utility that is credible.”<sup>249</sup>

---

244. See, e.g., *Festo Corp. v. Shoketsu Kinzoku Kogyo Kabushiki Co.*, 535 U.S. 722 (2002) (applying doctrine of equivalents standards); *In re Curtis*, 354 F.3d 1347, 1355 (Fed. Cir. 2004) (applying enablement standards); *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956 (Fed. Cir. 2002) (applying written description standards).

245. 35 U.S.C. § 101 (2000); see *Brenner v. Manson*, 383 U.S. 519 (1966).

246. See, e.g., U.S. PAT. & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PRACTICE, § 2107.01(I)(C) (8th ed. Rev. 7 2008) (discussing “Research Tools”).

247. See, e.g., *Fisher*, 421 F.3d at 1371.

248. Utility Examination Guidelines, 66 Fed. Reg. 1092 (Patent & Trademark Office Jan. 5, 2001).

249. *Id.* at 1098.

In 2005, in *In re Fisher*,<sup>250</sup> the Federal Circuit essentially affirmed the Utility Guidelines. The court held that claims directed to ESTs were unpatentable given that the functions of the underlying genes were unknown, that the only asserted uses for the ESTs at that stage were as research intermediates to isolate and experiment on the relevant genes, and that the asserted uses were only possibilities that any EST could achieve but which for these ESTs had not yet been used in the real world.<sup>251</sup> Further, following the Utility Guidelines, the court held that the status of an invention as a research tool is not dispositive; rather, the question is whether the invention has “a specifically identified substantial utility . . . [rather than an] asserted utility [that] requires further research to identify or reasonably confirm.”<sup>252</sup>

The utility standard articulated by the PTO and approved in *Fisher* should preclude patents for many of the most criticized patents claiming genes, as well as for other biomedical discoveries lacking an established use beyond that as a pure research tool. In particular, this utility standard should bar patents on gene fragments or genetic sequences of unknown function or significance.

## **B. The Patentable Subject Matter Requirement of Section 101**

The patentable subject matter doctrine, which limits the types of inventions that are patentable, also may be used in the future to restrict patenting of certain genetic and research tool inventions. The statutory language of section 101 defines the scope of inventions that are patentable in the United States as any new and useful “process, machine, manufacture, or composition of matter.”<sup>253</sup> While the Supreme Court has interpreted this language broadly to potentially encompass any product or process that is “made by man,”<sup>254</sup> it has also stressed on numerous occasions that it does not extend to “laws of nature, physical phenomena, and abstract ideas.”<sup>255</sup> Since 1981, when the Supreme Court last addressed patentable subject matter, the Federal Circuit dramatically altered the standard of patentability to allow protection of a wide range of new technologies and

---

250. *Fisher*, 421 F.3d at 1365.

251. *Id.* at 1373.

252. *Id.* at 1372.

253. 35 U.S.C. § 101 (2000).

254. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980) (quoting S. Rep. No. 82-1979 at 5 (1952), and H.R. Rep. No. 82-1923 at 6 (1952)).

255. *Id.* (citing *Parker v. Flook*, 437 U.S. 584 (1978), and *O'Reilly v. Morse*, 56 U.S. 62, 112-21 (1854)); see *Diamond v. Diehr*, 450 U.S. 175, 185 (1981).

practices.<sup>256</sup> In turn, this change in law required the PTO to grant patents for such inventions, including many new genetic and biomedical inventions used in research.<sup>257</sup> However, the Federal Circuit as a whole recently began to pull back from its earlier, broadest extensions of patentable subject matter in the *In re Bilski* case, which addressed a method for managing commodity sales risks.<sup>258</sup> *Bilski* clearly restricts patents on abstract methods of calculation or electronic transformations of data that are not claimed with regard to particular “physical and tangible objects,”<sup>259</sup> and thus could be used in many kinds of research. *Bilski* also reiterates that “[p]henomena of nature” are themselves unpatentable and that an invention incorporating such phenomena must involve more than “insignificant post solution activity” to become patentable.<sup>260</sup> This new emphasis could lead to further restrictions on some biotechnology patents, such as isolated and purified genetic sequences and diagnostic methods. It is also foreseeable that the Supreme Court will soon revisit the standards for patentable subject matter.

In 2006, the Supreme Court in *Laboratory Corp. of America Holdings, Inc. v. Metabolite Laboratories, Inc.* originally accepted and later dismissed without an opinion a case that raised significant questions regarding patentable subject matter.<sup>261</sup> The patent claim broadly recited a method for detecting a vitamin deficiency, involving the two steps of: (1) assaying a patient’s body fluid for an amino acid; and (2) mentally correlating the knowledge of an elevated level of the amino acid to the existence of the vitamin deficiency.<sup>262</sup> Although the Court as a whole decided not to decide the case (likely because of a failure to plead section 101 and because the issue had not been adequately addressed below), three Justices would have decided the case and would have found the patent invalid under the exclusion for laws of nature, natural phenomena, and abstract ideas.<sup>263</sup> These Justices voiced strong reservations with respect to patents

---

256. See, e.g., A. Samuel Oddi, *Assault on the Citadel: Judge Rich and Computer Related Inventions*, 39 HOUS. L. REV. 1033, 1040 (2002).

257. See, e.g., Eileen M. Kane, *Splitting the Gene: DNA Patents and the Genetic Code*, 71 TENN. L. REV. 707, 741-47 (2004).

258. See *In re Bilski*, 545 F.3d 943, 949, 954 (Fed. Cir. 2008).

259. *Id.* at 962.

260. *Id.* at 951, 957 (quoting *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972), and *Diamond v. Diehr*, 450 U.S. 175, 191-92 (1981)).

261. *Lab. Corp. of Am. Holdings, Inc. v. Metabolite Labs., Inc.*, 548 U.S. 124 (2006).

262. See, e.g., *Metabolite Labs., Inc. v. Lab. Corp. of Am. Holdings.*, 370 F.3d 1354, 1358-64 (Fed. Cir. 2004).

263. See *Lab. Corp. of Am. Holdings, Inc. v. Metabolite Labs., Inc.*, 548 U.S. at 126 (Breyer, J., dissenting).

broadly claiming biological correlations, and an eagerness to rein in, or even reverse, a trend in the lower courts towards an overly expansive definition of patentable subject matter. Further, they suggested constitutional concerns with such patents, implying that Congress lacks the power to authorize them.<sup>264</sup> If the Court is presented with another case raising patentable subject matter issues and in better condition for appellate review, the Court might decide it in a manner consistent with the views of the dissenting Justices.

If either the Federal Circuit or the Supreme Court further restricts what qualifies as patentable subject matter, the holdings may significantly affect the patentability of some genetic and research tool discoveries. Some genetic technology companies clearly recognized the potential for such a result in *Laboratory Corp.*, filing amicus briefs arguing that a decision could substantially affect genetic inventions, especially those involving “correlations.” For example, as amicus Perlegen (a personalized medicine company patenting discoveries regarding genetic disease correlations) argued:

Virtually every patent claim concerning a diagnostic method is based, explicitly or implicitly, on a correlation between a test result and a disease or medical condition. Thus, the repercussions for biotechnology, particularly diagnostics, if [the Court were to invalidate the claim at issue for encompassing unpatentable subject matter] would be staggering. Hundreds, if not thousands, of patents would at once be called into question.<sup>265</sup>

Similarly, amicus Affymetrix analogized the claim at issue to controversial patents on a breast cancer gene and to patents claiming SNPs, and urged the Court to invalidate the claim in a manner that would bar the patenting of what it characterized as “natural genetic phenomena.”<sup>266</sup>

Less than two months after the Supreme Court dismissed *Laboratory Corp.*, a district court in an unreported order held in *Classen Immunotherapies, Inc. v. Biogen IDEC*, that various method claims were invalid for encompassing unpatentable natural phenomena.<sup>267</sup> Specifically, the claims recited methods for determining vaccination protocols, based on compar-

---

264. *See id.*

265. Brief for Perlegen Sciences, Inc. & Mohr, Davidow Ventures as Amici Curiae Supporting Respondents, *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. 124 (2006), No. 04-607, 2006 WL 303908.

266. Brief for Affymetrix, Inc. & Professor John H. Barton Amici Curiae Supporting Petitioner, *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. at 126 (2006), No. 04-607, 2005 WL 3597814.

267. *Classen Immunotherapies, Inc. v. Biogen IDEC*, Memorandum Order, No. 04-2607 (D. Md. Aug. 16, 2006).

ing the incidence of immune disorders between two or more groups of subjects immunized under different schedules. The court characterized the claims as indirect attempts to patent the idea of a correlation between the vaccination schedules and chronic immune-mediated disorders. The decision was affirmed in late 2008 in a one-paragraph, unpublished opinion relying on the *Bilski* decision,<sup>268</sup> which may suggest the invalidity of many such correlation claims and may therefore reduce some of the concerns that have been voiced with regard to biomedical research tool patents.

### C. The Nonobviousness Requirement

Section 103 of the U.S. patent statute imposes a patentability requirement of a nonobvious invention (or inventive step), which also might restrict the patenting of many research tool inventions. Specifically, section 103 denies patentability “if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.”<sup>269</sup> For many years, the Federal Circuit and its predecessor court employed a restrictive approach to proving obviousness, requiring “a teaching, suggestion, or motivation to combine known elements” of the claimed invention that were found in the prior art.<sup>270</sup>

However, the Supreme Court in *KSR International Co. v. Teleflex, Inc.* held that a more flexible approach should be applied to determining obviousness. The Court criticized the Federal Circuit’s approach to determining whether there was “an apparent reason to combine” prior art elements of the claimed invention as “rigid,”<sup>271</sup> and noted four specific errors of the Federal Circuit’s approach in the case (which addressed a combination of an electronic sensor with an adjustable automotive foot pedal assembly). These were: (1) looking only to the problem that the patentee was trying to solve; (2) assuming the persons having ordinary skill in the art will look only to prior art designed to solve the same problem; (3) concluding that an invention cannot be proved obvious “merely by showing that the combination of elements was ‘obvious to try,’” at least when there is a design or market need and limited alternatives; and (4) seeking to prevent hind-

---

268. *Classen Immunotherapies, Inc. v. Biogen IDEC*, Nos. 2006-1634, 2006-1649 (Fed. Cir. Dec. 19, 2008), 2008 WL 5273107.

269. 35 USC § 103(a) (2000).

270. *KSR Int’l Co. v. Teleflex Inc.*, 127 S. Ct. 1727, 1741 (2007) (citing *In re Bergel*, 292 F.2d 955, 956-57 (C.C.P.A. 1961)).

271. *Id.* at 1739, 1741.

sight bias by adopting “[r]igid preventative rules that deny factfinders recourse to common sense.”<sup>272</sup>

Based on the *KSR International* decision, the PTO has adopted examination guidelines that provide many potentially expansive rationales for the PTO (and by extension courts) to find a claimed invention obvious.<sup>273</sup> These include:

(A) Combining prior art elements according to known methods to yield predictable results; (B) Simple substitution of one known element for another to obtain predictable results; (C) Use of known technique to improve similar devices (methods, or products) in the same way; (D) Applying a known technique to a known device (method, or product) ready for improvement to yield predictable results; (E) “Obvious to try”—choosing from a finite number of identified, predictable solutions, with a reasonable expectation of success; (F) Known work in one field of endeavor may prompt variations of it for use in either the same field or a different one based on design incentives or other market forces if the variations would have been predictable to one of ordinary skill in the art; [and] (G) Some teaching, suggestion, or motivation in the prior art that would have led one of ordinary skill to modify the prior art reference or to combine prior art reference teachings to arrive at the claimed invention.<sup>274</sup>

These rationales may have a significant effect on the patenting of research tool inventions, particularly given that a market motivation for creating such tools may exist, there may be limited alternatives, and the need for such tools may make the solution obvious to try.

With specific relevance to gene patents and other biotechnology inventions that can be used as research tools, the Federal Circuit’s 1995 decision in *In re Deuel*<sup>275</sup> (relied on by the Federal Circuit in its *KSR International* decision<sup>276</sup>) had been widely interpreted as creating an extremely high bar for the PTO and challengers to prove that claimed inventions are

---

272. *Id.* at 1742-43 (quoting *Teleflex, Inc. v. KSR International Co.*, 119 F. App’x. 282, 289 (Fed. Cir. 2005)); see Sarnoff, *supra* note 240 12, at 1032.

273. See Examination Guidelines for Determining Obviousness Under 35 U.S.C. § 103 in View of the Supreme Court Decision in *KSR International Co. v. Teleflex Inc.*, 72 Fed. Reg. 57526 (Patent & Trademark Office Oct. 10, 2007) [hereinafter Examination Guidelines for Determining Obviousness].

274. Examination Guidelines for Determining Obviousness, 72 Fed. Reg. at 57529.

275. 51 F.3d 1552 (Fed. Cir. 1995).

276. See *Teleflex, Inc.*, 119 F. App’x. at 289 (quoting *Deuel*, 51 F.3d at 1559).

obvious.<sup>277</sup> The Federal Circuit in *Deuel* had relied on earlier precedent rejecting the “obvious to try” approach to proving obviousness<sup>278</sup> to reverse a PTO determination of obviousness of claimed isolated and purified DNA and complementary DNA sequences relating to human and bovine growth factors.<sup>279</sup> The Federal Circuit had found that the prior art references teaching a method of gene cloning and a partial amino acid sequence of a protein were not sufficient to prove obviousness, as “the PTO has not cited a reference teaching cDNA molecules, but instead has improperly rejected the claims based on the alleged obviousness of a method of making the molecules.”<sup>280</sup> A dissenting opinion in *In re Fisher* (discussed above in regard to section 101) later argued that claims to isolated and purified genetic sequences (e.g., the ESTs at issue in *Fisher*) may not be sufficiently inventive to warrant patentability, but the *Deuel* precedent has precluded the PTO from rejecting such claims as obvious under section 103.<sup>281</sup>

The effects of *KSR* have yet to be felt or adequately assessed. However, since *KSR*, the PTO issued a decision in *Ex Parte Kubin*<sup>282</sup> that further calls into question the viability of the *Deuel* precedent. In *Kubin*, the PTO cited *KSR* and the obvious-to-try rationale in affirming a patent examiner’s rejection of a claim reciting a genus of novel genetic sequences in light of prior art that was analogous to the prior art at issue in *Deuel*.<sup>283</sup> The decision is on appeal to the Federal Circuit. Depending on how the Federal Circuit decides the case, a post-*KSR/Kubin* obviousness test might preclude the patentability of many genetic inventions that were once considered patentable. In any event, what is obvious to a person skilled in the relevant art changes over time, as does the scope of the prior art, and the *Deuel* precedent may now be obsolete as applied to modern genetic discoveries.

In summary, the standards for utility, patentable subject matter, and nonobviousness have been changing in ways that may make it more difficult to obtain patents for genetic and other inventions that are likely to be used in scientific research. It is possible that such changes may lead to al-

---

277. See, e.g., Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1178-81 (2002).

278. *In re Deuel*, 51 F.3d 1552, 1559 (citing *In re O’Farrell*, 853 F.2d 894, 903 (Fed. Cir. 1988)).

279. See *id.* at 1555, 1560.

280. *Id.* at 1557.

281. See *Fisher*, 421 F.3d at 1382.

282. See *Ex parte Kubin*, 83 U.S.P.Q.2d (BNA) 1410 (B.P.A.I. 2007).

283. See *id.* at 3-6.

ternative sources of funding to provide incentives for investment, invention, and disclosure of such new technologies. Similarly, as discussed below, changes to patent remedies may also affect the desire to patent and alternatives for funding research tools. If so, there may be corresponding changes to behaviors of the remaining patent holders regarding licensing and assertion of their patents against scientific researchers who use their technologies.

#### D. Injunctive Relief Under Section 283

The recent Supreme Court decision in *eBay, Inc. v. MercExchange, L.L.C.*,<sup>284</sup> and cases following *eBay* that deny injunctive relief to patent holders,<sup>285</sup> may help to alleviate concerns that patents on research tools will be used to restrict scientific research. Conversely, to the extent that denial of injunctive relief diminishes the commercial exclusivity of patent holders and reduces their revenue or their ability to bargain for higher licensing fees, *eBay* and its progeny may reduce incentives for the creation and patenting of research tools. Further, the denial of injunctive relief and the imposition of prospective compensatory damages in the form of ongoing royalty payments<sup>286</sup> may have a similar effect to the granting of a compulsory license on commercial terms determined by a judge through litigation. Because they directly affect commercial returns to patent holders, these changes to the available remedies for patent infringement also have the potential to change existing practices and working solutions.

Under section 283 of the Patent Act, district courts “may grant injunctions in accordance with the principles of equity.”<sup>287</sup> Prior to *eBay*, Federal Circuit precedent essentially mandated that, after finding patents to be valid and infringed, trial courts permanently enjoin future infringements,

---

284. 547 U.S. 388 (2006).

285. See, e.g., Andrew Beckerman-Rodau, *The Aftermath of eBay v. MercExchange*, 126 S.Ct. 1837 (2006): *A Review of Subsequent Judicial Decisions*, 89 J. PAT. & TRADE-MARK OFF. SOC'Y 631 (2007) (discussing the holdings of post-*eBay* decisions on patent law injunctions); Posting of Joseph S. Miller to Fire of Genius, Injunctions, <http://www.thefireofgenius.org/injunctions> (last updated Dec. 31 2007) (providing a comprehensive list through Dec. 31, 2007 of decisions regarding preliminary and permanent injunctive relief in patent, copyright, and trademark law that apply the *eBay* approach).

286. See, e.g., Paice LLC v. Toyota Motor Corp., 504 F.3d 1293, 1313-16 (Fed. Cir. 2007) (holding that ongoing royalty payments rather than injunctions may be appropriate, but vacating and remanding the ongoing royalty payment at issue).

287. 35 U.S.C. § 283 (2000).

at least absent “exceptional circumstances.”<sup>288</sup> The Supreme Court in *eBay* rejected this strong presumption in favor of granting injunctions in patent cases, holding that nothing in the Patent Act suggested that patent law should depart from traditional principles of equity law, and thus a patent holder can only obtain a permanent injunction as a remedy for infringement if he or she can demonstrate: (1) that the patent holder suffered an irreparable injury due to the infringement; (2) that remedies available at law, such as monetary damages, are inadequate to compensate for that irreparable injury; (3) that, considering the balance of hardships between the patent holder and the infringer, a remedy in equity is warranted; and (4) that the public interest would not be disserved by a permanent injunction.<sup>289</sup> However, the two concurring opinions of seven of the Justices reflect very different views about when injunctions are likely to be found appropriate after finding infringement of a valid patent.<sup>290</sup>

Thus, after *eBay*, a trial court has substantially more discretion to deny an injunction—a decision that can only be reversed under the highly deferential “abuse of discretion” standard.<sup>291</sup> Denial of injunctions is more likely to occur in cases where the patented technology makes up a relatively small portion of the infringing product or process, where the patent holder is not practicing the invention, money damages and ongoing royalty payments are sufficient to compensate the patent holder, or an injunction might unduly injure the infringer and/or adversely affect public interests.<sup>292</sup> For example, the Federal Circuit in *Innogenetics N.V. v. Abbott Laboratories*<sup>293</sup> held that a trial court abused its discretion in granting an injunction against an infringer of a gene patent, given that the patent holder had requested and obtained a jury verdict that included or contemplated

---

288. See, e.g., *Accumed LLC v. Stryker Corp.*, 483 F.3d 800, 811 (Fed. Cir. 2007) (recognizing that the Supreme Court in *eBay* “struck down” the Federal Circuit’s general rule); *MercExchange, L.L.C. v. eBay, Inc.*, 401 F.3d 1323, 1338 (Fed. Cir. 2005) (citing Federal Circuit precedents that established a “general rule . . . that a permanent injunction will issue once infringement and validity have been adjudged”).

289. *eBay*, 547 U.S. at 391.

290. Compare *id.* at 394-95 (Roberts, J. concurring), with *id.* at 395 (Kennedy, J., concurring).

291. *Id.* at 391 (majority opinion).

292. See, e.g., Beckerman-Rodau, *supra* note 285, at 653-57 (discussing some of these and other factors and noting that direct competition with the patent holder is the most significant predictive factor regarding whether a permanent injunction will issue); Andrew Beckerman-Rodau, *The Supreme Court Engages in Judicial Activism In Interpreting the Patent Law in eBay, Inc. v. MercExchange L.L.C.*, 10 TUL. J. TECH. & INTELL. PROP. 165, 201-02 (2007) (discussing the component product—or “complex invention”—concern) (citing *eBay*, 547 U.S. 388, 396-97 (Kennedy, J., concurring)).

293. 512 F.3d 1363 (Fed. Cir. 2008).

an ongoing royalty for continued use, and thus the patent holder could not be considered irreparably harmed by continued infringement.<sup>294</sup> The court remanded for further assessment of the terms of the ongoing royalty for continued access to the patented technology, which claimed methods of genotyping hepatitis C virus (which depending on the end use could be considered a research tool patent).<sup>295</sup>

In contrast, an injunction is more likely to issue if the patent holder is producing and selling the patented invention and if the infringer competes in the market for such sales. In such cases, courts may consider price erosion, loss of goodwill, potential reductions in workforce, and other factors which are difficult to quantify in terms of damages.<sup>296</sup> Such considerations are less likely to apply to research uses of patented inventions than to sales of inventions intended for use as research tools.

The Federal Circuit has yet to develop a clear understanding of the “public interest” consideration in granting or denying injunctive relief after *eBay*. For example, in the context of affirming a trial court’s grant of a preliminary injunction, one panel of Federal Circuit judges recently held that the public interest factor is neutral in regard to the competing public interests in the benefits of lower prices (for printer and facsimile machine toner cartridges) from free competition and in enforcing patent rights.<sup>297</sup> Conversely, a different panel of Federal Circuit judges held that there was no abuse of discretion in a trial court holding that the public interest in acquiring lower cost pharmaceuticals (and potential deaths that would result if consumers did not purchase them) was outweighed by the public’s interest in encouraging pharmaceutical research and development by enforcing patent rights.<sup>298</sup>

It is possible that a court would refuse to grant an injunction where a patented invention was used by an infringer as a research tool, particularly if the patent holder was engaged in a pattern of licensing its invention or if the research at issue was particularly important. As the 2004 NAS report suggested, injunctive relief “would rarely be an appropriate remedy in a research infringement case, because from these research uses there would rarely be ongoing commercial losses to the patent holder.”<sup>299</sup> Further, as the Supreme Court noted in *eBay*, the trial court had focused on the patent

---

294. *Id.* at 1380-81.

295. *Id.*

296. *See, e.g.,* *Sanofi-Synthelabo v. Apotex, Inc.*, 470 F.3d 1368, 1381-83 (Fed. Cir. 2006).

297. *See* *Canon, Inc. v. GCC Int’l Ltd.*, 263 F. App’x. 57 (Fed. Cir. 2008).

298. *See Sanofi-Synthelabo*, 470 F.3d at 1383-84.

299. PATENT SYSTEM, *supra* note 64, at 116.

holder's willingness to license the technology and its failure to itself practice the invention.<sup>300</sup> However, the Court nevertheless cautioned that no broad, categorical rule could be adopted, and for certain patent holders such as universities a willingness to license might not weigh against issuing the injunction.<sup>301</sup>

In summary, the four-part equitable test is highly sensitive to the facts of each case and to the discretionary judgments of particular judges. This renders the potential for obtaining injunctive relief in regard to research tool uses of patented inventions highly uncertain. Nevertheless, it is clear that the potential to obtain an injunction has been reduced since *eBay*, and consequently that the threat that scientific researchers will be prohibited from continuing to conduct experiments (or forced to negotiate licenses prior to or after litigation at higher rates, given the threat or grant of an injunction) is correspondingly reduced. Additional studies are needed to assess the extent to which these changes will affect incentives to develop and patent research tools, as well as the ability of scientific researchers to acquire and their willingness to use patented technologies as research tools.

#### **E. Potential Legislation Affecting Damages Remedies Under Section 284**

The U.S. Congress is considering as part of comprehensive legislation to reform the U.S. Patent Act a provision that would alter the existing rules governing calculation of royalty damages for infringement of patent rights.<sup>302</sup> The proposed change to the law would respond to perceived excesses in jury damage awards that are based on calculating royalty rates with regard to the entire value of the infringing product, even though the patent holder's invention may represent only a fraction of the patented and unpatented technologies included in the infringing product.<sup>303</sup> For example, the proposed changes in the U.S. Senate would: (1) limit reliance on the "entire market value" rule for calculating the royalty base to cases where the patent holder's invention was the predominant basis for the market demand for the infringing product; (2) permit royalties to be based on similar, nonexclusive licenses if enough such licenses indicate that the royalty terms are reasonable; and (3) if neither (1) nor (2) apply, limit the royalty base to the portion of the economic value of the infringing inven-

---

300. *eBay*, 547 U.S. 388 at 392.

301. *Id.*

302. See Patent Reform Act of 2007, H.R. 1908, 110th Cong. (2007); Patent Reform Act of 2007, S. 1145, 110th Cong. (2007).

303. See, e.g., S. Rep. No. 110-259, at 12 (2008).

tion attributable to the patented invention's contribution over the prior art (which for inventions consisting of novel combinations of prior art elements may consist of the additional function or enhanced value of the combination).<sup>304</sup>

Although it is difficult to predict whether such revisions will be enacted into law, they would clearly tend to limit recoverable royalty damages in regard to technologies incorporated into commercial products and to patents that are nonexclusively licensed. Thus, such changes could affect the damages recoverable for competing sales of research tool inventions (or products incorporating those inventions) for scientific research uses. Similarly, such changes could affect royalties recoverable for scientific uses of research tool inventions, as well as potential royalties for new products resulting from the research and incorporating the research tool (which therefore infringe the rights of making and of sale, as well as of use). Further, such changes could affect reach-through royalties that might be recoverable for scientific research uses of patented inventions to develop valuable information, products, or processes that do not infringe the patented invention. As with injunctive relief, reducing the potential scope of damage awards could affect incentives for investment in and invention and patenting of research tools, as well as willingness to use patented technologies in scientific research.

## VII. ALTERNATIVES TO EXPERIMENTAL-USE AND REGULATORY-APPROVAL EXCEPTIONS TO INFRINGEMENT

The previous parts of this Article have discussed the historic development of the experimental-use exception and regulatory-approval exception, the effects of these legal developments on the practices of seeking patents on research tools and of using patented technologies for scientific research and commercial development, and responses taken by the government, academic institutions, and industry to assure that patents do not restrict access to the technologies or their use for scientific research and commercial development. This Part addresses existing and proposed legal and practical alternatives to these exceptions, which can help to assure access and continued use of patented technologies in scientific research and commercial development. These alternatives include: (1) compulsory licensing and functional equivalents thereto; (2) government licenses and march-in rights regarding federally funded inventions; (3) reach-through licensing agreements; (4) patent pools; (5) antitrust remedies; and (6) off-

---

304. *See id.* at 13-14.

shoring of research activities. Additional legal development and studies are needed to determine the extent to which such alternatives can be and will be used to assure access to patented inventions for use in scientific research.

#### A. Compulsory Licensing

Compulsory licensing provisions were considered for possible incorporation into the 1952 revision of the U.S. patent laws—the most recent comprehensive revision to and codification of U.S. Patent Act. However, these provisions were removed from draft legislation before the final bill was introduced.<sup>305</sup> Since then, “[c]ompulsory licensing of patents often has been proposed, but it has never been enacted on a broad scale.”<sup>306</sup> As late as 2005, a bill was introduced in Congress that would have provided for compulsory licensing of certain patented inventions relating to health care emergencies, but the bill never became law.<sup>307</sup> The patent reform bills currently being considered by Congress include no compulsory licensing provisions.<sup>308</sup> As noted in a 2004 report of the NAS on the patent system, there is a prevalent hostility in industry and among patent holders generally to any form of compulsory licensing.<sup>309</sup>

Nevertheless, U.S. law does provide some limited forms of compulsory licensing of patented technologies. For example, the Clean Air Act provides for the compulsory licensing of patents on pollution control devices to those parties who cannot use substitutes to meet pollution control requirements imposed under the statute.<sup>310</sup> The existing compulsory licensing provisions, however, have little if any relevance to the use of patented research tools, particularly those used in the context of biomedical research.

Of greater relevance, use by the U.S. government of any and all patented inventions is fully authorized by statute (and is consistent with the World Trade Organization’s Agreement on Trade-Related Aspects of Intellectual Property Rights (the TRIPS Agreement), subject to the payment of adequate remuneration but taking into consideration any anti-

---

305. See H. COMM. ON THE JUDICIARY, 81ST CONG., PROPOSED REVISION AND AMENDMENT OF THE PATENT LAWS 91 (Comm. Print 1950).

306. Dawson Chem. Co. v. Rohm & Haas Co., 448 U.S. 176, 215 & n.21 (1980).

307. Public Health Emergency Medicines Act, H.R. 4131, 109th Cong. (2005).

308. Patent Reform Act of 2007, H.R. 1908, 110th Cong. (2007); Patent Reform Act of 2007, S. 1145, 110th Cong. (2007).

309. See PATENT SYSTEM, *supra* note 64 at 115.

310. See 42 U.S.C. § 7608 (2000).

competitive practices).<sup>311</sup> Under 28 U.S.C. § 1498(a), a patent holder's sole legal remedy for an infringing manufacture, use or sale of a patented invention by the U.S. government—or by any person or entity working under the “authorization and consent” of the U.S. government (i.e., a government contractor)—is a legal claim for “reasonable compensation.” This legal claim requires the patent holder to file a lawsuit against the U.S. Government in the U.S. Court of Claims to prove infringement (and where challenged to defend the validity of the patent). However, unlike a normal patent infringement lawsuit, the patent holder cannot obtain injunctive relief to prohibit continuing infringement by the government. (The patent holder may seek to prohibit a third-party's use by filing a lawsuit in a federal district court seeking an injunction, and the third party must prove authorization under section 1498 as an affirmative defense.)<sup>312</sup> Like ongoing royalty damages, section 1498 operates similarly to a compulsory license, particularly as the U.S. government might invoke its authorization on behalf of third parties.<sup>313</sup>

All of the research conducted by, and much of the research conducted for the U.S. government falls under the protection of section 1498(a).<sup>314</sup> The provision is often explicitly invoked on behalf of grantees or contractors to assure access to patented technologies.<sup>315</sup> The authors are unaware of any instance where section 1498(a) has been explicitly invoked to induce voluntary licensing of a patented research tool (although voluntary licensing of such tools may routinely occur given recognition that use without permission of the patent holder may be authorized by section 1498(a)). However, the government has on occasion explicitly threatened to invoke section 1498(a) in order to compel a patent holder to license its patent in rare cases where the patent is perceived to cover the only viable means to address a potential massive public health emergency.

---

311. See Agreement on Trade-Related Aspects of Intellectual Property Rights, art. 31, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994).

312. See *Madey v. Duke Univ.*, 413 F. Supp. 2d 601, 607 (M.D.N.C. 2006).

313. BRIAN T. YEH, CONG. RESEARCH SERV., CRS REPORT NO. RL33159, INFLUENZA ANTIVIRAL DRUGS AND PATENT LAW ISSUES (2005).

314. As discussed below, use by the government and its contractors also may be authorized by a statutory license arising from the use of federal funds in the development of the invention, and the existence of such a license and its scope in regard to infringing activity may only be resolved in a suit seeking compensation under section 1498. See *Madey*, 413 F. Supp. 2d at 608. Conversely, use by state governments is immunized from compensatory liability by the 11th Amendment to the U.S. Constitution, but injunctive relief may still be available. See U.S. CONST., amend. XI; Fla. Prepaid Postsecondary Educ. Expense Bd. v. College Sav. Bank, 527 U.S. 627, 633-635 (1999).

315. See *Madey*, 413 F. Supp. 2d at 607-08.

Notable recent examples involved Roche's Tamiflu and Bayer's Ciprofloxacin, thought to be critical in responding to fears of an avian flu pandemic and an anthrax bioterrorism attack, respectively.<sup>316</sup> In both cases, the government was reportedly able to use the threat to gain significant concessions from patent holders without actually authorizing third-party production under section 1498(a).<sup>317</sup>

Given that legislative enactment of a broad experimental-use exception might not occur, the 2004 NAS report on the patent system recommended that the federal government consider assuming liability under the "authorization and consent" provision of section 1498(a) for the infringement of research tool patents by investigators whose work it supports under contracts, grants, and cooperative agreements.<sup>318</sup> However, the report noted that authorization under section 1498(a) has not often been extended to federal grantees in this context, and has never been formally extended by the NIH (although reportedly the DOE has exercised this option).<sup>319</sup> One member of the NAS committee issuing the 2004 report recommended that the government consider providing authorization under section 1498(a) for scientific research uses of patented inventions only in cases where access to research tool technologies is not resolved in the marketplace by licensing on reasonable terms, and predicted that in all likelihood the threat of its use would lead to a negotiated solution.<sup>320</sup> The report itself recommended that federal agencies include explicit authorization and consent "as a reasonable step that addresses the need to maintain research tool access."<sup>321</sup>

Similarly, an ongoing royalty damage award (which may be considered a compulsory license<sup>322</sup>) can be achieved in instances where a court declines to enter an injunction against a party found liable for infringing a research tool patent. As noted above, *eBay* has significantly expanded the

---

316. YEH, *supra* note 313; JOHN R. THOMAS, CONG. RESEARCH SERV., CRS REPORT RL32051, INTELLECTUAL PROPERTY ISSUES IN HOMELAND SECURITY (2007).

317. James P. Love, KNOWLEDGE ECOLOGY INT'L, *Recent Examples of the Use of Compulsory Licensing of Patents* (2007), available at [http://www.keionline.org/misc-docs/recent\\_cls.pdf](http://www.keionline.org/misc-docs/recent_cls.pdf).

318. *See* PATENT SYSTEM, *supra* note 64 at 115.

319. *Id.*

320. *Id.* at 117.

321. *Id.*

322. *Compare* Paice LLC v. Toyota Motor Corp., 504 F.3d 1293, 1313 n.13 (Fed. Cir. 2007) (distinguishing the two because there is no authorization for third party use other than by the parties to the lawsuit), *with id.* at 1316 (Rader, J., concurring) ("[C]alling a compulsory license an 'ongoing royalty' does not make it any less a compulsory license.").

courts' discretion to deny injunctions, and courts may in the future do so for research uses of patented inventions. In *Genomic Best Practices*, the NRC recommended that "[c]ourts should continue to decline to enjoin patent infringement in those extraordinary situations in which the restricted availability of genomic or proteomic inventions threatens the public health or sound medical practice."<sup>323</sup>

Given that compulsory licensing, and its functional equivalents of governmental authorization under section 1498(a) and refusals to enjoin continued infringement, can assure research uses of patented inventions, a number of academic commentators have proposed that the U.S. institute some form of compulsory licensing (or codify an experimental-use exception either providing for compensation to patent holders or specifically targeting certain types of research uses) so as to promote access to patented research tools in certain situations. For example, Rebecca Eisenberg has proposed a compulsory licensing regime that would deny "patent holders an injunctive remedy to prevent subsequent researchers from using their inventions to make further advances in the same field," but would allow the patent holder a reasonable royalty.<sup>324</sup>

In contrast, Katherine Strandburg has proposed that the research tool inventor be granted an initial period of a few years of complete exclusivity, after which the technology would be subject to compulsory licensing.<sup>325</sup> This proposal is designed to provide adequate compensation for the inventor while ensuring that the research tool is not withheld from other researchers for the entire length of the patent term.<sup>326</sup> Strandburg has predicted that the compulsory license provision would rarely be invoked, but would incentivize the patent holder to negotiate a voluntary license during the initial period of complete exclusivity.<sup>327</sup>

Janice Mueller has proposed a "liability rule" model that permits the non-consensual "development use" of research tools not readily available for licensing or purchase, while providing an ex post royalty payment to the patent owner that would be correlated to the commercial value of the new product developed from the non-consensual use.<sup>328</sup> This reach-through royalty approach seeks to approximate the true worth of the research tool to its user. According to Mueller, her proposal would ensure a

---

323. REAPING THE BENEFITS, *supra* note 69, at 146-47.

324. Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1075-77 (1989).

325. Strandburg, *supra* note 78, at 143-44.

326. *Id.* at 143-45.

327. *Id.* at 141-42.

328. Mueller, *supra* note 8, at 58.

royalty award of sufficient amount to maintain incentives for the development and patenting of new research tools, yet alleviate the access restrictions and up-front costs currently associated with acquisition and use of many proprietary research tools.

Rochelle Dreyfuss has proposed a plan pursuant to which a university or other nonprofit research institution that wanted to use patented material and cannot obtain a license from the patentee on reasonable terms could use the technology without permission if it were willing to sign a waiver of potential patent rights.<sup>329</sup> The waiver would require the institution to promptly publish the results of work conducted with the patented technology and to refrain from patenting discoveries made in the course of that work.<sup>330</sup> Richard Nelson has proposed a modification of the Dreyfuss waiver plan, which would allow the researchers to patent their work but would require them to agree to license on a nonexclusive basis for reasonable royalties.<sup>331</sup>

Jordan Karp has proposed a

modified experimental use exception whereby an inventor is paid a “reasonable royalty” by those who experiment on her patented innovation. This type of scheme treats experimental use as a type of limited compulsory licensing. . . . Under this paradigm, the royalty payment required from the experimenter could be tied to the commercial success of any innovation resulting from the experimental activity on the patented invention. An experimenter would only have to compensate the patentee when the experimental activity actually resulted in a benefit to the experimenter (thus, allowing “pure” scientific research to continue unhindered).<sup>332</sup>

This proposal would effectively impose reach-through royalty licensing for research tool uses, which is a controversial approach.

David Parker has suggested that a statutory research exemption could undermine the value of patents covering basic research tools by rendering them essentially incapable of infringement.<sup>333</sup> Thus, Parker has proposed

---

329. Rochelle C. Dreyfuss, *Protecting the Public Domain of Science: Has the Time for an Experimental Use Defense Arrived*, 46 ARIZ. L. REV. 457, 471-72 (2004).

330. *Id.* at 471.

331. Richard R. Nelson, *The Market Economy, and the Scientific Commons*, 33 RES. POL'Y 455, 467 (2004).

332. Jordan P. Karp, Note, *Experimental Use as Patent Infringement: The Impropriety of a Broad Exception*, 100 YALE L.J. 2169, 2188 (1991).

333. David L. Parker, *Patent Infringement Exemptions for Life Science Research*, 16 HOUS. J. INT'L L. 615, 659 (1994).

that “[i]f an exception for ‘commercial’ research and development is warranted,” the approach should be

based upon the concept of allowing the commercial use of a patented invention in research and development and only making this commercial research activity subject to infringement once a decision has been made to commercialize the fruits of that endeavor. Of course, if the activity results in a product or process within the scope of the patented technology, the end product or process itself would be actionable without regard to the underlying technology used in its development. In short, only the research activities would receive the ‘limited-time’ protection, not the end result of that research.<sup>334</sup>

At a recent Congressional hearing relating to the patenting of human genes, Lawrence Sung proposed that the U.S. establish a research use exception limited to basic, noncommercial research.<sup>335</sup> Under his proposal, academic researchers and institutions would be exempt from infringement liability for noncommercial research activities, with the caveat that the researchers and institutions must provide actual notice to the patent holder of the open and notorious use of the patented technology for basic research uses, and agree to dedicate the results of the research to the public.<sup>336</sup>

#### **B. Government Rights to Inventions Patented Under the Bayh-Dole Act**

As summarized in a 1998 report by the NIH Working Group on Research Tools:

The Bayh-Dole Act [“Bayh-Dole”] provides the statutory basis and framework for federal technology transfer activities, including the patenting and licensing of federally funded inventions by recipient organizations. The Act permits recipients of federal grants and contracts to elect title to patentable “subject inventions” that arise with the use of federal funds. If recipients elect title, the Act requires them to file patent applications, seek commercialization opportunities, and report back to the funding agency on efforts to obtain utilization of their inventions. *The*

---

334. *Id.* at 659-60.

335. *Hearings, supra* note 58, at 10 (statement of Lawrence M. Sung, Law School Professor & Intellectual Property Law Program Director, University of Maryland School of Law).

336. *Id.* at 13-14.

*Act also retains for the funding agency certain residual rights in subject invention.*<sup>337</sup>

Bayh-Dole has led to dramatic changes in the economic structure of research and norms of open science, as well as to increased patenting of basic research discoveries by federally funded academic research institutions.<sup>338</sup>

Under Bayh-Dole, for all inventions made in the course of federally funded research, the federal government retains “a non-exclusive, non-transferable, irrevocable, paid-up license to practice or have practiced for or on behalf of the United States any subject invention throughout the world.”<sup>339</sup> However, the NIH Working Group noted that while

[t]his license gives the NIH, and any other agency of the Federal Government, the right to use any patented research tool arising in the course of federally-sponsored research without liability for patent infringement[, it] is not clear whether NIH's retained license [] allows NIH to authorize use of subject inventions by other recipients of NIH grants. Some agencies take the position that the activities of grantees are covered by the exemption, but NIH has considered it an open question.<sup>340</sup>

Bayh-Dole also provides that a federal agency engaged in research funding, such as NIH, can “march-in” and grant licenses to patented inventions arising out of funded research under certain specified circumstances, including when the agency determines that such action is necessary because the grantee has “not taken, or is not expected to take within a reasonable time, effective steps to achieve practical application of the subject invention in such field of use,” or when such “action is necessary to alleviate health or safety needs which are not reasonably satisfied by the” grantee.<sup>341</sup>

The NIH Working Group suggested that NIH might exercise the march-in right “on a case by case basis to improve access to particular research tools.” However, the Working Group noted that “[i]n order to exercise march-in rights, the funding agency must comply with a lengthy ad-

---

337. REPORT OF THE NATIONAL INSTITUTES OF HEALTH (NIH) WORKING GROUP ON RESEARCH TOOLS, App. D, <http://www.nih.gov/news/researchtools/appendd.htm> [hereinafter NIH RESEARCH TOOLS REPORT] (emphasis added).

338. See generally Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 L. & CONTEMP. PROBS. 289 (2003).

339. 35 U.S.C. § 202(c)(4) (2000 & Supp. II 2002).

340. NIH RESEARCH TOOLS REPORT, *supra* note 337.

341. 35 U.S.C. § 203(a) (2000 & Supp. II 2002).

ministrative process,” and that “[e]ach particular case can be expected to be lengthy and uncertain.” The NIH Working Group also noted that, because of this administrative burden the mechanism “does not lend itself to routine use.”

The NIH has never asserted its march-in rights in the nearly twenty-eight (28) years since the Act was enacted. It has denied at least three (3) formal requests to exercise the right (none of which was brought with respect to a patented research tool), concluding that the patented technologies were being made reasonably available under the patent.<sup>342</sup> In denying the requests, NIH noted that it was concerned that exercising its march-in rights would act as a disincentive for investment in the development of commercial products based on inventions patented under Bayh-Dole. It has also stated that the march-in right is not intended to be used to compel patent holders to make patented technology available at lower prices, and that “manufacture, practice, and operation . . . [by the patent holder providing for] availability and use by the public” is sufficient to meet the standard.<sup>343</sup>

Absent a sharp departure from past practice or a legislative change, it seems unlikely that the NIH or other federal agencies will exercise their march-in rights with respect to a research tool patent absent some showing that the restrictive practices of the patent holders are precluding all access to the technology or substantially impairing the “health or safety needs” of the U.S. public. This would likely be a difficult showing to make. However, a witness at a recent Congressional hearing on gene patents strongly urged Congress to consider legislation that would encourage more active use of the march-in provision to promote accessibility to genetic diagnostic testing services. If Congress acts on this proposal it could perhaps open the door to the use of march-in rights more broadly with respect to patented research tools.

### C. Reach Through Licensing Agreements

Under a reach-through licensing agreement (“RTLA”), the licensor receives a share of the profits generated by the ultimate commercial prod-

---

342. See, e.g., In the Case of Norvir® Manufactured by Abbott Laboratories, Inc. (Nat’l Insts. Health July 2, 2004), available at <http://www.ott.nih.gov/policy/March-In-Norvir.pdf> [hereinafter Norvir®]; In the Case of Xalatan® Manufactured by Pfizer, Inc. (Nat’l Insts. Health September 17, 2004), available at <http://ott.od.nih.gov/policy/March-in-xalatan.pdf>; DETERMINATION In the Case of PETITION OF CELLPRO, INC. (Nat’l Insts. Health August 1, 1997), available at <http://www.nih.gov/news/pr/aug97/nihb-01.htm>.

343. Norvir®, *supra* note 342, at 5; see *id.* at 5-6.

uct if and only if the research tool is used in the development of such a product.<sup>344</sup> However, RTLAs are controversial because they raise potential antitrust and patent misuse issues, given that the patent holder may require as a condition of use of the patented invention that the licensee provide compensation (at least in part) for uses or sales of unpatented aspects of the products developed with the patented invention.<sup>345</sup> The legal resolution may depend in part on the market power of the patent holder and the specific form of the licensing offer in conditioning access to the patented technology.<sup>346</sup> According to the 2003 FTC report on the patent system, some representatives of the biotechnology industry reported that RTLAs have been successfully employed to provide commercial researcher with access to patented research tools.<sup>347</sup> These representatives expressed the view that RTLAs can promote access to a wide range of research tools at low up-front cost, and facilitate risk-sharing between licensor and licensee. However, other panelists interviewed for the FTC report argued that RTLAs create anticommons problems, and might violate antitrust and patent misuse laws.<sup>348</sup>

#### D. Patent Pools

Patent pools involve “patents [from multiple patentees being] licensed in a package, either by one of the patent holders or by a new entity established for this purpose, usually to anyone willing to pay the associated royalties.”<sup>349</sup> The Biotechnology Industry Organization (BIO), a leading trade association representing biotechnology companies, has stated that voluntary patent pools are “one of the most important potential solutions to concerns regarding overlapping patents.”<sup>350</sup> Similarly, the PTO has released a report entitled “Patent Pools: A Solution to the Problem of Access and Biotechnology Patents?” which discusses the use of patent pools as a

---

344. See, e.g., Thomas J. Kowalski & Christian M. Smolizza, *Reach-Through Licensing; a US Perspective*, 6 J. COMMERCIAL BIOTECH. 349.

345. See, e.g., *id.* at 349 n.1; Research Tool Guidelines, *supra* note 205 (“[I]mposing reach-through royalty terms as a condition of use of a research tool is inconsistent with this principle [of ensuring appropriate distribution of NIH-funded tools].”).

346. See, e.g., *Zenith Radio Corp. v. Hazeltine Research, Inc.*, 395 U.S. 100, 139 (1969) (“While a licensee must pay if he uses the patent . . . he may insist upon paying only for use, and not on the basis of total sales”); *Bayer AG v. Housey Pharms., Inc.*, 228 F. Supp. 2d 467, 470-71 (D. Del. 2002) (rejecting allegation that reach through licensing agreement constituted patent misuse where the licensee voluntarily agreed to the royalty provision).

347. FTC, TO PROMOTE INNOVATION, *supra* note 60, ch. 3, at 26-28.

348. *Id.*

349. Carl Shapiro, *supra* note 131, at 119-150.

350. FTC, TO PROMOTE INNOVATION, *supra* note 60, ch. 3, at 27.

means of fostering access to patented research tools.<sup>351</sup> The 2003 FTC report on the patent system notes that the “centralized management that the patent pools entails may help in avoiding the royalty stacking/complements problem that economists have suggested may develop when multiple patents are needed for follow-on activities, and each patentee independently determines its own royalty rates.”<sup>352</sup>

Nevertheless, some have questioned whether high transaction costs might substantially limit the ability to form and the use of patent pools in the context of genetic inventions.<sup>353</sup> It has been noted that these technologies are fundamentally different from the electronics sector, in which patent pools are used more frequently because of the importance of standards and interoperability.<sup>354</sup> Further, the greater unpredictability of biotechnological inventions that may result in wider differences in valuation of patented technologies, and the potentially greater reliance of biotechnology companies on maximizing licensing revenues may reduce incentives for particular patent holders to join or to agree to standard licensing terms of patent pools.<sup>355</sup>

Nevertheless, various proposals have been put forward for creating specific research tool patent pools. For example, Affymetrix, a leading DNA microarray company, has been an outspoken advocate for the creation of gene patent pools.<sup>356</sup> A group of European scholars has published a series of articles discussing the potential use of patent pools to facilitate access to genetic technologies for use in diagnostic testing.<sup>357</sup> Merrill

---

351. JEANNE CLARK ET AL., PATENT & TRADEMARK OFFICE, U.S. PATENT AND TRADEMARK OFFICE, PATENT POOLS: A SOLUTION TO THE PROBLEM OF ACCESS IN BIOTECHNOLOGY PATENTS? (2000), available at <http://www.uspto.gov/web/offices/pac/dapp/opla/patentpool.pdf>.

352. FTC, TO PROMOTE INNOVATION, *supra* note 60, ch. 3, at 42.

353. *See id.* ch. 3, at 28.

354. *Id.*

355. *Cf.* Ted J. Ebersole et al., *Patent Pools as a Solution to the Licensing Problems of Diagnostic Genetics*, 17 INTELL. PROP. & TECH. L.J. 6, 10 (2005) (discussing differences among the genomics industry that make it difficult to identify “essential patents”).

356. Barbara Caulfield, Executive V.P. & Gen. Counsel, Affymetrix, Inc., Presentation at Personalized Medicine and Molecular Diagnostics Conference at Arizona State University (March 2, 2007), available at [http://www.law.asu.edu/files/Centers\\_and\\_Programs/LST/Conferences\\_&\\_Events/caulfield.pdf](http://www.law.asu.edu/files/Centers_and_Programs/LST/Conferences_&_Events/caulfield.pdf) (slide presentation).

357. Geetruui Van Overwalle et al., *Models for Facilitating Access to Patents on Genetic Inventions*, 7 NATURE REVIEWS: GENETICS 143 (2006); Brigit Verbeure et al., *Patent Pools and Diagnostic Testing*, 24 TRENDS IN BIOTECHNOLOGY 3 (2006) [hereinafter Verbeure, *Patent Pools*]; Brigit Verbeure et al., *Analyzing DNA Patents in Relation with Diagnostic Genetic Testing*, 14 EUR. J. HUM. GENETICS 1, 26-33 (2006); Esther van Zimmeren et al., *A Clearing House for Diagnostic Testing: The Solution to Ensure*

Goozner of the Center for Science in the Public Interest has proposed a patent pool for the California Institute of Regenerative Medicine and other funders of stem cell research.<sup>358</sup> Similar approaches could prove useful for biomedical research tools. However, to date patent pooling has not played a significant role in the biotechnology sector. The best known example of a biotechnology patent pool is probably the collection of patent rights cobbled together to provide freedom of operation to produce “Golden Rice” (a genetically engineered rice that produces  $\beta$ -carotene, the precursor to vitamin A, which give the rice grains a yellow hue).<sup>359</sup> Golden Rice is not considered a commercially relevant crop, and licenses under the pool were granted free of charge, essentially for humanitarian reasons.<sup>360</sup> There has also been an attempt to create a pool of patents relating to SARS research, but so far there appears to have been no report that this attempt has been consummated.<sup>361</sup>

### E. Antitrust approaches

Some commentators, including Rochelle Dreyfuss, have argued that competition law should be invoked in certain circumstances to compel patent holders to make patented research tools available, particularly where the patent holder is effectively blocking downstream research on a biologic target of significant clinical importance, e.g., the BRCA breast cancer genes.<sup>362</sup> There is a long history in the United States of judicially imposed compulsory licenses to remedy antitrust violations or concerns, where patent holders exercise or seek to acquire monopoly market power or engage in other prohibited practices. There are also compulsory licenses imposed or agreed to in regard to administrative reviews (in the context of merger and acquisition reviews by the Federal Trade Commission, the U.S. agency that formulates and enforces much of the U.S. antitrust law and policy).<sup>363</sup>

---

*Access to and Use of Patented Genetic Inventions?*, 84 BULL. WORLD HEALTH ORG. 5, 337 (2006).

358. See Merrill Goozner, *Innovation in Biomedicine: Can Stem Cell Research Lead the Way to Affordability?*, 3 PLOS MED. 126, 612 (2006).

359. Verbeure, *Patent Pools*, *supra* note 357.

360. *Id.*

361. *Id.*

362. Rochelle C. Dreyfuss, *Unique Works/Unique Challenges at the Intellectual Property/Competition Law Interface* (N.Y.U. L. & Econ. Research Paper Series, Working Paper No. 05-13, 2005), available at <http://ssrn.com/abstract=763688>.

363. JEROME H. REICHMAN, *COMPULSORY LICENSING OF PATENTED INVENTIONS: COMPARING UNITED STATES LAW AND PRACTICE WITH OPTIONS UNDER THE TRIPS AGREEMENT* (May 14, 2006), available at <http://www.aals.org/documents/2006intprop/JeromeReichmanOutline.pdf> (presented at AALS Mid-Year Workshop on Intellectual

The FTC (along with the U.S. Department of Justice (DOJ)) recently indicated that they are unlikely to impose compulsory licenses. Their view is that although unilateral refusals to license are permissible, conditional refusals will be reviewed for antitrust violations under a “rule of reason” analysis.<sup>364</sup> Nevertheless, the FTC and DOJ have shown some willingness in the merger context to require licensing of patented research tool technology in cases where the merger has the potential to decrease the number of firms researching in a particular area.<sup>365</sup> For example, when the large biotechnology companies Amgen and Immunex merged, the FTC required them to agree to license out some of their patented research tools relating to the development of drugs targeting interleukin-1.<sup>366</sup>

However, U.S. courts, while willing to impose compulsory licenses to remedy antitrust violations, have shown little if any inclination to apply the antitrust laws to compel access to research tools. For example, in *Digene Corporation v. Third Wave Technologies Inc.*,<sup>367</sup> a district court recently rejected allegations that a patent infringement plaintiff violated the Sherman Act<sup>368</sup> by monopolizing the market for human papilloma virus (HPV) testing.<sup>369</sup>

Federal Circuit and Supreme Court precedents effectively preclude using antitrust and misuse law to address unilateral refusals to license, as well as conditional refusals to license, so long as the conditions are within the scope of patent rights. This is true even when the patent holder is not actively exploiting the technology, or is even suppressing it. For example,

---

Property, Vancouver, Canada, June 14-16, 2006); *see generally* JEROME H. REICHMAN & CATHERINE HASENZAHN, NON-VOLUNTARY LICENSING OF PATENTED INVENTIONS: HISTORICAL PERSPECTIVE, LEGAL FRAMEWORK UNDER TRIPS, AND AN OVERVIEW OF THE PRACTICE IN CANADA AND THE USA (2003), [http://ictsd.net/downloads/2008/06/cs\\_reichman\\_hasenzahl.pdf](http://ictsd.net/downloads/2008/06/cs_reichman_hasenzahl.pdf).

364. *See* DOJ & FTC, ANTITRUST ENFORCEMENT AND INTELLECTUAL PROPERTY RIGHTS: PROMOTING INNOVATION AND COMPETITION 15-32 (2007) (citing, *inter alia*, *Image Technical Servs., Inc. v. Eastman Kodak Co.*, 125 F.3d 1195 (9th Cir. 1997), and *In re Indep. Serv. Org. Antitrust Litig.*, 203 F.3d 1322 (Fed. Cir. 2000)).

365. *See* Dreyfuss, *supra* note 362, at 12.

366. Press Release, FTC, Resolving Anticompetitive Concerns, FTC Clears \$16 Billion Acquisition of Immunex Corp. by Amgen Inc. (July 12, 2002), *available at* <http://www.ftc.gov/opa/2002/07/amgen.shtm> (reporting consent agreement requiring Amgen and Immunex to license intellectual property rights relating to IL-1 inhibitors in view of the potential therapeutic relevance of these drugs).

367. *Digene Corp. v. Third Wave Techs., Inc.*, No. 07-0022, 2008 WL 450467 (W.D. Wis. 2008).

368. 15 U.S.C. § 2 (2000 & Supp. IV 2004).

369. *Digene*, 2008 WL 450467 at \*8-\*10.

in *Rite-Hite Corp. v. Kelley Co.*,<sup>370</sup> an en banc panel of the Federal Circuit held that “[t]here is no requirement in this country that a patentee make, use or sell its patented invention.”<sup>371</sup> The *Rite-Hite* Court did suggest, however, the court might in some circumstances refuse to enjoin patent infringement in cases of non-use, in effect creating a compulsory license: “If a patentee’s failure to practice a patented invention frustrates an important public need for the invention, a court need not enjoin infringement.”<sup>372</sup> After *eBay*, courts have more discretion to act upon this suggestion.

The Federal Circuit held in *Monsanto Co. v. McFarling*<sup>373</sup> that its earlier decision in *Mallinkrodt, Inc. v. Medipart, Inc.*<sup>374</sup> established that in “the cases in which the [conditional licensing] restriction is reasonably within the patent grant, the patent misuse defense can never succeed,” because such conditions cannot extend the patent right beyond the patent’s scope.<sup>375</sup> Similarly, in *Virginia Panel Corp. v. Mac Panel Co.*<sup>376</sup> the Federal Circuit held that violation of the antitrust laws “requires more exacting proof than suffices to demonstrate patent misuse.”<sup>377</sup> However, the continuing validity of any rule relying on *Mallinkrodt* and its progeny was recently called into question by the Supreme Court’s decision in *Quanta Computer Inc. v. LG Electronics, Inc.*,<sup>378</sup> which arguably effectively overruled *Mallinkrodt*.<sup>379</sup> Finally, the Supreme Court recently held in *Verizon Communications Inc. v. Law Offices of Curtis v. Trinko, LLP*<sup>380</sup> that the right to refuse to deal is not unqualified, but that it has “been very cautious in recognizing [abuse of dominant position, essential facilities, or other] exceptions, because of the uncertain virtue of forced sharing and the difficulty of identifying and remedying anticompetitive conduct by a single firm.”<sup>381</sup> This is in contrast with the European Union, where doctrines

---

370. 56 F.3d 1538 (Fed. Cir. 1995) (en banc).

371. *Id.* at 1547 (citing *Cont’l Paper Bag Co. v. E. Paper Bag Co.*, 210 U.S. 405, 424-30 (1908)); *see also, e.g.*, *Cygnus Therapeutics Sys. v. ALZA Corp.*, 92 F.3d 1153 (Fed. Cir. 1996).

372. *Rite-Hite*, 56 F.3d at 1547.

373. 363 F.3d 1336 (Fed. Cir. 2004).

374. 976 F.2d 700, 708 (Fed. Cir. 1992).

375. *Monsanto*, 363 F.3d at 1341.

376. 133 F.3d 860 (Fed. Cir. 1997).

377. *Id.* at 872.

378. 128 S.Ct. 2109, 2121-22 (2008); *see also* Transcript of Oral Argument, *Quanta Computer, Inc. v. LG Elecs., Inc.*, 128 S.Ct. 2109 (2008) (No. 06-937), *available at* [http://www.supremecourtus.gov/oral\\_arguments/argument\\_transcripts/06-937.pdf](http://www.supremecourtus.gov/oral_arguments/argument_transcripts/06-937.pdf).

379. *Cf.* *LG Elec., Inc. v. Bizcom Elec., Inc.*, 453 F.3d 1364 (Fed. Cir. 2006).

380. 540 U.S. 398 (2004).

381. *Id.* at 408.

such as essential facilities and abuse of dominant position tend to hold greater sway.<sup>382</sup>

Absent a substantial shift in U.S. policy, it seems unlikely that antitrust law will play a significant role in compelling research tool patent holders to expand access to the patented technology. To the contrary, some have expressed the concern that antitrust laws could restrict the availability of certain private ordering approaches to deal with the effect of research tool patents, such as patent pools or licensing arrangements.<sup>383</sup>

#### F. Off-shoring Research

One commentator has argued that “current U.S. jurisprudence is forcing U.S. drug companies to outsource their early stage drug research” to other countries.<sup>384</sup> Indeed, U.S. patent law would allow many research tool patents to be avoided by off-shoring certain uses of research tools to other countries where the tool is not patented, where patent enforcement is more difficult, or where use of the research tool would be more likely to fall under an experimental or research use exception. In general, U.S. patent law only reaches activities performed within the United States, and the Supreme Court recently expressed its view that U.S. patent law should generally be interpreted in a manner that minimizes the impact of U.S. law on extra-territorial activities.<sup>385</sup> However, U.S. patent law does include certain exceptions to this general principle, some of which could be relevant with respect to the susceptibility of U.S. patents to avoidance by off-shoring of research activities.

For example, section 271(g) of the Patent Act<sup>386</sup> provides that, under certain circumstances, a party can be held liable for infringement based on the importation into the U.S., or use or sale in the U.S., of a product produced outside the country by a process covered by a U.S. patent. Thus, in some cases the extraterritorial use of patented research tool process could result in liability for infringement under section 271(g) if a physical product of the process is imported into the United States. An example might be a cell line created outside the United States by a process patented in the United States. However, a 2003 decision by the Federal Circuit makes clear that section 271(g) only applies to physical products, and does not apply to information generated by a patented process.<sup>387</sup> Thus, a U.S.

---

382. *Dreyfuss, supra* note 362 at 13.

383. FTC, TO PROMOTE INNOVATION, *supra* note 60, ch. 3, at 26-28.

384. Helm, *supra* note 93.

385. *Microsoft Corp. v. AT&T Corp.*, 127 S.Ct. 1746, 1751 (2007).

386. 35 U.S.C. § 271(g) (2000).

387. *Bayer AG v. Housey Pharms.*, 340 F.3d 1367 (Fed. Cir. 2003).

company should be free to off-shore certain research activities to avoid a U.S. patent, and then bring the resulting data and insights back into the United States for subsequent drug development activities.

Conversely, a U.S. firm might be liable for patent infringement under section 271(f)<sup>388</sup> for exporting a component of a patented research tool that is subsequently incorporated into the patented research tool extraterritorially. For example, export of a noninfringing DNA vector which is subsequently used to create a cell line that would infringe a U.S. patent might, under certain circumstances as limited by the language of the statute, be the basis for a finding of infringement under section 271(f). However, a recent Supreme Court decision, *Microsoft v. AT&T*, indicates that the export of information, or software, which is later incorporated extraterritorially into a research tool covered by a U.S. patent will not infringe under section 271(f), which requires at least the export of tangible embodiments of the information that are capable of being used in a claimed process or product.<sup>389</sup> In *Microsoft*, the Supreme Court held that section 271(f) was not applicable where computer software was first sent from the United States to a foreign computer manufacturer on a master disk, or by electronic transmission, and then copied by the foreign recipient for installation on computers made and sold abroad, since the copies, as “components” installed on the foreign made computers, were not supplied from the United States.<sup>390</sup>

In summary, to the extent that the failure to provide a broad experimental-use or regulatory-approval exception provides incentives for off-shoring of research using patented technologies, current law does not meaningfully restrict the ability to develop and import into the U.S. new products or processes that do not themselves infringe the claims of the patent. There is no current consensus on whether broader exceptions are desirable to prevent such off-shoring of research.

## VIII. CONCLUSIONS

The law regarding the experimental-use and regulatory-approval exceptions to patent infringement has changed over time. In recent years, the Federal Circuit has narrowly construed the scope of the experimental-use exception in ways that largely preclude its application to patented research tools used in academic or commercial scientific research. In contrast, the Supreme Court and the Federal Circuit have construed the regulatory-

---

388. 35 U.S.C. § 271(f) (2000).

389. *See Microsoft*, 127 S.Ct. at 1746.

390. *Id.* at 1755-59.

approval exception broadly, and district courts have determined that the exception applies to at least some research tools and may soon determine that it applies to sales for research tool uses.

These legal developments have led to varied practical responses by academic and commercial scientists. Although the effects of the developments on access to patented technologies and on scientific research and development are uncertain, large-scale adverse effects have to date been avoided by adoption of working solutions to restrictions on access. These solutions include perceived widespread infringing activity and consequent forbearance from assertion of patents by patent holders. Nevertheless, the discontinuity between the law on the books and the law in practice continues to pose concerns that more serious problems of access may develop.

Further, the stability of the existing working solutions is uncertain, particularly in light of significant changes that are occurring to various patent law doctrines and to governmental, academic, and industrial licensing practices. The sensitivity of existing practices to these changes also is uncertain. Consequently, it is difficult to predict whether these changes, and possible consequential or extrinsic changes to patenting behaviors, funding for innovation, and patent holders' licensing behaviors, will alleviate or further exacerbate access problems regarding research uses of patented inventions. What is certain is that the issues of the scope of experimental-use and regulatory-approval exceptions, their application to research tools, practical responses and the social consequences of the rules and practices, and alternative legal and practical means for assuring access to patented inventions for research uses will remain a focus of concern and will continue to warrant careful scrutiny and empirical and theoretical analysis.

# CONGRESS SHOULD AMEND THE COPYRIGHT ACT TO PROTECT TRANSACTIONAL WATERMARKS

*By Matt Williams<sup>†</sup>*

## TABLE OF CONTENTS

I. INTRODUCTION .....	1368
II. DESPITE PERVASIVE P2P INFRINGEMENT, RECORD LABELS ARE REDUCING THEIR USE OF TPMS .....	1375
A. PERVASIVE P2P INFRINGEMENT .....	1376
B. RECORD LABELS DECIDE TO OFFER “TPM-FREE” MUSIC ONLINE.....	1379
III. WATERMARKING CAN BENEFIT COPYRIGHT OWNERS AND CONSUMERS .....	1381
IV. INADEQUATE LEGAL PROTECTION FOR WATERMARKS .....	1384
A. THE DEFINITION OF CMI: SECTION 1202’S HISTORY AND TEXT .....	1384
B. THE ABSENCE OF PROHIBITIONS AGAINST CIRCUMVENTION DEVICES AND SERVICES.....	1391
V. PRIVACY AND SPEECH CONCERNS: PROTECTING WATERMARKS WOULD ENCOURAGE MORE SPEECH THAN IT WOULD CHILL .....	1394
A. THE FIRST AMENDMENT PROTECTS ANONYMOUS SPEECH .....	1396
B. THE TRADITIONAL CONTOURS OF COPYRIGHT PROTECTION.....	1400
C. PROTECTING TRANSACTIONAL WATERMARKS WOULD LIKELY ALTER THE TRADITIONAL CONTOURS OF COPYRIGHT PROTECTION BUT THE SPEECH RELATED BENEFITS OF EXPANDING THE DEFINITION OF CMI WOULD OUTWEIGH THE POTENTIAL CHILLING EFFECTS .....	1404
VI. CONCLUSION .....	1409

## I. INTRODUCTION

Congress passed the Digital Millennium Copyright Act (“DMCA”)<sup>1</sup> in 1998 “to encourage new ways of disseminating works in the digital era”<sup>2</sup> by providing copyright owners with new legal mechanisms to prevent on-line infringement.<sup>3</sup> Although the DMCA’s prohibitions regarding the circumvention of technological protection measures (“TPMs”)<sup>4</sup> have thus far succeeded in the former statutory objective,<sup>5</sup> they have proven less than adequate in the latter.<sup>6</sup> Despite the proliferation of lawful online services that provide consumers with access to copyrighted material, copyright infringement is still rampant online. This is especially true in the context of peer-to-peer (“p2p”) networks, where infringements of sound recordings

---

† Associate, Mitchell Silberberg & Knupp LLP, Washington D.C. The views expressed herein are the author’s alone and may not reflect the views of MS&K or its clients. The author would like to thank Alex Curtis, Peter Jaszi, Robert Kasunic, Steven Metalitz, Eric Schwartz, Coriell Wright, and the staff of BTLJ for their comments and assistance.

1. Digital Millennium Copyright Act, Pub. L. No. 105-304, 112 Stat. 2860 (1998) [hereinafter DMCA].

2. Marybeth Peters, Register of Copyrights, Copyright Enters the Public Domain, The 33rd Donald C. Brace Memorial Lecture Delivered at New York University School of Law (Apr. 29, 2004), in 51 J. COPYRIGHT SOC’Y U.S.A. 701, 722 (2004).

3. See Jane C. Ginsburg, *Copyright and Control Over New Technologies of Dissemination*, 101 COLUM. L. REV. 1613, 1618 (2001) (explaining that copyright holders “persuaded Congress that it could foster participation in digital communication only by reinforcing copyright owners’ control over the distribution of their works”).

4. This article will use the term “TPM” to refer to technological protection measures currently protected by section 1201 of title 17, including technological measures that effectively control access to works and technological measures that effectively protect rights of copyright owners. “[A] technological measure ‘effectively controls access to a work’ if the measure, in the ordinary course of its operation, requires the application of information, or a process or a treatment, with the authority of the copyright owner, to gain access to the work.” 17 U.S.C. § 1201(a)(3)(B) (2000). “[A] technological measure ‘effectively protects a right of a copyright owner. . . .’ if the measure, in the ordinary course of its operation, prevents, restricts, or otherwise limits the exercise of a right of a copyright owner. . . .” *Id.* § 1201(b)(2)(B). This article will use the term “DRM” to refer to all kinds of technologies used by copyright owners to protect their rights in digital copies of works, including technologies that are not protected by section 1201.

5. See June M. Besek, *Anti-Circumvention Laws and Copyright: A Report from the Kernochan Center for Law, Media and the Arts*, 27 COLUM. J.L. & ARTS 385, 496 (2004) (finding that “[s]ection 1201 has been successful in stimulating new means of distribution and promoting consumer choices with respect to a variety of works, particularly sound recordings, motion pictures and television programming, and literary works”); Peters, *supra* note 2, at 722; see also 17 U.S.C. § 1201 (codifying prohibitions).

6. See generally Fred von Lohmann, *Measuring the Digital Millennium Copyright Act Against the Darknet: Implications for the Regulation of Technological Protection Measures*, 24 LOY. L.A. ENT. L. REV. 635 (2004).

still occur in large numbers.<sup>7</sup> Legislative action is needed to address this problem.

Although legal protection for TPMs has not prevented this widespread infringement, it has frustrated consumers, technologists, and academics who feel that TPMs prevent them from engaging in noninfringing uses of works.<sup>8</sup> Evidence indicates that the impact of TPMs on noninfringing uses is slight.<sup>9</sup> In addition, many of the types of “personal copying” that consumers believe to be noninfringing, do not necessarily qualify as fair use.<sup>10</sup> Nevertheless, the major record labels have decided to begin selling

---

7. INT’L FED’N OF THE PHONOGRAPHIC INDUS., IFPI DIGITAL MUSIC REPORT 2008: REVOLUTION, INNOVATION, RESPONSIBILITY 18 (2008), available at <http://www.ifpi.org/content/library/DMR2008.pdf> (“Tens of billions of illegal music files are traded annually worldwide at an estimated ratio of 20 illegal downloads for every track sold. This has had a major impact on the development of legal services, holding back growth in the whole digital sector.”).

8. See, e.g., Deirdre K. Mulligan & Aaron K. Perzanowski, *The Magnificence of the Disaster: Reconstructing the Sony BMG Rootkit Incident*, 22 BERKELEY TECH. L.J. 1157, 1185 (2007) (“The constraints imposed by DRM generally reduce the value to consumers of protected content.”); Pamela Samuelson & Jason Schultz, *Should Copyright Owners Have to Give Notice of Their Use of Technical Protection Measures?*, 6 J. TELECOMM. & HIGH TECH. L. 41, 42 (2007) (“TPMs may inhibit playful and creative uses of digital works and other non-infringing uses of the content, such as time- or platform-shifting.”); Pamela Samuelson, *Intellectual Property and the Digital Economy: Why the Anti-Circumvention Regulations Need to Be Revised*, 14 BERKELEY TECH. L.J. 519, 519 (1999) (“Either Congress or the courts will be forced to constrain the reach of the anti-device rules so as not to undermine Congressional intent to preserve fair uses and so as not to harm competition and innovation in the information technology sector.”).

9. See Peters, *supra* note 2, at 723 (explaining that rulemakings regarding DMCA have shown that “by and large technological measures ha[ve] not been used in a heavy-handed or inappropriate way”); Jane C. Ginsburg, *Legal Protection of Technological Measures Protecting Works of Authorship: International Obligations and the US Experience*, 29 COLUM. J.L. & ARTS 11, 12 (2005) (“The US experience to date indicates that legal protection for technological measures has helped foster new business models that make works available to the public at a variety of price points and enjoyment options, without engendering the ‘digital lockup’ and other copyright owner abuses that many had feared.”).

10. See Sydney Aaron Beckman, *From CD to MP3: Compression in the New Age of Technology Overlooked Infringement or Fair Use?*, 42 GONZ. L. REV. 469, 499 (2007) (“The application of Shape-Shifting to music CD’s owned by individuals is an infringing use which violates the United States Copyright Act.”); Memorandum from Marybeth Peters, Register of Copyrights, Recommendation of the Register of Copyrights in RM 2002-4; Rulemaking on Exemptions from Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, to James H. Billington, Librarian of Cong. 130 (Oct. 27, 2003), available at <http://www.copyright.gov/1201/docs/registers-recommendation.pdf> (“[N]o court has held that ‘space-shifting’ is a fair use.”); Wendy M. Pollack, Note, *Tuning In: The Future of Copyright Protection for Online Music in the*

songs online without TPMs in order to accommodate the wishes of their customers.<sup>11</sup>

This is good news for consumers, who will be able to move their music libraries from device to device and enjoy their favorite songs on-the-go. It is also good news for artists that transform<sup>12</sup> works by sampling them or creating mash-ups.<sup>13</sup> However, all indications are that the labels plan to continue to utilize filtering and watermarking to protect their rights online.<sup>14</sup> Whereas filtering involves screening databases of protected material that has been “fingerprinted” against online content to spot infringement and then remove or block infringing material, watermarking involves embedding information that identifies a copyright owner, relevant terms of use and/or a purchaser of a copy in copies of works.<sup>15</sup>

---

*Digital Millennium*, 68 FORDHAM L. REV. 2445, 2482 (2000) (“[S]imply because a use is private does not necessarily make it fair.”).

11. See generally Monika Roth, Note, *Entering the DRM-Free Zone: An Intellectual Property and Antitrust Analysis of the Online Music Industry*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 515 (2008); Press Release, EMI Group, EMI Music Launches DRM-Free Superior Sound Quality Downloads Across Its Entire Digital Repertoire (Apr. 2, 2007), available at <http://www.emigroup.com/Press/2007/press18.htm>; Jeff Leeds, *Universal Music to Sell Some Music Without the Copy Protection*, N.Y. TIMES, Aug. 10, 2007, at C6; Caroline McCarthy, *Sony BMG Signs on to Amazon's DRM Free Music Store*, CNET NEWS.COM, Jan. 10, 2008, [http://www.news.com/8301-10784\\_3-9848258-7.html](http://www.news.com/8301-10784_3-9848258-7.html); Ken Fisher, *Music DRM in Critical Condition: Universal Tests DRM Free Music Sales*, ARS TECHNICA, Aug. 9, 2007, <http://arstechnica.com/news.ars/post/20070809-music-drm-in-critical-condition-universal-tests-drm-free-music-sales.html>.

12. For a discussion of the meaning of the term “transformative,” see Matt Williams, *Recent Second Circuit Opinions Indicate That Google's Library Project Is Not Transformative*, 25 CARDOZO ARTS & ENT. L.J. 303, 305 (2007) [hereinafter Williams, *Transformative*] (arguing that a transformative use involves “the creation of original expression that contains commentary”).

13. See Pat Aufderheide & Peter Jaszi, Am. Univ. Ctr. for Soc. Media, *Recut, Reframe, Recycle: Quoting Copyrighted Material in User-Generated Video*, at 7 (2008), available at [http://www.centerforsocialmedia.org/files/pdf/CSM\\_Recut\\_Reframe\\_Recycle\\_report.pdf](http://www.centerforsocialmedia.org/files/pdf/CSM_Recut_Reframe_Recycle_report.pdf) (describing mash-up videos posted to online websites); Noah Shachtman, *Copyright Enters a Gray Area*, WIRED, Feb. 14, 2004, available at <http://www.wired.com/entertainment/music/news/2004/02/62276> (describing mash-up album that combined music from the Beatles' *White Album* with Jay-Z lyrics).

14. See David Kravets, *DRM Is Dead, But Watermarks Rise From Its Ashes*, WIRED, Jan. 11, 2008, available at [http://www.wired.com/print/entertainment/music/news/2008/01/sony\\_music](http://www.wired.com/print/entertainment/music/news/2008/01/sony_music); Bill Rosenblatt, *Music Industry Accelerating Watermarking Adoption*, DRM WATCH, Aug. 16, 2007 [hereinafter Rosenblatt, *Accelerating*], <http://www.drmwatch.com/drmtech/article.php/3694781>; Bill Rosenblatt, *New Market Study Predicts Growth in Watermarking and Fingerprinting Markets*, DRM WATCH, Jan. 24, 2008, <http://www.drmwatch.com/watermarking/article.php/3723626>.

15. Bill Rosenblatt of DRM Watch has articulated the distinction between watermarks and filtering technologies that rely on fingerprinting:

A debate regarding the efficacy and benefits of filtering is already well underway among interested parties<sup>16</sup> and in the courts.<sup>17</sup> While some hope that filtering will help save the music industry from the ongoing reduction in sales partly caused by infringement, others fear that it will deprive us of myriad noninfringing uses.<sup>18</sup> With much of the copyright community's focus on filtering, watermarking has thus far received less attention from policy-minded advocates.

Watermarking has the potential to benefit consumers and copyright owners by allowing consumers to engage in personal and transformative copying while at the same time enabling copyright owners to police infringement on peer-to-peer networks.<sup>19</sup> This is especially true of "transactional watermarking" that imprints information about the consumer who purchases songs on the digital files delivered by online music services.<sup>20</sup>

---

Watermarking and fingerprinting are two forms of technology known generically as *content identification*. Watermarking works by embedding data into digital images, audio, or video in such a way that the data is very difficult to remove and the effect on a user's perception of the content is (usually) nonexistent. The data embedded in a watermark is often the identity of the content, though it could also include the identity of a user or device that downloaded it, or of a retailer that sold it. Fingerprinting is a set of techniques for analyzing content, reducing its unique characteristics to a set of one or more numbers that serve as 'fingerprints,' and looking those fingerprints up in a database to determine the identity of the content.

Bill Rosenblatt, *New DRM Watch Section on Watermarking and Fingerprinting*, DRM WATCH, Oct. 24, 2007, <http://www.drmwatch.com/watermarking/article.php/3706996>; see also Dean L. Fanelli et al., *2007 Patent Law Decisions of the Federal Circuit*, 57 AM. U. L. REV. 821, 914 (2008) ("[A] 'watermark' is data embedded into audio, video, or image signals, or data files to identify a source or copyright status of the signals or data files.").

16. *Compare* Principles for User Generated Content Services, <http://www.ugcprinciples.com/> (last visited Sept. 12, 2008) ("UGC Services should use effective content identification technology . . . with the goal of eliminating from their services all infringing user-uploaded audio and video content . . ."), with Fair Use Principles for User Generated Video Content, <http://www.eff.org/issues/ip-and-free-speech/fair-use-principles-usergen> (last visited Sept. 12, 2008) ("Human creators should be afforded the opportunity to dispute the conclusions of automated filters.").

17. See, e.g., *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 518 F. Supp. 2d 1197 (C.D. Cal. 2007) (issuing injunction including filtering mandate regarding p2p service).

18. See Alexandra Berzon, *Filtering Fair Use?*, RED HERRING, Mar. 4, 2007, <http://www.redherring.com/Home/21516>.

19. See Eliot Van Buskirk, *Are Digital Music Watermarks a Blessing or a Curse?*, WIRED, Aug. 20, 2007, [http://www.wired.com/entertainment/music/commentary/listeningpost/2007/08/listening\\_post\\_0820](http://www.wired.com/entertainment/music/commentary/listeningpost/2007/08/listening_post_0820).

20. See Rosenblatt, *Accelerating*, *supra* note 14.

Such watermarks could, as the technology improves, allow a copyright owner to determine who is making copies of sound recordings available on p2p networks in a much more specific manner than the current techniques used by record labels.<sup>21</sup> The presence of a transactional watermark identifying the original purchaser of the file would provide evidence pointing to the source of the infringing copy. This could facilitate more effective enforcement of rights in court because record labels could prove that a particular copy of a song purchased by a specific consumer was the original source of a downloaded copy available in another p2p user's folder of shared files.<sup>22</sup> In addition, a private system of curbing infringers' Internet access may develop that penalizes Internet users who distribute copies of purchased songs.<sup>23</sup>

Commentators and policy makers have seen the potential in watermarks for some time.<sup>24</sup> Unfortunately, the statutory section that Congress

---

21. *See id.*

22. Just before this Article went to press, the Recording Industry Association of America ("RIAA") announced that it plans to scale back lawsuits against individual p2p infringers and pursue agreements with Internet access providers pursuant to which infringers could have their Internet access terminated after repeated warnings regarding infringement. *See* Sarah McBride & Ethan Smith, *Music Industry to Abandon Mass Suits*, WALL ST. J., Dec. 19, 2008. Nevertheless, watermarks can provide evidence of direct infringement in suits involving allegations of secondary liability and can also provide record labels and Internet access providers with more concrete evidence of p2p infringement.

23. France, Japan and the UK are considering policies that would threaten Internet users with loss of access for repeated infringements, but such proposals have received considerable criticism. *See* Posting of Danny O'Brien to EFF Deeplinks Blog, *Three Strikes, Three Countries: France, Japan, and Sweden*, <http://www.eff.org/deeplinks/2008/03/three-strikes-three-countries> (Mar. 18, 2008); Posting of Sherwin Siy to Public Knowledge Policy Blog, *Banned from Life: Why Copyright Shouldn't Control Online Connectivity*, <http://www.publicknowledge.org/node/1416> (Feb. 27, 2008 11:48 EST); *see also* Susan P. Crawford, *The Internet and the Project of Communications Law*, 55 UCLA L. REV. 359, 360 (2007) ("[A]ccess [to the Internet] is nearly as necessary as oxygen."). This article is not proposing such legislative action. However, there may be a way for copyright owners and online music service providers to reach reasonable agreements regarding terminating the accounts of repeat infringers.

24. For example, Jonathan Band has argued that:

A prohibition on tampering with copyright management information obviously benefits the copyright owner because it helps ensure that users receive accurate information about the terms and conditions governing the use of the work. At the same time, preserving the integrity of CMI also helps consumers by reducing counterfeiting and misrepresentations by middlemen concerning the contents of a digital envelope.

Jonathan Band, *The Digital Millennium Copyright Act: A Balanced Result*, 1999 STAN. TECH. L. REV. 12, 12 (1999).

created in the DMCA to protect “copyright management information” (“CMI”) such as watermarks, 17 U.S.C. § 1202, likely fails to prohibit circumvention of transactional watermarks for the purpose of removing information about the user (in this case, the purchaser) of a work. The statute explicitly excludes from the definition of CMI “any personally identifying information about a user of a work or of a copy” and prohibits the Register of Copyrights from prescribing any regulation that would amend the definition of CMI to include “any information concerning the user of a copyrighted work.”<sup>25</sup> These limitations on the protection of CMI were inserted in order to protect against a perceived threat to privacy.<sup>26</sup> However, recent studies indicate that transactional watermarks can be utilized without significantly undermining privacy interests.<sup>27</sup>

Shortly after the DMCA was signed into law, Senator Orrin Hatch, who was the Chairman of the Senate Judiciary Committee, wrote:

Because new technology is constantly changing, it is difficult to legislate intelligently. We want to be on the cutting edge, but not the *bleeding* edge of new technology. Therefore, some useful principles for judging what to do in the case of “high tech” issues are: (1) to refrain from legislation unless it is really necessary; (2) to address only those problems that have already identified themselves or that can be seen on the horizon; and (3) to make sure that technology is free to develop in any direction.<sup>28</sup>

Given that p2p networks were relatively obscure when the DMCA was enacted,<sup>29</sup> it made sense for Congress to limit the scope of the definition

---

Similarly, Senator Orrin Hatch has argued that CMI “should make it easier for the Internet’s electronic marketplace to function and self-regulate.” See Orrin G. Hatch, *Toward a Principled Approach to Copyright Legislation at the Turn of the Millennium*, 59 U. PITT. L. REV. 719, 756 (1998).

25. 17 U.S.C. § 1202 (2000).

26. See Staff of the H. Comm. on the Judiciary, 105th Cong., Section-by-Section Analysis of H.R. 2281 As Passed By the United States House of Representatives 34 (Comm. Print 1998) [hereinafter House Manager’s Report], reprinted in 46 J. Copyright Soc’y 635, 653 (1999); S. Rep. No. 105-190, at 36 (1998) (“To protect the privacy of users of copyrighted works, however, the Register of Copyrights may not include within the definition of CMI any information concerning *users* of copyrighted works.”) (emphasis in original).

27. See generally CTR. FOR DEMOCRACY AND TECH., PRIVACY PRINCIPLES FOR DIGITAL WATERMARKING (June 2, 2008), <http://cdt.org/publications/policyposts/2008/8> (proposing methods for protecting privacy).

28. Hatch, *supra* note 24, at 727 (emphasis in original).

29. See *In re Verizon Internet Servs., Inc.*, 240 F. Supp. 2d 24, 38 (D.D.C. 2003) (p2p was “not even a glimmer in anyone’s eye when the DMCA was enacted”). In fact, P2P technology could itself be seen, in part, as a reactionary response to the DMCA. See

of CMI in order to protect potential threats to consumer privacy. However, in the current context, where legislation to reduce infringement on p2p networks is necessary, and a problem relative to circumvention of transactional watermarks is clearly on the horizon, if not already present, Congress should consider amending the definition of CMI to protect transactional watermarks. Congress should also consider adding anti-trafficking provisions to section 1202 similar to those in section 1201 related to TPMs.<sup>30</sup> Such protections will help technologists develop watermarks that will benefit consumers who wish to engage in personal and transformative copying as well as copyright owners who seek to enforce their rights online.

Many scholars, technologists, and activists would probably object to this proposal given the widespread criticism of the DMCA's anti-circumvention provisions<sup>31</sup> and the very substantial literature on privacy and speech issues in the information economy.<sup>32</sup> Any suggestion that the

---

Joel R. Reidenberg, *The Rule of Intellectual Property Law in the Internet Economy*, 44 HOUS. L. REV. 1073, 1084 (2007) ("The rejection of the democratically chosen rule of law is well illustrated by the development of peer-to-peer ('P2P') technology.").

30. See e.g., 17 U.S.C. § 1201(a)(2).

No person shall manufacture, import, offer to the public, provide, or otherwise traffic in any technology, product, service, device, component, or part thereof, that—(A) is primarily designed or produced for the purpose of circumventing a technological measure that effectively controls access to a work protected under this title; (B) has only limited commercially significant purpose or use other than to circumvent a technological measure that effectively controls access to a work protected under this title; or (C) is marketed by that person or another acting in concert with that person with that person's knowledge for use in circumventing a technological measure that effectively controls access to a work protected under this title.

*Id.*

31. See, e.g., Neil W. Netanel, *Why Has Copyright Expanded? Analysis and Critique*, in 6 NEW DIRECTIONS IN COPYRIGHT LAW 3 (Fiona Macmillan ed., 2008); JESSICA LITMAN, DIGITAL COPYRIGHT: PROTECTING INTELLECTUAL PROPERTY ON THE INTERNET (2001); Brief Amicus Curiae of Intellectual Property Law Professors in Support of Defendants-Appellants, *Universal City Studios v. Corley*, 273 F.3d 429 (2d Cir. 2001) (No. 00-9185), 2001 WL 34105194; David Nimmer, *A Riff on Fair Use in the Digital Millennium Copyright Act*, 148 U. PA. L. REV. 673 (2000); David Nimmer, *Aus Der Neuen Welt*, 93 NW. U. L. REV. 195 (1998).

32. See, e.g., Julie E. Cohen, *Privacy, Visibility, Transparency, and Exposure*, 75 U. CHI. L. REV. 181 (2008); Julie E. Cohen, *Cyberspace as/and Space*, 107 COLUM. L. REV. 210 (2007); Sonia Katyal, *Privacy vs. Piracy*, 7 YALE J. LAW & TECH. 222 (2004) [hereinafter Katyal, *Privacy*]; Julie E. Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L.J. 575 (2003) [hereinafter Cohen, *DRM*]; Sonia K. Katyal, *The New Surveillance*, 54 CASE W. RES. L. REV. 297 (2003); Jerry Kang, *Information Privacy in Cyberspace Transactions*,

prohibitions of chapter 12 should be expanded should be approached with caution. However, when the benefits of watermarks are balanced against their potential detriments, a legislative proposal aimed at broadening the definition of CMI appears to best serve the purpose of copyright, which is to promote learning by encouraging creative output and the availability of creative works.<sup>33</sup> Thus, crafting such a proposal and pushing it through to passage deserves the focus of the copyright community.

Part II of this Article contains a brief overview of the scope of infringement on p2p networks and the inefficacy of attempts to decrease it, as well as a summary of the recent decisions of the major record labels to reduce their utilization of use-restrictive TPMs. Part III focuses on watermarking technology as a potential remedy to the problems discussed in Part II. Part IV analyzes the language and legislative history of section 1202 of the Copyright Act, and proposes amending the section to protect transactional watermarks. Finally, Part V discusses potential objections to amending section 1202, and concludes that the speech-related benefits of protecting transactional watermarks outweigh the potential threats to privacy and speech when viewed in light of copyright law's purpose and the online marketplace's current attitudes.

## II. DESPITE PERVASIVE P2P INFRINGEMENT, RECORD LABELS ARE REDUCING THEIR USE OF TPMS

In order to understand the need for greater protection for transactional watermarks, it is necessary to examine, briefly, the rise and continued pre-

---

50 STAN. L. REV. 1193 (1998); Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981 (1996) [hereinafter Cohen, *Right to Read*].

33. See Reidenberg, *supra* note 29, at 1077 ("Intellectual property law has a critical normative role. The allocation of rights to assure the balance of public values in the dissemination of knowledge, the incentive to create, and the freedom of expression are political choices."); Matt Williams, Note, *Making Encouraged Expression Imperceptible: The Family Movie Act of 2005 is Inconsistent with the Purpose of American Copyright*, 5 VA. SPORTS & ENT. L.J. 233, 235 (2006) ("[T]he purpose of American copyright law is to encourage and enable learning."); Paul Goldstein, *Copyright's Commons*, 29 COLUM. J.L. & ARTS 1, 10 (2005) ("The correct cause for advocacy is copyright itself . . . a system that takes as its balance wheel the need at once to promise authors protection for the product of their labors and to ensure them the freedom to borrow unprotected elements from the works of others."); Peters, *supra* note 2, at 722 ("Striking the balance between meeting consumer expectations and limiting harmful copying and distribution is the key to preserving copyright's standing in the eyes of the public."); Hatch, *supra* note 24, at 723 ("[C]opyright rights should be protected, unless it can be shown that the extent of protection is hampering creativity or the wide dissemination of works.").

valence of p2p networks. Although copyright owners have attempted to combat the large-scale infringement that occurs on these networks through successful litigation and competitive legal services, such as iTunes, musicians and the music industry still suffer myriad copyright violations. After maintaining for some time that it was necessary to utilize TPMs when distributing song files to consumers through online music services, the major record labels have concluded that it is in their interest to try an alternative route to protecting their content despite pervasive p2p infringement.<sup>34</sup>

### A. Pervasive p2p Infringement

Shawn Fanning launched the first prominent “file-sharing” network, Napster, in 1999.<sup>35</sup> Within months, over ten million people used Napster.<sup>36</sup> Not long after Napster’s arrival, other p2p networks and software, such as KaZaA, LimeWire, Bit Torrent, Aimster, Morpheus, Fast Track, Gnutella and Grokster, sprang up, each utilizing very different, amorphous delivery mechanisms.<sup>37</sup>

---

34. See *supra* note 11 and accompanying text.

35. LAWRENCE LESSIG, *FREE CULTURE* 67 (2004).

36. *Id.*

37. The Supreme Court in *Grokster* described the architecture of such p2p networks: [P]eer-to-peer networks [are] so called because users’ computers communicate directly with each other, not through central servers. The advantage of peer-to-peer networks over information networks of other types shows up in their substantial and growing popularity. Because they need no central computer server to mediate the exchange of information or files among users, the high-bandwidth communications capacity for a server may be dispensed with, and the need for costly server storage space is eliminated. Since copies of a file (particularly a popular one) are available on many users’ computers, file requests and retrievals may be faster than on other types of networks, and since file exchanges do not travel through a server, communications can take place between any computers that remain connected to the network without risk that a glitch in the server will disable the network in its entirety.

*Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913, 919-20 (2005). Peter Yu described it this way:

Unlike MP3.com and Napster, all of the new P2P technologies, such as Grokster, iMesh, KaZaA, and Morpheus, do not have centralized servers. Instead, they allow users to transfer files from one location to another while accommodating users’ needs to employ different hardware and software. As a result, enforcement is likely to become difficult. There will be no deep pocket to sue, no chokepoint to target, and no human face to blame.

Peter K. Yu, *P2P and the Future of Private Copying*, 76 U. COLO. L. REV. 653, 676 (2005).

Tim Wu has described p2p networks as “the most ambitious effort to undermine an existing legal system using computer code.”<sup>38</sup> Copyright owners, and especially the record labels, were to a large extent defenseless against the fast-rising threat that unrestricted file sharing represented to their business. Litigation, however, led to several legal victories, including at the Supreme Court, in which courts rejected the technological attempt to subvert copyright law.<sup>39</sup>

Nevertheless, p2p users continue to access p2p networks and to utilize them for infringement.<sup>40</sup> This has contributed to a rapid decline in music sales and very difficult times for record labels and artists.<sup>41</sup> Online sales of digital music are way up, but they have not yet come close to balancing out the impact of infringement.<sup>42</sup> For example, the record label EMI recently announced that it would have to reorganize itself in order to remain profitable.<sup>43</sup>

---

38. Timothy Wu, *When Code Isn't Law*, 89 VA. L. REV. 679, 683 (2003); see also Lior Jacob Strahilevitz, *Charismatic Code, Social Norms, and the Emergence of Cooperation on the File-Swapping Networks*, 89 VA. L. REV. 505, 535 (2003) (arguing that p2p networks “represent a particularly brazen and successful attack on intellectual property rights”); Reidenberg, *supra* note 29, at 1084–86 (analyzing Wu and Strahilevitz articles). Of course, p2p networks have valuable purposes other than infringement. See *Grokster*, 545 U.S. at 952 (Breyer, J. dissenting) (describing lawful uses of p2p software).

39. See, e.g., *Grokster*, 545 U.S. at 936–37 (2005) (holding that p2p services that intentionally induced copyright infringement were secondarily liable for infringement); *A&M Records, Inc. v. Napster, Inc.* 239 F.3d 1004, 1014 (9th Cir. 2001) (holding Napster p2p service liable for contributory infringement and vicarious infringement); *In re Aimster Copyright Litig.*, 334 F.3d 643, 645–46 (7th Cir. 2003) (holding Aimster p2p service liable for contributory infringement).

40. See Press Release, PRWeb, LimeWire Now Found on One-Third of All PCs Worldwide (Dec. 13, 2007), available at <http://www.prweb.com/releases/2007/12/prweb576418.htm>.

41. See STEPHEN E. SIWEK, INST. FOR POLICY INNOVATION, *THE TRUE COST OF SOUND RECORDING PIRACY TO THE U.S. ECONOMY* 14 (2007) (arguing that if one assumes that one out of five illegal downloads over p2p networks represents a lost sale at 99 cents, then U.S. companies lost over six billion dollars in 2005 alone); INT'L FED'N OF THE PHONOGRAPHIC INDUS., *IFPI DIGITAL MUSIC REPORT 2008: REVOLUTION, INNOVATION, RESPONSIBILITY* (2008), available at <http://www.ifpi.org/content/library/DMR2008.pdf>; *Napster*, 239 F.3d at 1018 (“The district court further found that both the market for audio CDs and market for online distribution are adversely affected by Napster's service.”).

42. Raphael G. Satter, *Digital Music Sales Up Worldwide*, ASSOCIATED PRESS, Jan. 24, 2008.

43. See Posting of Daniel Kreps to Rolling Stone Rock & Roll Daily Blog, EMI Chairman Confirms Cutbacks, Says Bands May be Sponsored Like Football Teams (Jan. 15, 2008, 5:25 EST), <http://www.rollingstone.com/rockdaily/index.php/2008/01/15/emi-chairman-confirms-cutbacks-says-bands-may-be-sponsored-like-football-teams/>.

Although the relationship between p2p usage and the current downturn in the music industry is often disputed by copyright skeptics,<sup>44</sup> even former advocates for unauthorized p2p distribution of copyrighted material have concluded that rampant infringement has damaged creative output. For example, Jaron Lanier, the musician and computer scientist who reportedly coined the term “virtual reality,” published an essay in the *New York Times* in 1999 entitled *Piracy Is Your Friend*.<sup>45</sup> There, he argued that “you can make more money [as a musician] in the new era of ‘free’ digital music.”<sup>46</sup> However, in November of 2007, Lanier, again in the *New York Times*, stated the following in an essay entitled *Pay Me For My Content*:

Internet idealists like me have long had an easy answer for creative types—like the striking screenwriters in Hollywood—who feel threatened by the unremunerative nature of our new Eden: stop whining and figure out how to join the party! That’s the line I spouted when I was part of the birthing celebrations for the Web. I even wrote a manifesto titled “Piracy Is Your Friend.” But I was wrong. We were all wrong.<sup>47</sup>

The sustained high levels of infringement despite the record labels’ and artists’ courtroom victories have led some governments to consider imposing policing obligations on Internet Service Providers (“ISPs”).<sup>48</sup> In France, President Sarkozy has proposed requiring ISPs to terminate the Internet access of repeat infringers.<sup>49</sup> Such proposals have received harsh criticism.<sup>50</sup> And yet, clearly something must be done.<sup>51</sup>

---

44. See, e.g., LESSIG, *supra* note 35, at 69.

45. Jaron Lanier, *Piracy Is Your Friend*, reprinted in Neil Strauss, *Music: A Chance to Break the Pop Stranglehold*, N.Y. TIMES, May 9, 1999, § 2 (Arts & Leisure), at 1.

46. *Id.*

47. Jaron Lanier, *Pay Me For My Content*, N.Y. TIMES, Nov. 20, 2007, at A23.

48. See, e.g., Dep’t for Bus. Enter. & Regulatory Reform, Consultation on Legislative Options to Address Illicit Peer-to-Peer (P2P) File-Sharing 5–6 (2008), available at <http://www.berr.gov.uk/files/file47139.pdf> (proposing a “co-regulatory approach” and an alternative regulatory approach that may include requirements for ISPs to disclose the Internet protocol addresses of alleged infringers); Dep’t for Culture, Media & Sport, Creative Britain: New Talents for the New Economy 10 (2008), available at <http://www.culture.gov.uk/images/publications/CEPFeb2008.pdf>.

49. Eric Pfanner, Effort to Combat Internet Piracy Gains Strength in France, N.Y. TIMES, Dec. 3, 2007, at C8.

50. See, e.g., Siy, *supra* note 23.

51. See Billy Bragg, *The Royalty Scam*, N.Y. TIMES, Mar. 22, 2008, at A13 (“If young musicians are to have a chance of enjoying a fruitful career, then we need to establish the principle of artists’ rights throughout the Internet—and we need to do it now.”).

## B. Record Labels Decide to Offer “TPM-Free” Music Online

Record labels have endured persistent criticism for using TPMs to restrict the unauthorized copying and distribution of songs purchased online.<sup>52</sup> This criticism became highly publicized when Steve Jobs, the CEO of Apple, published an essay called *Thoughts On Music* in early 2007.<sup>53</sup> The essay, which made public Jobs’ frustration with negotiating licenses for content sold on Apple’s iTunes platform, repeated a mantra that had been previously put forward, first by Microsoft employees in an article called *The Darknet and the Future of Content Distribution*,<sup>54</sup> and then by technology advocate Fred von Lohmann:<sup>55</sup> “There is no theory of protecting content other than keeping secrets. . . . The problem, of course, is that there are many smart people in the world, some with a lot of time on their hands, who love to discover such secrets and publish a way for everyone to get free (and stolen) music.”<sup>56</sup> In other words, Jobs argued that TPMs would inevitably be hacked, therefore failing to provide record labels with any protection against p2p infringement, and at the same time, would frustrate customers who would not be able to engage in personal and transformative copying.

Jobs asked the record labels to “imagine a world where every online store sells [TPM]-free music encoded in open licensable formats [such that] any player can play music purchased from any store, and any store can sell music which is playable on all players.”<sup>57</sup> By April of 2007, EMI announced that it would allow iTunes to sell TPM-free song files.<sup>58</sup> Then, in August, Universal Music Group made a similar announcement, although it decided not to offer TPM-free songs through iTunes and to instead do so through services such as Amazon.com.<sup>59</sup> In December of 2007 and January of 2008, respectively, Warner and Sony BMG also decided to

---

52. See, e.g., Mulligan & Perzanowski, *supra* note 8.

53. Steve Jobs, *Thoughts On Music*, Feb. 6, 2007, <http://www.apple.com/hotnews/thoughtsonmusic/>.

54. Peter Biddle et al., *The Darknet and the Future of Content Distribution* (2002), presented at the 2002 ACM Workshop on Digital Rights Management, available at <http://crypto.stanford.edu/DRM2002/darknet5.doc>.

55. Von Lohmann, *supra* note 6.

56. Jobs, *supra* note 53.

57. *Id.*

58. Press Release, EMI Group, EMI Music Launches DRM-Free Superior Sound Quality Downloads Across Its Entire Digital Repertoire (Apr. 2, 2007), available at <http://www.emigroup.com/Press/2007/press18.htm> [hereinafter EMI Group].

59. Leeds, *supra* note 11.

sell TPM-free songs online.<sup>60</sup> Finally, in January of 2009, Apple ended the debate by announcing that its iTunes store will sell only TPM-free songs.<sup>61</sup>

When EMI first announced its decision, it stated: “EMI is releasing the [TPM-free] downloads in response to consumer demand for high fidelity digital music for use on home music systems, mobile phones and digital music players. EMI’s new [TPM]-free products will enable full interoperability of digital music across all devices and platforms.”<sup>62</sup> An Apple press release reiterated this goal: “With [TPM]-free music from the EMI catalog, iTunes customers will have the ability to download tracks from their favorite EMI artists without any usage restrictions that limit the types of devices or number of computers that purchased songs can be played on.”<sup>63</sup>

Such responsiveness to consumer demands will hopefully help to improve consumers’ opinions of record labels and copyright law. Prominent members of the copyright community have recently encouraged copyright owners to make efforts in this regard,<sup>64</sup> and the record labels appear to be doing so despite, or perhaps in part because of, the losses they continue to suffer from infringement on p2p networks.

TPM-free downloads will enable consumers to engage in personal and transformative copying, and may result in increased sales.<sup>65</sup> However, record labels have reportedly been watermarking the TPM-free downloads in an effort to keep tabs on the number that are made available on p2p networks.<sup>66</sup> Thus, in the current music marketplace, protecting watermarks against circumvention may be even more important than protecting TPMs.

---

60. McCarthy, *supra* note 11; Posting of Matt Rosoff to Digital Noise, DRM Deathwatch: Warner on Amazon (Dec. 27, 2007, 10:05 PST), [http://blogs.cnet.com/-8301-13526\\_1-9837994-27.html](http://blogs.cnet.com/-8301-13526_1-9837994-27.html).

61. See Brad Stone, *Want to Copy iTunes Music? Go Ahead, Apple Says*, N.Y. TIMES, Jan. 7, 2009. There was not time to make substantial revisions to the article to reflect this breaking news.

62. EMI Group, *supra* note 58.

63. Press Release, Apple, Inc., Apple Unveils Higher Quality DRM-Free Music on the iTunes Store (Apr. 2, 2007), *available at* <http://www.apple.com/pr/library/2007/04/02itunes.html>.

64. See, e.g., Peters, *supra* note 2, at 717 (proposing ways to “help copyright retain its good standing in the eyes of the public”); Goldstein, *supra* note 33, at 2–3 (discussing importance of public education regarding copyright’s public benefits).

65. But see Bill Rosenblatt, *Is EMI’s DRM-Free Strategy Working?*, DRM WATCH, Aug. 8, 2007, <http://www.drmwatch.com/ocr/article.php/3693316> (suggesting that there does not appear to be evidence of such an increase).

66. Rosenblatt, *supra* note 14.

### III. WATERMARKING CAN BENEFIT COPYRIGHT OWNERS AND CONSUMERS

A digital watermark consists of information inserted into a digital file that may identify or explain the copyright owner of the content of the file, the license applicable to the file, the person who purchased the file, and/or other facts related to use of the file.<sup>67</sup> Most digital watermarks are imperceptible by human eyes and ears.<sup>68</sup> In order to “view a watermark, an investigator needs a special program or device (i.e., a ‘detector’) that can extract the watermark data” and reveal the information.<sup>69</sup>

Watermarking comes in at least two types: transactional (also known as media serialization) and non-transactional.<sup>70</sup> Transactional watermarks contain the identity of the device or user that downloads or purchases the content, in addition to information about the originating service and copyright owner. When the content is found somewhere that it isn’t supposed to be (e.g., residing on a p2p network or YouTube), content owners can trace the content back to its original purchaser or downloader to determine who is responsible for the infringing distribution.<sup>71</sup> Non-transactional watermarks do not contain information about the individual user or user’s device, but instead simply contain information about the copyright owner, the terms of use, and the service from which the content was purchased.<sup>72</sup> Universal Music Group is reportedly using non-transactional watermarking technology in connection with its offerings, and iTunes is reportedly using transactional watermarking on its TPM-free songs.<sup>73</sup>

---

67. See *Stealthy Audio Watermarking*, U.S. Patent No. 7,266,697 col.2 l.65 (filed May 3, 2004) (“ In general, a ‘digital watermark’ is a pattern of bits inserted into a digital image, audio, or video file that identifies the file’s copyright information (author, rights, etc.). The name comes from the faintly visible watermarks imprinted on stationary that identify the manufacturer of the stationary.”).

68. *Id.* at col.3 l.6–8.

69. *Id.* at col.3 l.18–20.

70. See Rosenblatt, *Accelerating*, *supra* note 14.

71. See Kravets, *supra* note 14 (“Watermarking offers copyright protection by letting a company track music that finds its way to illegal peer-to-peer networks. At its most precise, a watermark could encode a unique serial number that a music company could match to the original purchaser.”); ’697 Patent, at col.3 l.50 (detailing watermarks that do not degrade when content is transferred into different formats).

72. Rosenblatt, *Accelerating*, *supra* note 14.

73. See Ken Fisher, *Apple Hides Account Info In DRM Free Music, Too*, ARS TECHNICA (May 30, 2007), <http://arstechnica.com/news.ars/post/20070530-apple-hides-account-info-in-drm-free-music-too.html>.

Watermarks generally do not impose restrictions on how consumers can use copies of works.<sup>74</sup> They nevertheless help copyright owners protect their rights. For example, Random House Audio Publishing Group recently announced that it watermarked eBooks sold in the fall of 2007 without TPMs through the online service eMusic.com.<sup>75</sup> The watermarks enabled Random House to monitor p2p traffic to determine whether selling eBooks without TPMs lead to increased infringement of its works. The test results indicated that the watermarked copies were not being distributed on p2p networks. In fact, copies of eBooks that contained circumvented TPMs were available on p2p networks whereas Random House was unable to find “a single instance of the eMusic watermarked titles being distributed illegally.”<sup>76</sup> Random House decided, based on the test, to continue selling TPM-free watermarked eBooks through multiple online platforms.

Allowing copyright owners to test the benefits and security of new business models is only one benefit of watermarks. Record labels have faced difficult evidentiary issues in their lawsuits against individual p2p infringers.<sup>77</sup> Defendants and their supporters often argue, *inter alia*, that their Internet access accounts have been used by others to make song files available on p2p networks.<sup>78</sup> Proving that a particular defendant is respon-

---

74. See ACTIVATED CONTENT, THE FUTURE OF WATERMARKING, *available at* [http://www.activatedcontent.com/whitepapers/ActivatedContent\\_FutureOfWatermarking.pdf](http://www.activatedcontent.com/whitepapers/ActivatedContent_FutureOfWatermarking.pdf) (last visited Oct. 20, 2008) (discussing survival or robust transfers).

75. Letter from Madeline McIntosh, Senior Vice President & Publisher, Random House Audio Publ'g Group, to Publ'g Partners (Feb. 21, 2008) [hereinafter McIntosh Letter] (on file with beyondthebookcast.com). It is unclear whether these watermarks were transactional watermarks or not.

76. *Id.*

77. See, e.g., *Virgin Records America, Inc. v. Thompson*, 512 F.3d 724, 725 (5th Cir. 2008) (detailing how defendant alleged that his daughter used his Internet access account for p2p infringement); *Capitol Records, Inc. v. Foster*, 86 U.S.P.Q.2d (BNA) 1203, 1206 (W.D. Okla. 2007) (discussing defendant's allegation that her Internet access account was being used for p2p infringement by a member of the household other than the original defendant); see also MICHAEL PIATEK ET AL., UNIV. OF WASH., DEP'T OF COMPUTER SCIENCE & ENG'G, CHALLENGES AND DIRECTIONS FOR MONITORING P2P FILE SHARING NETWORKS—OR—WHY MY PRINTER RECEIVED A DMCA TAKEDOWN NOTICE 1 (2008), *available at* [http://dmca.cs.washington.edu/uwcse\\_dmca\\_tr.pdf](http://dmca.cs.washington.edu/uwcse_dmca_tr.pdf) (describing methods used to “frame” innocent computers for copyright infringement); Robert Kasunic, *Making Circumstantial Proof of Distribution Available*, 18 *FORDHAM INTELL. PROP. MEDIA & ENT. L.J.* 1145, 1156 (2008) (discussing possibility that files are unknowingly made available for copying through p2p networks).

78. See, e.g., Brief of Amici Curiae American Ass'n of Law Libraries et al. in Support of Defendant Debbie Foster's Motion for Attorneys' Fees, *Capital Records, Inc. v.*

sible for the infringement observed by record label investigators can be problematic. Transactional watermarks could help to reduce these evidentiary difficulties by more clearly connecting individual copies with individual users. An investigator searching a p2p network for infringing copies could locate a downloadable copy of a popular song in a person's share folder, download that copy, and review the watermark information to determine who originally purchased the copy from which the infringing copy was made. This could be done in a variety of ways, including by embedding a serial number unique to users of online music services rather than including names or other personal information.<sup>79</sup>

Making the labels better able to detect an individual's posting of song files on p2p networks may even reduce the need for litigation. If copyright owners could clearly identify individual infringers, they could seek agreements with online services that would result in account terminations for repeat infringers.<sup>80</sup> A realistic threat of loss of access to such services could prove to be a powerful deterrent against infringement.<sup>81</sup>

Unfortunately, watermarks, like TPMs, are not invulnerable to attack. Although developers are producing watermarks that persist through a variety of format conversions, such as moving a song from a portable device

---

Foster, No 04-1569-W, slip op. at 6 (W.D. Okla. 2006). Amici for the defendant in *Capital v. Foster* argued:

The RIAA's investigators sign into file sharing networks hoping to identify users who are sharing particular songs. However, users on P2P networks are difficult to identify. Each user has a 'screenname' that represents her presence on the network. This screenname is usually some kind of vague or anonymous nickname, e.g. 'musicfan21.' Moreover, on many systems, multiple users can have the same screenname, further obfuscating association with a particular identity. Thus, neither that screenname nor anything else available from the P2P network alone can tie a virtual-world user directly to a specific real-world person.

*Id.* at 9.

79. Kravets, *supra* note 14. Of course, watermarks would not eliminate all evidentiary difficulties. For example, someone may purchase a song embedded with a watermark related to his or her own account. Someone else in the household may then use the same computer to distribute the song on a p2p network, unbeknownst to the purchaser. Nevertheless, watermarks would reduce much of the confusion surrounding p2p screen names.

80. Such agreements represent an alternative to the legislative proposals pending abroad that would result in loss of Internet access for repeat infringers. *See* note 23 *supra*.

81. *Cf.* S. REP. NO. 105-90, at 52 (1998) ("[T]hose who repeatedly or flagrantly abuse their access to the Internet through disrespect for the intellectual property rights of others should know that there is a realistic threat of losing that access.").

to a desktop, and that appear to be very difficult to circumvent,<sup>82</sup> recent history indicates that “smart people . . . with a lot of time on their hands”<sup>83</sup> can thwart nearly any technological effort to protect works.<sup>84</sup>

Copyright owners will be hesitant to continue the trend of TPM-free songs if watermarks prove ineffective.<sup>85</sup> A reversal of this trend would negatively impact consumers who want to engage in personal and transformative copying. Thus, in order to benefit consumers and decrease infringement, watermarks need adequate legal protection.

#### IV. INADEQUATE LEGAL PROTECTION FOR WATERMARKS

Transactional watermarks that include information about the users of works could help copyright owners fight infringement on p2p networks. But section 1202 of the Copyright Act, which prohibits altering or removing CMI, likely fails to protect transactional watermarks against circumvention. In addition, section 1202 does not contain any provisions prohibiting trafficking in services or devices designed or marketed to circumvent watermarks of any kind. Congress should consider remedying these inadequacies.

##### A. The Definition of CMI: Section 1202’s History and Text

Although the DMCA became law in 1998, its beginnings lie five years earlier. In 1993, President Clinton created the Information Infrastructure Task Force (“IITF”) within the Department of Commerce to craft policies that would assist the development of the national information infrastructure (“NII”, i.e., Internet).<sup>86</sup> Two years later, The Working Group on Intel-

---

82. *See, e.g.*, ’697 Patent.

83. Jobs, *supra* note 53.

84. *Cf.* Brief Amici Curiae of Computer Science Professors Harold Abelson et al. Suggesting Affirmance of the Judgment, 2005 WL 497760, at \*14, *Metro-Goldwyn-Mayer Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005) (No. 04-480) (discussing methods of defeating filtering technologies).

85. In addition, new advertising-based business models that appear to be attracting support from content owners are dependent on watermarks functioning properly. *See ACTIVATED CONTENT*, *supra* note 74. These business models will be unable to develop if watermarks are unreliable. Efforts to provide consumers with additional features by using watermarks are also being considered. *See* Antony Bruno, *High Watermark: New DRM Technology Could Flood Consumers with Bonus Features*, BILLBOARD, Jan. 19, 2008, at 14. Thus, copyright owners and consumers will benefit in many ways from adequate legal protection.

86. *IQ Group, Ltd. v. Wiesner Publ’g, LLC*, 409 F. Supp. 2d 587, 594 (D.N.J. 2006).

lectual Property Rights, a division within the IITF, released a report describing the likely impact of the NII on intellectual property rights and recommending appropriate changes to U.S. law.<sup>87</sup> This report, often referred to as the “White Paper,” proposed draft legislation to create legal protections for “copyright management information” and “copyright protection systems.”<sup>88</sup> The draft legislation would have prohibited dealing in devices or services that enabled circumvention of any “mechanism or system which prevents or inhibits the violation of any of the exclusive rights [of the copyright owner] under section 106.”<sup>89</sup> It also would have prohibited the knowing removal or alteration of copyright management information. The draft legislation defined CMI as follows: “‘Copyright management information’ means the name and other identifying information of the author of a work, the name and other identifying information of the copyright owner, terms and conditions for uses of the work, and such other information as the Register of Copyrights may prescribe by regulation.”<sup>90</sup>

Subsequent to the release of the White Paper, the National Information Infrastructure Copyright Protection Act, which contained the White Paper’s proposals verbatim, was introduced in Congress.<sup>91</sup> The legislation did not pass. However, the international community was considering addressing the developing Internet at the same time, and in 1996, the World Intellectual Property Organization (“WIPO”) Copyright Treaty (“WCT”) and the WIPO Performances and Phonograms Treaty (“WPPT”) were agreed upon.<sup>92</sup>

These treaties “enhance[d] the exploitation and enforcement of exclusive rights in the digital environment”<sup>93</sup> in response to “the questions raised by new economic, social, cultural and technological developments.”<sup>94</sup> In other words, the treaties are the product of an international “recognition that works made available in digital formats may be especial-

---

87. Bruce A. Lehman, THE REPORT OF THE WORKING GROUP ON INTELLECTUAL PROPERTY RIGHTS, INTELLECTUAL PROP. & THE NAT’L INFO. INFRASTRUCTURE (1995).

88. *Id.* at 233–35.

89. *Id.* at 230.

90. *Id.* at 7.

91. See H.R. REP. NO. 104-879 (1997) (noting the introduction of H.R. 2441).

92. World Intellectual Prop. Org. [WIPO] Copyright Treaty, Dec. 20, 1996, S. Treaty Doc. No. 105–17, 36 I.L.M. 65 (1997); WIPO Performances and Phonograms Treaty, Dec. 20, 1996, S. Treaty Doc. No. 105–17, 36 I.L.M. 76 (1997).

93. SAM RICKETSON & JANE C. GINSBURG, INTERNATIONAL COPYRIGHT AND NEIGHBOURING RIGHTS: THE BERNE CONVENTION AND BEYOND VOL. II (2d ed. 2006).

94. WIPO Copyright Treaty, *supra* note 92, at 4 (reciting preamble); see also WIPO, GUIDE TO THE COPYRIGHT AND RELATED RIGHTS TREATIES ADMINISTERED BY WIPO AND GLOSSARY OF RELATED RIGHTS TERMS (2003).

ly vulnerable to unauthorized copying and redistribution; unless the digital file can be secured against these acts, its susceptibility to unauthorized recirculation may discourage authors from making it digitally available to the general public.”<sup>95</sup>

The WIPO treaties require signatories to provide “adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection with the exercise of their rights”<sup>96</sup> as well as “adequate and effective legal remedies against any person knowingly” removing or altering rights management information.<sup>97</sup> Although debate regarding Article 11 of the WCT, which covers TPMs, was heated, the provision covering rights management information “did not raise the kind of controversy raised by those of Article 11.”<sup>98</sup> The WIPO treaties defined “rights management information” to mean:

information which identifies the work, the author of the work, the owner of any right in the work, or information about the terms and conditions of use of the work, and any numbers or codes that represent such information, when any of these items of information is attached to a copy of a work or appears in connection with the communication of a work to the public.<sup>99</sup>

After the United States signed the WIPO treaties, legislation was introduced in the Senate to implement them.<sup>100</sup> The legislation used the term “copyright management information” rather than “rights management information” and included a slightly different definition:

---

95. RICKETSON & GINSBURG, *supra* note 93, at 966.

96. WIPO Copyright Treaty, *supra* note 92, at 10 (quoting Article 11).

97. *Id.* at 11 (quoting Article 12). Dr. Mihaly Ficsor has explained the difference in wording between Article 11 and Article 12 of the WIPO Copyright Treaty:

It is to be noted that while Article 11 speaks about the obligation to provide “adequate legal protection and effective legal remedies” this Article [12] “only” obliges Contracting Parties to provide “adequate and effective legal remedies.” It seems, however, that the disparity between the two texts is the result of a mere drafting inadvertence, and that the basic nature of the obligations of the Contracting Parties is practically the same under the two provisions.

Mihaly Ficsor, *The Law of Copyright and the Internet* 564 (2002).

98. *Id.*

99. WIPO Copyright Treaty, *supra* note 92, at art. 12.

100. WIPO Copyright and Performances and Phonograms Treaty Implementation Act, S. 1121, 105th Cong. (1997); *see also* 105 CONG. REC. S8582 (daily ed. July, 31 1997) (statement of Sen. Hatch) (introducing the bill).

the following information conveyed in connection with copies or phonorecords of a work or performances or displays of a work, including in digital form—

- (1) the title and other information identifying the work, including the information set forth on a notice of copyright;
- (2) the name of, and other identifying information about, the author of a work;
- (3) the name of, and other identifying information about, the copyright owner of the work, including the information set forth in a notice of copyright;
- (4) terms and conditions for use of the work;
- (5) identifying numbers of symbols referring to such information or links to such information; or
- (6) such other information as the Register of Copyrights may prescribe by regulation, except that the Register of Copyrights may not require the provision of any information concerning the user of a copyrighted work.<sup>101</sup>

Subparagraph 6 of the definition allowed the Register to broaden the definition through regulations, if needed, but limited her ability to do so where “any information concerning a user of a copyrighted work” was involved.<sup>102</sup> Although the floor statements made related to the introduction of the legislation and the early legislative reports did not indicate why the Register’s ability to proscribe regulations was limited in this manner, language from the White Paper expressed a concern that protecting CMI could “unduly burden use of the work by consumers or compromise their privacy.”<sup>103</sup>

Protecting citizens’ privacy online was a major focus of the Clinton Administration during development of the NII agenda. In fact, the National Telecommunications and Information Administration within the Department of Commerce released a report called *Privacy and the NII: Safeguarding Telecommunications-Related Personal Information* one month

---

101. CONG. REC. S. 1121, at section 3.

102. *Id.*

103. Lehman, *supra* note 87, at 191.

after the White Paper was released.<sup>104</sup> This desire to protect privacy is also evident in early versions of the DMCA and resulted in specific privacy exceptions to section 1201.<sup>105</sup> It appears that this concern also led to the limitation on the Register's power.

This limitation remained part of the definition of CMI throughout the legislative process that resulted in passage of the DMCA. In addition, the House and Senate Conferees added an additional privacy related limitation into the definition before passage. This limitation explicitly excludes "any personally identifiable information about a user of a work or of a copy" from the definition.<sup>106</sup> Senator John Ashcroft, who introduced this change, explained that he intended the carve out to "help preserve the critical balance that we must maintain between the interests of copyright owners and the privacy interests of information users."<sup>107</sup>

The legislative history is, for the most part, very unclear regarding the exact meaning of the privacy carve out. The clearest statement is found in the House Manager's Report:

It also should be noted that the definition of "copyright management information" does not encompass, nor is it intended to encompass, tracking or usage information relating to the identity of users of works. It would be inconsistent with the purpose and construction of this bill and contrary to the protection of privacy to include tracking and usage information within the definition of CMI.<sup>108</sup>

Based on this language, it appears clear that Congress intended the carve out to apply broadly. Although the language of the statute could be

---

104. NATIONAL TELECOMMUNICATIONS AND INFORMATION ADMINISTRATION, DEPARTMENT OF COMMERCE, PRIVACY AND THE NII: SAFEGUARDING TELECOMMUNICATIONS-RELATED PERSONAL INFORMATION, *available at* <http://www.ntia.doc.gov/ntiahome/privwhitepaper.html>.

105. For example, the DMCA contained an exception allowing circumvention of access controls where personally identifiable information is involved, 17 U.S.C. § 1201(i) (2002), as well as an entire section stating that:

[n]othing in this chapter abrogates, diminishes, or weakens the provisions of, nor provides any defense or element of mitigation in a criminal prosecution or civil action under, any Federal or State law that prevents the violation of the privacy of an individual in connection with the individual's use of the Internet.

17 U.S.C. § 1205 (2000).

106. H.R. REP. NO. 105-796, at 14 (1998).

107. 144 CONG. REC. S. 11888, 105th Cong. (daily ed. Oct. 2, 1998) (statement of Sen. Ashcroft).

108. House Manager's Report, *supra* note 26, at 653.

interpreted to only apply to specific types of information about users, i.e., “personally identifiable information”,<sup>109</sup> the Manager’s Report says that “tracking or usage information *relating* to the identity of users of works” is excluded from the definition of CMI.<sup>110</sup> Under this interpretation, even serial numbers that correspond to a specific user’s music service account but otherwise contain no information about the user fall outside of the meaning of CMI. The broad limitation on the Register’s power which prevents her from including “*any* information concerning the user of a copyrighted work” also seems to back up this interpretation.<sup>111</sup>

Thus far, no court has considered the scope of the privacy carve out. But the trend in cases involving section 1202 claims has been to interpret the definition of CMI narrowly.<sup>112</sup> In fact, one court, in *IQ Group v.*

---

109. The term “personally identifiable information” is not defined in section 1202. However, it is defined elsewhere in the U.S. Code. For example, 11 U.S.C. § 101 (2000) defines it as follows:

The term “personally identifiable information” means—

(A) if provided by an individual to the debtor in connection with obtaining a product or a service from the debtor primarily for personal, family, or household purposes—

(i) the first name (or initial) and last name of such individual, whether given at birth or time of adoption, or resulting from a lawful change of name;

(ii) the geographical address of a physical place of residence of such individual;

(iii) an electronic address (including an e-mail address) of such individual;

(iv) a telephone number dedicated to contacting such individual at such physical place of residence;

(v) a social security account number issued to such individual;

or

(vi) the account number of a credit card issued to such individual; or

(B) if identified in connection with 1 or more of the items of information specified in subparagraph (A)—

(i) a birth date, the number of a certificate of birth or adoption, or a place of birth; or

(ii) any other information concerning an identified individual that, if disclosed, will result in contacting or identifying such individual physically or electronically.

*Id.* at 41A (A)–(B).

110. House Manager’s Report, *supra* note 26, at 653 (emphasis added).

111. 17 U.S.C. § 1202(c)(8) (2000) (emphasis added).

112. *See, e.g.*, *Thron v. HarperCollins Publishers, Inc.*, 64 U.S.P.Q.2d (BNA) 1221, 1222 (S.D.N.Y. 2002) (granting summary judgment for defendant on section 1202 claim because plaintiff’s registration was invalid); *Kelly v. Arriba Soft Corp.*, 77 F. Supp. 2d 1116, 1122 (C.D. Cal. 1999), *aff’d in part, rev’d in part*, 336 F.3d 811 (9th Cir. 2003)

*Weisner Publishing*, interpreted section 1202 only to apply to “copyright management performed by the technological measures of automated systems.”<sup>113</sup> Although that court acknowledged that Congress intended section 1202 to protect digital watermarks,<sup>114</sup> it went on to imply that watermarks are only protected if they are mingled with a TPM protected by section 1201: “Chapter 12, as a whole, appears to protect automated systems which protect and manage copyrights. The systems themselves are protected by § 1201 and the copyright information used in the functioning of the systems is protected in § 1202.”<sup>115</sup>

If this opinion is read expansively to exclude any watermark that is not coupled with a TPM from protection, then none of the watermarks used by the recording industry in their TPM-free online offerings are covered by section 1202. This is an entirely unsatisfactory result that does not benefit consumers or copyright owners.<sup>116</sup> Fortunately, *IQ Group* is a district court opinion that has not thus far discouraged copyright owners from uti-

---

(holding that CMI must be on the work itself rather than alongside the work); *Schiffer Publ'g, Ltd. v. Chronicle Books, LLC*, 73 U.S.P.Q.2d (BNA) 1090, 1102 (E.D. Pa. 2004) (finding that CMI must be removed “from the ‘body’ of, or area around, plaintiff’s work itself”); *Textile Secrets Int’l, Inc. v. Ya-Ya Brand Inc.*, 524 F. Supp. 2d 1184 (C.D. Cal. 2007) (holding that CMI must be digital, so fabric containing copyright information unprotected). *But see* *Med. Broad. Co. v. Flaiz*, No. 02-8554, 2003 U.S. Dist. LEXIS 22185, at \*8 (E.D. Pa. Nov. 25, 2003) (finding that no registration was required for a section 1202-related claim); *McClatchey v. Associated Press*, 82 U.S.P.Q.2d (BNA) 1190, 1195 (W.D. Pa. 2007) (holding that CMI need not be digital); *see also* Rebecca Tushnet, *Naming Rights: Attribution and Law*, 2007 UTAH L. REV. 789, 790 (2007) (endorsing view that section 1202 has proven ineffective); Greg Lastowka, *Digital Attribution: Copyright and the Right to Credit*, 87 B.U. L. REV. 41, 70–73 (2007) (discussing cases).

113. *IQ Group, Ltd. v. Wiesner Publ'g, LLC*, 409 F. Supp. 2d 587, 597 (D.N.J. 2006).

114. *Id.* at 596.

115. *Id.* at 597.

116. Congress and commentators recognized from the start that CMI could benefit consumers by informing them of valuable information regarding copyright owners. *See, e.g.*, Band, *supra* note 24 at 12. Recently, significant problems related to licensing of works where copyright owners are unknown have received public attention during Copyright Office and Congressional consideration of “orphan works.” *See* REGISTER OF COPYRIGHTS, REPORT ON ORPHAN WORKS (2006), available at <http://www.copyright.gov/orphan/orphan-report-full.pdf>; *see also* Eric J. Schwartz & Matt Williams, *Access to Orphan Works: Copyright Law, Preservation, and Politics*, 46 CINEMA J. 139 (2007). CMI can help resolve these problems. *See* Coree Thompson, Note, *Orphan Works, U.S. Copyright Law, and International Treaties: Reconciling Differences to Create a Brighter Future for Orphans Everywhere*, 23 ARIZ. J. INT’L & COMP. L. 787, 808 (2006) (“With fairly strong civil and criminal penalties backing the DMCA’s copyright management information provisions, individuals may find that much-desired ownership information will become more readily provided and more reliable.”).

lizing watermarks without TPMs. But if subsequent courts adopt an expansive reading of the opinion, copyright owners may be forced to use TPMs in order to obtain protection for their watermarks.<sup>117</sup>

### **B. The Absence of Prohibitions Against Circumvention Devices and Services**

Another inadequacy of the prohibitions of section 1202 that may discourage copyright owners from abandoning TPMs in some contexts is the lack of provisions prohibiting trafficking in devices and services that enable consumers to strip CMI from copies of works. Section 1201 prohibits circumvention of access controls by consumers as well as dealing in circumvention devices or services that enable circumvention of access controls or copy controls.<sup>118</sup>

In contrast, section 1202 only prohibits removing or altering CMI. This results in rather ineffective protection because it is very difficult to know who stripped a copy of CMI once the CMI is removed. It is much more practical to prohibit dealing in devices (such as software) and services designed to remove or alter CMI.

Judge Lewis Kaplan of the U.S. District Court for the Southern District of New York explained the problem in relation to TPMs as follows:

Dissemination of decryption technology over the Internet is analogous to spreading an infectious disease. Every downloaded copy of the technology is itself capable both of performing acts of infringement and of yielding other copies of the decryption technology just as every person carrying an infectious pathogen may get ill and may also infect others with whom the person comes into contact. Once a decryption program is disseminated, finding its source cannot stop its use for infringement. And given the speed and size of the Internet, dissemination is tantamount to use.<sup>119</sup>

---

117. A recent opinion explicitly followed *IQ Group*, but included language that appears to retreat from the more expansive reading of *IQ Group*. See *Textile Secrets*, 524 F. Supp. 2d at 1201–02 (“[T]he Court nevertheless cannot find that the provision was intended to apply to circumstances that have no relation to the Internet, electronic commerce, automated copyright protections or management systems, public registers, or other technological measures or processes as contemplated in the DMCA as a whole.”).

118. 17 U.S.C. § 1201 (2000).

119. Lewis A. Kaplan, *Copyright in the Digital Age: The 2001 Donald C. Brace Memorial Lecture Delivered at Fordham University School of Law on Nov. 12, 2001*, 49 J. COPYRIGHT SOC’Y U.S.A. 1, 15–16 (2001).

Professor Neal Netanel has also concluded that “the copyright industries accurately contend, if technological controls are to have any chance of being broadly effective, the law must prohibit the dissemination of software and other devices capable of skirting DRM technology.”<sup>120</sup> Similarly, as long as devices and services aimed at removing transactional watermarks are allowed to proliferate, such watermarks will fail to be “broadly effective” infringement deterrents.<sup>121</sup>

The legislative history fails to illuminate why Congress did not choose to prohibit devices and services aimed at enabling violations of section 1202. The White Paper did not propose doing so without explanation. However, failing to prohibit such devices and services arguably places the United States in conflict with its obligations under the WIPO Treaties.

The United States, and other nations, chose to prohibit devices and services that enable circumvention of TPMs because doing otherwise would have violated our treaty obligations.<sup>122</sup> As Dr. Mihaly Ficsor, the Assistant Director General of WIPO at the time the treaties were drafted and signed, has said:

It should be taken into account that, in general, the acts of circumvention of technological protection measures will be carried out by individuals in private homes or offices, where enforce-

---

120. Neil Weinstock Netanel, *Impose a Noncommercial Use Levy to Allow Free Peer-to-Peer File Sharing*, 17 HARV. J.L. & TECH. 1, 10 (2003).

121. Marybeth Peters, the Register of Copyrights, has explained that:

For copyright owners, technological protection can never be more than half of the answer. Technology can always be matched and surpassed by technology; the most ingenious anti-copying system will eventually be circumvented by the development of ingenious anti-anti-copying systems. In the area of computer programs, for example, every program developed to prevent unauthorized copying has ultimately been defeated by a program that instructs the computer to ignore the first program. If technology is to provide a solution to the enforcement challenge, it is imperative to devise some method to put an end to the cycle.

Marybeth Peters, *The Spring 1996 Horace S. Manges Lecture—The National Information Infrastructure: A Copyright Office Perspective*, 20 COLUM.-VLA J.L. & ARTS 341, 343 (1996).

122. See H.R. REP. NO. 105-551 (1998). Congress was mindful of the United States’s obligations under international law to prevent circumvention:

There will be those who will try to profit from the works of others by decoding the encrypted codes protecting copyrighted works, or engaging in the business of providing devices or services to enable others to do so. A new “Section 1201” to the Copyright Act is required by both WIPO Treaties to make it unlawful to engage in such activity.

*Id.* at pt.1, at 10.

ment will be very much more difficult, *inter alia*, because of objections thrown up by some privacy considerations. Thus, if legislation tries only to cover the acts of circumvention themselves, it cannot provide adequate legal protection and effective legal remedies against such acts which, in spite of the treaty obligations, would continue uncontrolled. It is, however, still possible to provide such protection and remedies. Considering the complexity of the technology involved, in most cases, such acts may only take place after the acquisition of the necessary circumvention device or service. Thus, the possible direction of providing the protection and remedies in harmony with the obligations is as follows: to stop the unauthorized acts of circumvention by cutting the supply line of illicit circumvention devices and services by prohibiting the manufacture, importation and distribution of such devices and the offering of such services.<sup>123</sup>

Dr. Ficsor's conclusion regarding this approach to implementing the treaties' requirement has been endorsed by other respected scholars, such as Jane Ginsburg, Sam Ricketson, Jorg Reinbothe, and Silke von Lewinski, as well.<sup>124</sup> Given that WIPO Treaties signatories are obligated to provide the same level of protection to rights management information,<sup>125</sup> or CMI as it is called in the United States, similar device and services prohibitions appear to be required.

Even if such prohibitions are not required, they are desirable. Many of the objections that commentators have raised about prohibitions around enabling circumvention of TPMs do not apply to prohibitions against dealing in devices or services that enable removal or alteration of copyright management information. For example, scholars such as David Nimmer have expressed concern that the prohibitions related to TPMs will prevent people from engaging in activities that qualify as fair use.<sup>126</sup> Preventing

---

123. Ficsor, *supra* note 97, at 549.

124. See Ricketson & Ginsburg, *supra* note 93, at 976–77 (concluding that it should not be inferred from the absence of an explicit requirement in article 11 of the WCT that signatories are not required to prohibit dealing in circumvention devices and services); JÖRG REINBOTHE & SILKE VON LEWINSKI, *THE WIPO TREATIES* 1996 144 (2002) (limiting prohibition to acts of circumvention “would not correspond to the objective of the provision” because “the manufacturing and distribution of devices which permit of facilitate circumvention may potentially cause more important prejudice to rightholders than acts of circumvention”).

125. See Ficsor, *supra* note 97, at 564.

126. See generally David Nimmer, *A Riff on Fair Use in the Digital Millennium Copyright Act*, 148 U. PA. L. REV. 673 (2000).

people from removing or altering watermarks would not prevent anyone from engaging in fair use or any other form of copying.<sup>127</sup>

Adding prohibitions against dealing in devices or services aimed at altering or removing CMI (including transactional watermarks) would provide copyright owners a valuable new tool and discourage infringement without unduly burdening users of works, who will be able to engage in transformative and personal copying. Applying Senator Hatch's standard for judging copyright legislation aimed at technologies, this proposal meets the test:<sup>128</sup> (1) such legislation is necessary to freeing copyright holders from reliance upon TPMs; (2) the circumvention of transactional watermarks is an identifiable problem, or at the very least is clearly on the horizon; and (3) the proposal may be implemented while still allowing watermarking technology to develop in any direction.

## V. PRIVACY AND SPEECH CONCERNS: PROTECTING WATERMARKS WOULD ENCOURAGE MORE SPEECH THAN IT WOULD CHILL

This Part addresses several legitimate privacy and speech concerns, concluding that the proposal's benefits ultimately outweigh them.

Anonymous speech is protected by the First Amendment in many contexts, including online.<sup>129</sup> If the definition of CMI were expanded to in-

---

127. It may prevent someone from anonymously engaging in fair use. *See infra* Part V.

128. *See* Hatch, *supra* note 24, at 727. I believe the proposal also satisfies, to the extent that it is applicable, the test for determining whether proposed intellectual property laws deserve passage put forward by Congressman Robert W. Kastenmeier and Michael J. Remington in their article, *The Semiconductor Chip Protection Act of 1984: A Swamp or Firm Ground?*, 70 MINN. L. REV. 417 (1985), which has been endorsed by Lawrence Lessig in his article, *The Balance of Robert Kastenmeier*, 2004 WIS. L. REV. 1015, 1029 (2004). That test requires copyright legislation to meet the following requirements:

First, the proponent of a new interest ought to show that the interest can fit harmoniously within the existing legal framework without violating existing principles or basic concepts. . . . Second, the proponent of a new intellectual property interest must be able to commit the new expression to a reasonably clear and satisfactory definition. . . . Third, the proponent of change should present an honest analysis of all the costs and benefits of the proposed legislation. . . . Fourth, any advocate of a new protectable interest should show on the record how giving protection to that interest will enrich or enhance the aggregate public domain.

Kastenmeier & Remington, at 440–41.

129. *See In re Verizon Internet Servs., Inc.*, 257 F. Supp. 2d 244, 258 (D.D.C. 2003) (“The First Amendment Protects Anonymous Expression on the Internet.”); *London-Sire Records, Inc. v. Doe 1*, 542 F. Supp. 2d 153, 163 (D. Mass. 2008) (“[W]hile the aspect of

clude transactional watermarks that contain information about the users (e.g., purchasers) of digital files, it would be unlawful for individuals to remove CMI in order to use copyrighted works in anonymous speech, including speech authorized by the fair use doctrine, such as political parody or satire.<sup>130</sup> Commentators have argued that courts should strike down legislative efforts to protect copyrights online that inhibit anonymous speech.<sup>131</sup> Moreover, the Supreme Court has instructed the lower courts to apply some level of First Amendment scrutiny when Congress has “altered the traditional contours of copyright protection. . . .”<sup>132</sup>

However, the Supreme Court has stressed that copyright laws themselves provide speech-related benefits. “[T]he Framers intended copyright itself to be the engine of free expression. By establishing a marketable right to use of one’s expression, copyright supplies the economic incentive to create and disseminate ideas.”<sup>133</sup> Given that copyright simultaneously benefits and inhibits speech, it deserves a unique place in our First Amendment jurisprudence.<sup>134</sup> Justice Stephen Breyer, a copyright skeptic,<sup>135</sup> dissenting in *Eldred v. Ashcroft*, suggested that the Court should develop a copyright-specific standard of First Amendment review to determine whether a statute lacks “rational support.”<sup>136</sup> This standard would set a lower threshold than traditional “strict” or “intermediate” scrutiny, which is appropriate given that copyright laws are a constitutionally endorsed mechanism of creating speech related benefits that sometimes simultaneously burden some speech.<sup>137</sup> Although the test comes from a dis-

---

a file-sharer’s act that is infringing is not entitled to First Amendment protection, other aspects of it are.”).

130. Of course, Congress could choose to explicitly create a fair use exception to the prohibition against removing CMI. However, this may undermine the overall effect of expanding the definition.

131. See, e.g., Cohen, *Right to Read*, *supra* note 32; Katyal, *Privacy*, *supra* note 32.

132. *Eldred v. Ashcroft*, 537 U.S. 186, 221 (2003).

133. *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985).

134. See *cf.* Melville B. Nimmer, *The Right to Speak From Times to Time: First Amendment Theory Applied to Libel and Misapplied to Privacy*, 56 CALIF. L. REV. 935, 953 (1968) (discussing the speech benefits of properly restrained defamation laws, and the unique approach to defamation articulated by the Supreme Court in *New York Times v. Sullivan*).

135. See Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies and Computer Programs*, 84 HARV. L. REV. 281, 329 (1970) (“[T]he harms that [copyright] causes grow more rapidly than the benefits that it yields.”).

136. *Eldred v. Ashcroft*, 537 U.S. at 245 (Breyer, J., dissenting).

137. If a statute qualifies as “content based,” courts apply a test referred to as strict scrutiny to determine whether the statute is constitutional. See *U.S. v. Playboy Entm’t Group, Inc.*, 529 U.S. 803, 813 (2000) (“If a statute regulates speech based on its content, it must be narrowly tailored to promote a compelling government interest.”). If a statute

senting opinion, no Supreme Court majority opinion has articulated what the appropriate level of scrutiny is for a copyright statute that alters the traditional contours of copyright protection.<sup>138</sup> The Court has set lower levels of scrutiny in other contexts,<sup>139</sup> and if courts applied such a standard to legislation expanding the definition of CMI, the legislation should satisfy the requirements of the First Amendment.<sup>140</sup>

#### A. The First Amendment Protects Anonymous Speech

The Supreme Court has eloquently protected anonymous literary and political speech. In *Talley v. California*, the Court struck down a Los Angeles City ordinance restricting the distribution of handbills without printing the name of the creators and distributors on the covers.<sup>141</sup> California tried to defend the ordinance by arguing that disclosure of names was required to “identify those responsible for fraud, false advertising and libel.”<sup>142</sup> But Justice Black’s opinion explained that such a justification is insufficient given that “[a]nonymous pamphlets, leaflets, brochures and even books have played an important role in the progress of mankind. Persecuted groups . . . have been able to criticize oppressive practices and laws either anonymously or not at all.”<sup>143</sup>

Thirty-five years later, in *McIntyre v. Ohio Elections Commission*, the Court once again confronted a statute that required leafleters to disclose their identities.<sup>144</sup> The Ohio law at issue only imposed such a requirement, however, where the leaflets were designed to influence voters in an elec-

---

qualifies as “content neutral,” courts apply intermediate scrutiny. *See* *Turner Broad. Sys. v. FCC*, 520 U.S. 180, 189 (1997) (holding that intermediate scrutiny requires that a statute “[1] advances important government interests unrelated to the suppression of free speech and [2] does not burden substantially more speech than necessary to further those interests.”).

138. Lower courts have applied intermediate scrutiny in some cases. *See, e.g.*, *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 450 (2d Cir. 2001).

139. *See* Catherine J.K. Sandoval, *Antitrust Language Barriers: First Amendment Constraints on Defining an Antitrust Market by a Broadcast’s Language, and its Implications for Audiences, Competition, and Democracy*, 60 *FED. COMM. L.J.* 407, 414–20 (2008) (discussing lower levels of scrutiny applied to commercial speech and broadcast regulations).

140. In my opinion, legislation protecting transactional watermarks would also pass intermediate scrutiny if it was properly drafted to preserve to the extent practicable the privacy required for anonymous speech.

141. *Talley v. California*, 362 U.S. 60, 65 (1960).

142. *Id.* at 64.

143. *Id.*

144. *McIntyre v. Ohio Elections Comm’n*, 514 U.S. 334 (1995).

tion. Nevertheless, the Court again struck down the law. Justice Stevens concluded:

Under our Constitution, anonymous pamphleteering is not a pernicious, fraudulent practice, but an honorable tradition of advocacy and of dissent. Anonymity is a shield from the tyranny of the majority. It thus exemplifies the purpose behind the Bill of Rights, and of the First Amendment in particular: to protect unpopular individuals from retaliation—and their ideas from suppression—at the hand of an intolerant society. The right to remain anonymous may be abused when it shields fraudulent conduct. But political speech by its nature will sometimes have unpalatable consequences, and, in general, our society accords greater weight to the value of free speech than to the dangers of its misuse.<sup>145</sup>

The Court picked up on this stirring language once again in *Buckley v. American Constitutional Law Foundation*.<sup>146</sup> There, the Court considered the constitutionality of a Colorado law that required, inter alia, petition distributors to wear name badges. Justice Ginsburg reasoned that the “badge requirement compels personal name identification at the precise moment when the circulator’s interest in anonymity is greatest,” and held the provision invalid.<sup>147</sup>

Scholars have relied on cases like *Talley*, *McIntyre*, and *Buckley*, among others,<sup>148</sup> in thoughtful and important articles about the relationship between copyright, anonymity, and privacy.<sup>149</sup> In a 1996 article entitled *A Right to Read Anonymously*, Julie Cohen argued “that reading is so intimately connected with speech, and so expressive in its own right, that the freedom to read anonymously must be considered a right that the First Amendment protects.”<sup>150</sup> Professor Cohen also concluded that any law that prohibits a reader from tampering “with copyright management systems solely to preserve their own anonymity or the anonymity of others” would be unconstitutional.<sup>151</sup>

David Nimmer seconded Professor Cohen’s objections to decreased anonymity in the name of copyright protection in his 1998 article *Aus Der*

---

145. *Id.* at 357.

146. *Buckley v. Am. Constitutional Law Found.*, 525 U.S. 182 (1999).

147. *Id.* at 199.

148. *See, e.g.*, *Lamont v. Postmaster Gen.*, 381 U.S. 301 (1965); *Stanley v. Georgia*, 394 U.S. 557 (1969).

149. *See supra* note 32.

150. Cohen, *Right to Read*, *supra* note 32 at 1038–39.

151. *Id.* at 1039.

*Neuen Welt*.<sup>152</sup> Professor Nimmer argued that “[w]hen the government has available to it an unexpurgated printout of each and every word, note, and image that has entered my brain for the past dozen years, then the era of Big Brother will have dawned.”<sup>153</sup>

More recently, Sonia Katyal has argued that copyright owners who use technological means to monitor online activities for infringement create a panopticon of sorts for Internet users that chills free expression, such as transformative uses of copyrighted material.<sup>154</sup> In Professor Katyal’s scenario, the inability of creators to confidently use the works of other creators in new works of authorship without fear of exposure discourages the production and distribution of the very expression copyright laws are intended to promote.<sup>155</sup>

The concerns expressed by Professors Cohen, Nimmer, and Katyal are rooted in the Supreme Court precedents that justify protection for anonymity and privacy in terms similar to those used to justify the existence of copyright protection. For example:

Great works of literature have frequently been produced by authors writing under assumed names. Despite readers’ curiosity and the public’s interest in identifying the creator of a work of art, an author generally is free to decide whether or not to disclose his or her true identity. The decision in favor of anonymity may be motivated by fear of economic or official retaliation, by concern about social ostracism, or merely by a desire to preserve as much of one’s privacy as possible. Whatever the motivation may be, at least in the field of literary endeavor, the interest in having anonymous works enter the marketplace of ideas unquestionably outweighs any public interest in requiring disclosure as a condition of entry. Accordingly, an author’s decision to remain anonymous, like other decisions concerning omissions or addi-

---

152. Nimmer, *Aus Der Neuen Welt*, *supra* note 31, at 210.

153. *Id.*

154. Katyal, *Privacy*, *supra* note 32, at 318 (“The Panopticon refers to the design of a prison that facilitates constant surveillance by placing guards in a central tower, thereby creating a sense of ‘conscious and permanent visibility that assures the automatic functioning of power.’”) (quoting OSCAR H. GANDY, JR., *THE PANOPTIC SORT: A POLITICAL ECONOMY OF PERSONAL INFORMATION* 9 (1993)); *see also* MICHEL FOUCAULT, *DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON* 195 (Alan Sheridan trans., Vintage Books 2d ed. 1995) (1975).

155. *See* *Golan v. Gonzales*, 501 F.3d 1179, 1188 (10th Cir. 2007) (“It is clear that the Copyright Clause is meant to foster values enshrined in the First Amendment. The Clause’s primary purpose is to provide authors with incentives to produce works that will benefit the public.”) (internal citation omitted).

tions to the content of a publication, is an aspect of the freedom of speech protected by the First Amendment.<sup>156</sup>

Thus, privacy and the right to speak anonymously, or to choose not to speak at all, are clearly linked to creative output, and have a similar underlying justification to copyright laws.<sup>157</sup>

Nevertheless, privacy rights, and copyright rights are not always pulling in the same direction. As Professor Cohen has written: “Privacy rights in information about intellectual activities and preferences preserve the privacy interest in (metaphoric) breathing space for thought, exploration, and personal growth.”<sup>158</sup> Her choice of words connects privacy rights and anonymity to the “breathing space” that the Supreme Court has said the fair use doctrine creates “within the confines of copyright.”<sup>159</sup> Such breathing space is required by the First Amendment.<sup>160</sup> Although speech related privacy protection and copyright protection have similar rationales,

---

156. *McIntyre v. Ohio Elections Comm’n*, 514 U.S. 334, 341–42 (1995).

157. In fact, copyright law itself has been used to protect privacy interests. *See, e.g.*, Roger J. Miner, *Exploiting Stolen Text: Fair Use or Foul Play?*, 37 J. COPYRIGHT SOC’Y 1, 6–11 (1989) (arguing that copyright should prevent exposure of unpublished letters); Jonathan Zittrain, *What the Publisher Can Teach the Patient: Intellectual Property and Privacy in the Era of Trusted Privication*, 52 STAN. L. REV. 1201, 1203 (2000) (arguing that “there is a profound relationship between those who wish to protect intellectual property and those who wish to protect privacy”); Cohen, *DRM*, *supra* note 32, at 593 (stating that “[t]he argument that effective privacy protection should include control over the spaces of intellectual consumption finds support . . . within both the substantive provisions and the overall structure of copyright law”).

158. Cohen, *DRM*, *supra* note 32, at 578; *see also* Julie E. Cohen et al., *Copyright & Privacy—Through the Privacy Lens*, 4 J. MARSHALL. REV. INTELL. PROP. L. 273, 282 (2004).

159. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 579 (1994). This is an interesting connection given that fair use has been used to reduce the scope of privacy rights pertaining to unpublished works. As Judge Pierre Leval has argued:

I do not argue that a writer of private documents has no legal entitlement to privacy. He may well have such an entitlement. The law of privacy, however, and not the law of copyright supplies such protection. Placing all unpublished private papers under lock and key, immune from any fair use, for periods of fifty to one hundred years, conflicts with the purposes of the copyright clause. Such a rule would use copyright to further secrecy and concealment instead of public illumination.

Pierre N. Leval, Commentary, *Toward a Fair Use Standard*, 103 HARV. L. REV. 1105, 1119 (1990). Joseph Liu also has written about the connection between the “breathing space” created by the fair use doctrine and the “breathing space” created by First Amendment exceptions to defamation and tort law. *See generally* Joseph P. Liu, *Copyright and Breathing Space*, 30 COLUM. J.L. & ARTS 429 (2007).

160. *See Eldred v. Ashcroft*, 537 U.S. 186, 221 (2003) (describing ways in which the fair use doctrine protects speech).

there are spaces in which they collide. In such circumstances, it is necessary to consider which type of protection better serves speech interests.<sup>161</sup>

## B. The Traditional Contours of Copyright Protection

Given that scholars like Professors Nimmer and Cohen were urging the importance of anonymity and privacy online while Congress was in the process of implementing the WIPO treaties, Congress may have had the Supreme Court's emphatic protection of anonymity and privacy related to creative and political expression in mind when it chose to exclude identifying information from the definition of CMI. However, the Supreme Court has consistently viewed copyright laws as "the engine of free expression."<sup>162</sup> Most recently, in *Eldred v. Ashcroft*, where the constitutionality of the Sonny Bono Copyright Term Extension Act<sup>163</sup> was at issue, Justice Ginsburg's majority opinion explained that when copyright revisions do not "alter[] the traditional contours of copyright protection," further First Amendment review is unnecessary because copyright laws themselves function alongside the First Amendment to increase public speech.<sup>164</sup>

Commentators have struggled to determine what the traditional contours of copyright protection are.<sup>165</sup> Although *Eldred* clearly labeled the fair use doctrine and the idea/expression distinction as "built in free speech safeguards" that make up part, if not all, of the traditional contours, the opinion left the exact scope of the phrase unpronounced.<sup>166</sup>

---

161. Similarly, there are circumstances in which freedom of speech collides with privacy protections. The Supreme Court has fashioned tests to determine when the First Amendment trumps state privacy statutes. *See, e.g., Time, Inc. v. Hill*, 385 U.S. 374, 388 (1967) ("Exposure of the self to others in varying degrees is a concomitant of life in a civilized community. The risk of this exposure is an essential incident of life in a society which places a primary value on freedom of speech and of press.").

162. *See, e.g., Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539, 558 (1985).

163. Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified at 17 U.S.C. §§ 108, 203, 301, 304).

164. *Eldred*, 537 U.S. at 221.

165. *See, e.g.,* Robert Kasunic, *Preserving the Traditional Contours of Copyright*, 30 COLUM. J.L. & ARTS 397 (2007); Matt Williams, Comment, *Balancing Free Speech Interests: The Traditional Contours of Copyright Protection and the Visual Artists' Rights Act*, 13 UCLA ENT. L. REV. 105 (2005) [hereinafter Williams, *Balancing*]; Marshall Leaffer, *Life After Eldred: The Supreme Court and the Future of Copyright*, 30 WM. MITCHELL L. REV. 1597 (2004); Michael D. Birnhack, *Copyright Law and Free Speech After Eldred v. Ashcroft*, 76 S. CAL. L. REV. 1275 (2003).

166. *Eldred*, 537 U.S. at 221.

Recently, a Circuit split has arguably developed between the Ninth Circuit and the Tenth Circuit on this point.<sup>167</sup> In *Kahle v. Gonzales*, the plaintiff argued that Congress altered the traditional contours of copyright protection by eliminating the copyright renewal requirement for works created between 1964 and 1977.<sup>168</sup> The government argued in response that the traditional contours of copyright protection consist of the fair use doctrine and the idea/expression distinction alone.<sup>169</sup> The Ninth Circuit concluded in May of 2007 that “extending existing copyrights while preserving speech-protective measures does not alter the ‘traditional contours of copyright protection.’”<sup>170</sup>

Four months later, the Tenth Circuit, in *Golan v. Gonzales*, confronted whether Congress altered the traditional contours of copyright protection by restoring copyright protection for works of foreign authorship that had entered the public domain due to U.S. formalities such as renewal, registration, and publication with notice.<sup>171</sup> The court concluded, without referencing *Kahle*, that “one of the[] traditional contours is the principle that once a work enters the public domain, no individual—not even the creator—may copyright it.”<sup>172</sup>

In a much more detailed analysis of the relationship between the First Amendment and copyright laws than found in *Kahle*, the *Golan* opinion stated: “It is clear that the Copyright Clause is meant to foster values enshrined in the First Amendment. The Clause’s primary purpose is to provide authors with incentives to produce works that will benefit the pub-

---

167. See Christopher A. Mohr, *Traditional Contours of Copyright, Silver Lining or Storm Clouds*, 1(1) LANDSLIDE 30, 33 (2008) (arguing that the Tenth Circuit has given a “far broader reading to the *Eldred*’s traditional contours pronouncement” than the Ninth Circuit).

168. *Kahle v. Gonzales*, 474 F.3d 665, 668, *reh’g en banc denied, amended by*, 487 F.3d 697 (9th Cir. 2007), *cert. denied*, *Kahle v. Mukasey*, 128 S. Ct. 958 (2008) (analyzing the constitutionality of the Copyright Renewal Act of 1992, Pub. L. No. 102-307, 106 Stat. 264 (codified at 17 U.S.C. § 304 (2000)), and the Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, 112 Stat. 2827 (1998)).

169. See Brief for the Appellee, 2005 WL 926823, at \*15, *Kahle v. Gonzales*, 474 F.3d 665, *reh’g en banc denied, amended by*, 487 F.3d 697 (9th Cir. 2007), *cert. denied*, 128 S. Ct. 958 (2008) (No. 04-17434) (“[N]either the 1992 Act nor the CTEA alters ‘the traditional contours of copyright protection’—in particular, the ‘traditional First Amendment safeguards,’ i.e., the ‘idea/expression dichotomy’ and the ‘fair use’ defense comprising copyright law’s ‘built-in First Amendment accommodations’—and thus these statutes require no further First Amendment scrutiny.”).

170. *Kahle*, 487 F.3d at 700.

171. *Golan v. Gonzales*, 501 F.3d 1179 (10th Cir. 2007).

172. *Id.* at 1184.

lic.”<sup>173</sup> Nevertheless, the court read *Eldred* to limit Congress’ ability to pass laws designed to create such incentives if Congress alters the “functional” or the “historical” aspects of copyright protection. By functional, the Tenth Circuit meant “the outline” or “the general form or structure” of copyright protection. By historical, the Court referred to the “bedrock principle” by which Congress has traditionally legislated in the copyright area.<sup>174</sup> Thus, in the Tenth Circuit, legislation alters the traditional contours of copyright protection if it revises the scope of protection in a manner inconsistent with the historical progression of copyright laws.

Regardless of whether the Tenth Circuit, the Ninth Circuit, or the U.S. Government is correct regarding the meaning of the mysterious *Eldred* phrase, a further question exists regarding the proper method of First Amendment review of copyright laws that do alter the traditional contours of copyright protection: the appropriate level of scrutiny. Even if a statute clearly alters the traditional contours, what test should be used to determine constitutionality? The *Golan* opinion instructed the district court on remand to determine whether the statute at issue is content based or content neutral, and then, based on that determination, to apply either strict scrutiny or intermediate scrutiny.<sup>175</sup> However, this instruction ignores Justice Breyer’s dissent in *Eldred*, which proposed a form of rational basis review applicable exclusively to copyright statutes.<sup>176</sup> That level of review would set a less strenuous standard for constitutionality based on copyright’s relationship to the First Amendment: copyright benefits speech while at the same time restricting speech.

Although Justice Breyer believed that the statute at issue in *Eldred*, the Copyright Term Extension Act,<sup>177</sup> was unconstitutional, he recognized that copyright laws have a unique place in our Constitutional system.

The Copyright Clause and the First Amendment seek related objectives—the creation and dissemination of information. When working in tandem, these provisions mutually reinforce each other, the first serving as an “engine of free expression,” the second assuring that government throws up no obstacle to its dissemination. At the same time, a particular statute that exceeds proper Copyright Clause bounds may set Clause and Amendment at cross-purposes, thereby depriving the public of the

---

173. *Golan*, 501 F.3d at 1188.

174. *Id.* at 1187.

175. *Id.* at 1196.

176. *Eldred v. Ashcroft*, 537 U.S. 186, 245 (2003) (Breyer, J., dissenting).

177. Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, 112 Stat. 2827 (1998) (codified at 17 U.S.C. §§ 108, 203, 301, 304).

speech-related benefits that the Founders, through both, have promised.<sup>178</sup>

Because of copyright's unique legal role, Justice Breyer suggested that traditional standards of scrutiny designed for other types of statutes are inapplicable to copyright statutes.

There is no need in this case to characterize that review as a search for “ ‘congruence and proportionality,’ ” or as some other variation of what this Court has called “intermediate scrutiny[.]” Rather, it is necessary only to recognize that this statute involves not pure economic regulation, but regulation of expression, and what may count as rational where economic regulation is at issue is not necessarily rational where we focus on expression-in a Nation constitutionally dedicated to the free dissemination of speech, information, learning, and culture. In this sense only, and where line-drawing among constitutional interests is at issue, I would look harder than does the majority at the statute's rationality-though less hard than precedent might justify. Thus, I would find that the statute lacks the constitutionally necessary rational support (1) if the significant benefits that it bestows are private, not public; (2) if it threatens seriously to undermine the expressive values that the Copyright Clause embodies; and (3) if it cannot find justification in any significant Clause-related objective. Where, after examination of the statute, it becomes difficult, if not impossible, even to dispute these characterizations, Congress' “choice is clearly wrong.”<sup>179</sup>

Given that Justice Breyer has a tendency to question copyright's underlying premise, it is likely safe to assume that the other Justices would not suggest applying a more strenuous level of scrutiny than the one he set out in *Eldred*.<sup>180</sup> Thus, the Tenth Circuit's instruction to the district court in *Golan* may have been inconsistent with the Supreme Court's approach to speech related copyright statutes. After all, Justice Ginsburg's majority opinion in *Eldred* rejected the “imposition of uncommonly strict scrutiny on a copyright scheme that incorporates its own speech protective purposes. . . .”<sup>181</sup> Even if we assume that hypothetical legislation expanding the scope of the definition of CMI to protect transactional watermarks would alter the traditional contours of copyright protection, and thus require a

---

178. *Eldred*, 537 U.S. at 244 (internal citations omitted).

179. *Id.* at 244–45 (internal citations omitted).

180. See Breyer, *supra* note 135, at 350 (“[T]he general case for copyright protection is weak.”).

181. *Eldred*, 537 U.S. at 218–19.

court to apply some form of First Amendment scrutiny to determine its constitutionality, no Supreme Court precedent suggests that intermediate or strict scrutiny is the appropriate standard. If we apply Justice Breyer's test instead, as this article does in the next Section, then such legislation should be found compliant with the First Amendment.

**C. Protecting Transactional Watermarks Would Likely Alter the Traditional Contours of Copyright Protection but The Speech Related Benefits of Expanding the Definition of CMI Would Outweigh the Potential Chilling Effects**

As discussed above, First Amendment protection for anonymous speech creates "breathing space" for artistic and political expression in a manner analogous to the fair use doctrine's creation of "breathing space within the confines of copyright."<sup>182</sup> Regardless of the precise scope of the traditional contours of copyright protection, it is clear that Justice Ginsburg included the speech related benefits of the fair use doctrine within that scope.<sup>183</sup> The fair use doctrine prevents copyright from discouraging artistic and political speech because it declares some transformative works, such as certain satires and parodies, noninfringing.<sup>184</sup> Similarly, protecting anonymous speech prevents identity disclosure from discouraging distribution of such satires and parodies.<sup>185</sup> If the Copyright Act prohibited me from removing a watermark from a copy of a work that disclosed my identity, and I used that copy in a new work of authorship that criticized the underlying work, I may be discouraged from distributing the critical transformative work. This discouragement likely alters the traditional contours of copyright protection. Central to Justice Ginsburg's rea-

---

182. See *supra* notes 158–162 and accompanying text.

183. See *Eldred*, 537 U.S. at 221 (stating that "copyright's built-in free speech safeguards are generally adequate" to address First Amendment challenges).

184. See Williams, *Transformative*, *supra* note 12, at 315–16. Labeling a work a parody or satire does not allow the author to reproduce another's creative expression gratuitously:

[A] parody is likely fair where it 'needs to mimic an original to make its point.' However, where such a need does not exist and an author uses another author's expression to criticize a general societal ailment (e.g., a satire) by grabbing the audience's attention with the pre-existing use, less justification exists, and other factors may prove the transformative use unfair. This is not to say that parodies are always fair, or that satires are not, but only that the two types of transformative uses are of differing weights.

*Id.*

185. See Cohen, *DRM*, *supra* note 32, at 598 ("Anonymity . . . allows fair users to decide later whether to reveal their identities when releasing their work.").

soning in *Eldred* was that “[t]he First Amendment securely protects the freedom to make—or decline to make—one’s own speeches; it bears less heavily when speakers assert the right to make other people’s speeches.”<sup>186</sup> Requiring creators of transformative works to disclose their identities arguably forces them to speak in a manner they would prefer not to. Doing so likely alters the traditional contours of copyright protection.<sup>187</sup>

Nevertheless, even assuming that to be true, expanding the definition of CMI to protect transactional watermarks is likely constitutional under Justice Breyer’s test. Under that test, a statute is unconstitutional “(1) if the significant benefits that it bestows are private, not public; (2) if it threatens seriously to undermine the expressive values that the Copyright Clause embodies; and (3) if it cannot find justification in any significant Clause-related objective.”<sup>188</sup>

First, the significant benefits of protecting transactional watermarks are both private and public. The benefits are arguably private in that they assist copyright owners in enforcing their copyrights. However, they are public in that protecting transactional watermarks will encourage copyright owners to reduce the use of use-restrictive TPMs that prevent consumers from engaging in personal and transformative copying. The benefits are also public in that the p2p distribution of copies threatens the continued viability of copyright-based business models and in so doing undermines copyright’s foundational premise; that granting exclusive rights in works leads to increased creative expression and distribution thereof.<sup>189</sup> As Erwin Chemerinsky has stated, “Copyright laws are permitted, even in a society that deeply values freedom of speech, because they are seen overall as enhancing expression. . . . Without copyright protections, less speech would occur.”<sup>190</sup> If we allow infringement on p2p networks to go unchecked, copyright based industries will suffer further, and the public will likely obtain access to fewer works as a result.<sup>191</sup>

---

186. *Eldred*, 537 U.S. at 221.

187. See generally Williams, *Balancing*, *supra* note 165.

188. *Eldred*, 537 U.S. at 245 (Breyer, J., dissenting).

189. *Id.* at 219 (“The Copyright Clause and the First Amendment were adopted close in time. This proximity indicates that, in the Framers’ view, copyright’s limited monopolies are compatible with free speech principles. Indeed, copyright’s purpose is to *promote* the creation and publication of free expression.”) (emphasis in original).

190. Erwin Chemerinsky, *Balancing Copyright Protections and Freedom of Speech: Why the Copyright Extension Act is Unconstitutional*, 36 LOY. L.A. L. REV. 83 (2002).

191. See *id.* at 84 (“I wrote an amicus brief for the Ninth Circuit in the Napster case, arguing that applying copyright laws in that situation enhanced speech. Allowing for distribution of copyrighted music without paying royalties would lessen the incentives for musicians, their producers, and distributors to engage in speech activities.”); see also

Second, protecting transactional watermarks would not seriously undermine the expressive values that the Copyright Clause embodies. Enabling copyright owners to protect their rights is an essential part of providing adequate copyright protection.<sup>192</sup> Prohibiting removal of transactional watermarks, as discussed above, would reduce evidentiary problems associated with proving infringement by p2p users.<sup>193</sup> Courts have consistently held that the First Amendment does not prevent copyright owners from obtaining the identity of alleged infringers.<sup>194</sup> Although protecting transactional watermarks may discourage some creative expression in the form of anonymous fair use, it would not seriously undermine the expressive values of the Copyright Clause when this discouragement is compared to the encouragement that protecting transactional watermarks would provide to

---

Brief of Amici Curiae Motion Picture Association of America, Inc. (MPAA) et al. in Support of Appellee's Position Seeking Affirmance, 2005 WL 926824, at \*6, *Kahle v. Gonzales*, 487 F.3d 697 (9th Cir. 2007) (No. 04-17434) ("Congress could have reasonably concluded that inadvertent forfeiture resulting from complex and technical renewal rules actually inhibited the dissemination of speech by interfering with copyright's incentives to produce expressive works, and that automatic renewal was necessary to protect free speech values.").

192. *See* *Universal City Studios, Inc. v. Reimerdes*, 82 F. Supp. 2d 211, 220 (S.D.N.Y. 2000) ("[The Supreme Court] has made it unmistakably clear that the First Amendment does not shield copyright infringement.").

193. *See supra* notes 77–81 and accompanying text.

194. Courts have upheld a provision of the DMCA, 17 U.S.C. § 512(h) (2000), that provides copyright owners with an expedited process for obtaining subpoenas ordering disclosure of the identities of alleged online infringers. For example, the U.S. District Court for the District of Columbia stated the following in *In re Verizon Internet Servs., Inc.*, 257 F. Supp. 2d 244, 261 (D.D.C. 2003) :

[T]he DMCA neither authorizes governmental censorship nor involves prior restraint of potentially protected expression. Section 512(h) merely allows a private copyright owner to obtain the identity of an alleged copyright infringer in order to protect constitutionally-recognized rights in creative works; it does not even directly seek or restrain the underlying expression (the sharing of copyrighted material).

*Id.*; *see also* *Sony Music Entm't Inc. v. Does 1–40*, 326 F. Supp. 2d 556, 567 (S.D.N.Y. 2004) (holding that the subpoena granted in the John Doe suit was constitutional because "defendants' First Amendment right to remain anonymous must give way to plaintiffs' right to use the judicial process to pursue what appear to be meritorious copyright infringement claims"). Courts have also upheld state statutes requiring distributors of recordings and audiovisual materials to place the name of the distributor on the goods. *See, e.g., Anderson v. Nidorf*, 26 F.3d 100, 103 (9th Cir. 1994) ("Disclosure of the manufacturer . . . protects against piracy."); *see also* Brian McFarlin, *From the Fringes of Copyright Law: Examining California's "True Name and Address" Internet Piracy Statute*, 35 HASTINGS CONST. L.Q. 547, 564 (2008). These statutes are arguably analogous to a statute prohibiting removal or alteration of transactional watermarks.

artists and distributors of copyrightable expression.<sup>195</sup> In addition, the facilitation of personal and transformative copying that transactional watermarks provide, by lessening reliance on TPMs, also encourages expressive activities. These benefits would come with relatively minor privacy intrusions if a thoughtful approach to protecting watermarks is implemented. The Center for Democracy and Technology has proposed methods for including privacy protections such as notice to users of the presence of watermarks<sup>196</sup> within the design of transactional watermarks. Such privacy protections would best protect expressive values.

Finally, as already discussed, protecting transactional watermarks is related to the objective of the Copyright Clause because it facilitates enforcement of existing copyrights and thereby encourages creative expression. Justice Breyer said that a copyright statute is unconstitutional, or “clearly wrong,” under the First Amendment if it is “difficult, if not impossible, even to dispute” that the statute fails all three parts of his test.<sup>197</sup> Clearly, it is easy to dispute that protecting transactional watermarks fails any one part, much less all three.

Nevertheless, some may argue that protecting transactional watermarks would be unconstitutional in certain circumstances. If legislation became law and a court was confronted with a specific case of an individual who removed a transactional watermark before creating a transforma-

---

195. See Goldstein, *supra* note 33. Anecdotal evidence suggests that casual infringers are a far more prevalent threat to speech than over-protective copyright owners:

For every songwriter, biographer or documentary maker who is mistakenly told that her use exceeds what the law considers fair or de minimis or what lies in the public domain, there is at least one other author who has been given a right but no practical remedy. . . . I would venture that the incidence of unrequited infringing uses outnumbers the incidence of unjustified demands by no less than a thousand to one.

*Id.* at 8.

196. See generally CTR. FOR DEMOCRACY AND TECH., *supra* note 27 (proposing methods for protecting privacy).

197. *Eldred v. Ashcroft*, 537 U.S. 186, 245 (2003) (Breyer, J., dissenting). One reading of Justice Breyer’s dissent would see his three-part test as merely a tool for discovering obvious unconstitutionality rather than a test to determine constitutionality. Such a reading would distinguish between a circumstance in which a congressional choice is “clearly wrong” and a circumstance in which a congressional choice is unconstitutional. However, Justice Breyer took the words “clearly wrong” from *Helvering v. Davis*, 301 U.S. 619, 640 (1937), a Supreme Court opinion involving when Congress can tax and spend for the “general welfare.” The reference to this opinion indicates that Justice Breyer sees copyright as an area in which the Court should defer to Congress unless Congress’ decisions are “clearly wrong.” Thus, Justice Breyer appears to have been equating a “clearly wrong” choice with an unconstitutional one.

tive work, would the court see the encouraging forest when confronted with the discouraged tree?

Robert Kasunic, among others, has suggested that courts may create a fair use doctrine of sorts applicable to the DMCA's TPM related anti-circumvention provisions if confronted with specific scenarios in which individuals circumvented TPMs in order to engage in journalism regarding corporate whistle blowing, for example.<sup>198</sup> Perhaps courts would pursue a similar path if confronted with a similar case involving removal of a transactional watermark. However, the adoption of such a path should at least be limited to actual acts of removal. If courts applied a limiting doctrine to prohibitions against distributing devices or services aimed at enabling removal of CMI, the exception would likely swallow the rule by providing every device distributor or service provider with a defense based on a hypothetical circumstance in which their devices or services enabled fair uses. If such arguments arise, courts should reject them, as they have in the TPM context.<sup>199</sup>

---

198. See Kasunic, *supra* note 165, at 412. Professor Kasunic envisions the creation of a safe harbor for certain acts of free expression, like whistleblowing, that open societies need to protect:

Looking to *Eldred* for guidance, it would appear that the optimal way to preserve the traditional contours of copyright so as to safeguard free speech values would be for a court to apply or expand the existing free speech safeguards, or, alternatively, to create an appropriate common law doctrine to address the conflict.

*Id.* at 413; see also Timothy K. Armstrong, *Fair Circumvention*, 74 BROOK. L. REV. (forthcoming 2008) (“[C]ourts are borrowing (and should borrow) factors and criteria that have developed under fair use en route to creating what I have labeled as the doctrine of ‘fair circumvention’ under the DMCA.”); Jane C. Ginsburg, *The Pros and Cons of Strengthening Intellectual Property Protection: Technological Protection Measures and Section 1201 of the U.S. Copyright Act*, 16 INFO. & COMM. TECH. L. 191, 209 (2007). Professor Ginsberg argues that the courts must strike a delicate balance between protecting creative expression and allowing fair use:

Absent evidence that the expanded intellectual property right cannot co-exist with free expression, courts should not introduce exceptions that eviscerate the statute, but recognition of the residual role of fair use in intellectual property law in general suggests that, in appropriate circumstances, courts may temper section 1201 with carefully tailored fair use-equivalent limitations.

*Id.*

199. See, e.g., *Macrovision v. Sima Products Corp.*, 81 U.S.P.Q.2d (BNA) 1923, 1924 (S.D.N.Y. 2006) (“Sima’s defense that it only intends to enable ‘fair use’ copying of copyrighted works is no defense at all—as stated above, the DMCA provides no exception to its prohibition of the manufacture of these devices.”). As Professor Denicola has argued:

## VI. CONCLUSION

Transactional watermarks benefit copyright owners and users of copyrighted material by providing protections for authors and distributors while at the same time facilitating transformative and personal copying. Unfortunately, the law currently fails to provide adequate and effective protection for such watermarks. Congress should consider amending 17 U.S.C. § 1202 to explicitly prohibit the removal or alteration of transactional watermarks as well as dealing in devices or services aimed at enabling such removal or alteration. Doing so would encourage more expression than it would chill, and would be entirely consistent with the First Amendment and our engine of free expression, the Copyright Act.

---

[A] device capable of circumventing technological protection for the purpose of fair use is usually also capable of circumventing for the purpose of infringement. Any generalized exception to the anti-trafficking rules would thus leave copyright owners vulnerable to the same threat of piracy that prompted the passage of the DMCA.

Robert C. Denicola, *Access Controls, Rights Protection, and Circumvention: Interpreting the Digital Millennium Copyright Act to Preserve Noninfringing Use*, 31 COLUM. J.L. & ARTS 209, 216 (2008).

# EXPLORING EUROPEAN UNION COPYRIGHT POLICY THROUGH THE LENS OF THE DATABASE DIRECTIVE

By *Miriam Bitton*<sup>†</sup>

## TABLE OF CONTENTS

I.	<b>INTRODUCTION</b> .....	1412
II.	<b>OVERVIEW OF EUROPEAN UNION COPYRIGHT POLICY</b> .....	1413
	A. ANGLO-AMERICAN COPYRIGHT AND THE DEVELOPMENT OF AUTHOR'S RIGHTS IN CONTINENTAL COPYRIGHT .....	1413
	B. DEVELOPMENT OF THE INFORMATION SOCIETY .....	1415
III.	<b>THE FORMULATION OF THE DATABASE DIRECTIVE</b> .....	1419
	A. THE STATE OF PROTECTION FOR COMPILATIONS PRIOR TO THE DATABASE DIRECTIVE .....	1420
	B. THE EMERGENCE OF DATABASE PROTECTION AS A POLICY MATTER .....	1422
	C. THE FORMULATION PROCESS: FROM PROPOSAL TO DIRECTIVE .....	1427
	1. <i>Settling on Sui Generis Protection</i> .....	1427
	2. <i>Pre-Proposal Discussions</i> .....	1430
	3. <i>The Commission Database Proposal</i> .....	1432
	4. <i>The Proposal and Its Amended Text</i> .....	1436
	5. <i>The Council Common Position</i> .....	1439
	6. <i>The Adopted Database Directive</i> .....	1443
	D. TREATMENT OF DATABASES UNDER THE DATABASE DIRECTIVE .....	1445

---

© 2008 Miriam Bitton

<sup>†</sup> Assistant Professor of Law, Bar-Ilan University Faculty of Law, Israel. LL.B., M.A., Bar-Ilan University Faculty of Law, Israel. LL.M, S.J.D., University of Michigan Law School. Many thanks to the following individuals who provided me with detailed comments to different versions of this Article: Professors Rebecca S. Eisenberg, Professor Rosemarie Ziedoniz from the University of Michigan Business School, Professor Emily M. Morris from the University of Indiana in Indianapolis, and Sarala Nagala (UC Berkeley School of Law, J.D. 2008). Thanks are due to Dr. Jens Gaster, Principal Administrator, Industrial Property Unit, Directorate General D2, European Commission, for his assistance. Thanks are also due to Microsoft Inc. for supporting my work on this Article as a Microsoft Research Fellow at the University of California, Berkeley, Boalt Hall School of Law, Berkeley Center for Law and Technology.

IV. QUESTIONING THE WISDOM OF EU POLICY CONCERNING DATABASE PROTECTION .....	1447
A. DOES THE DATABASE DIRECTIVE REALLY MATTER? .....	1448
1. <i>The Availability of Market-Based, Technological, and Legal Mechanisms for Protection of Databases</i> .....	1448
2. <i>Evaluating the Empirical Evidence to Date: No Evidence of Market Failure</i> .....	1449
a) Empirical Part of the Working Paper .....	1451
b) Policy Options of the Working Paper .....	1456
3. <i>The Futility of Reciprocity</i> .....	1456
4. <i>Non-Identical Implementation of the Directive</i> .....	1457
5. <i>Confusion in the Courts</i> .....	1460
6. <i>Persistence of Other Barriers</i> .....	1463
B. POSSIBLE CHALLENGES TO THE DATABASE DIRECTIVE.....	1463
1. <i>National and EU Constitutional Validity</i> .....	1464
2. <i>Invalidity under International IP Conventions</i> .....	1467
V. CONCLUSION .....	1469

## I. INTRODUCTION

This Article explores and evaluates the European Union's ("EU") efforts to integrate its various copyright laws, using the EU's 1996 Database Directive as a case study. Legal scholarship on this subject to date has focused on and criticized only isolated legislative initiatives and directives at the EU level but has not attempted to explore European copyright law policy in a comprehensive manner. Therefore, this Article aims to evaluate overall EU copyright law policy through the lens of the Database Directive.

Part II provides an overview of EU copyright policy and harmonization efforts, specifically touching upon the different copyright traditions among its Member States and the development of the Information Society within both the EU and the wider international arena.

Part III tells the story of how the Database Directive was formulated, starting with how compilations were protected prior to the Directive, how database protection emerged as a policy matter, and how the Directive evolved through the formal policy process. Among other things, the analysis illustrates the Directive's origins within European copyright policy, highlights the Directive's dual copyright/*sui generis* approach, provides a detailed overview of the main revisions to the text of the Directive, and

indicates the key policy players involved in the formulation process. Most importantly, the analysis critically examines EU copyright policy.

Finally, Part IV turns to EU copyright policy as illustrated by the Database Directive, showing the policy's inherent inefficiency as well as the flawed assumptions underlying it.

## II. OVERVIEW OF EUROPEAN UNION COPYRIGHT POLICY

The following overview provides the necessary background for understanding the origins of EU copyright policy in general and database protection in particular. The differences between the copyright and *droit d'auteur* systems are first highlighted in Section II.A. Such differences within the EU proved to be a major hurdle in the EU's harmonization efforts in general and with regard to database protection in particular. Section II.B discusses the emergence of the information society in Europe and illustrates how EU copyright policy was influenced by international developments.

### A. Anglo-American Copyright and the Development of Author's Rights in Continental Copyright

The emergence of commercial markets for literary, artistic, and technical works resulted in the development of copyright law regimes in Europe. Copyright in England was a state-created monopoly right granted to publishers to print, publish, and sell works in return for censorship of certain content.<sup>1</sup> This situation continued until the passage of the Statute of Anne in 1710.<sup>2</sup> In response to lack of competition in the field of publishing, the British Parliament enacted the Statute of Anne in 1710. The statute formed the basis for copyright laws in all Anglo-American countries, such as the United States. The Statute of Anne transformed copyright into a state-created monopoly for a limited term of fourteen years that was available to authors and others named in the Act.<sup>3</sup> Copyright law expanded in its scope and term of protection in the years following the Statute of Anne. By the beginning of the twentieth century, Anglo-American copyright law had developed into a system of economic rights justified as

---

1. See LYMAN RAY PATTERSON, COPYRIGHT IN HISTORICAL PERSPECTIVE 28–30 (1968) (discussing the charter of the Stationer's Company and the state's motive for granting it); see also MARK ROSE, AUTHORS AND OWNERS: THE INVENTION OF COPYRIGHT 12–13 (1993) (discussing the state's motive for granting monopoly to the Stationer's Company).

2. ROSE, *supra* note 1, at 13.

3. PATTERSON, *supra* note 1, at 143.

a necessary incentive to bring about the creation of and investment in works for society's well-being.<sup>4</sup>

In other countries in Europe the tradition of authors' rights had developed since the sixteenth through the eighteenth centuries in an almost identical manner to Anglo-American copyright law.<sup>5</sup> For example, in France the Crown granted in return for censorship of content a right to publish.<sup>6</sup> However, after the 1789 French Revolution all the Crown's rights were abolished and authors' rights were subsequently perceived as natural rights granted to authors as a reward for their work.<sup>7</sup> Author's rights, unlike the rights granted under the Statute of Anne, did not depend on formalities such as registration or publication and were available for the life of the author and beyond.<sup>8</sup>

The author's rights system not only provided authors with economic rights to exploit the protected work's value for a limited time<sup>9</sup> like its Anglo-American counterpart, but it also gradually provided authors with moral rights. Moral rights reflected the extension of protection to the author's personality.<sup>10</sup> Such a development ultimately brought about the continental *droit d'auteur* systems.<sup>11</sup> This emphasis on the author's personality as the basis for rights had an impact on the need to balance author's rights with the public interest, which was not as important as in the copyright system.<sup>12</sup>

---

4. See Roberta Rosenthal Kwall, *Inspiration and Innovation: The Intrinsic Dimension of the Artistic Soul*, 81 NOTRE DAME L. REV. 1945, 1982–88 (2006).

5. See, e.g., Jane C. Ginsburg, *A Tale of Two Copyrights: Literary Property in Revolutionary France and America*, 64 TUL. L. REV. 991, 997, 1006–14 (1990).

6. *Id.* at 997.

7. *Id.* at 1018–20.

8. *Id.* at 997–998.

9. Neil Netanel, *Copyright Alienability Restrictions and the Enhancement of Author Autonomy: A Normative Evaluation*, 24 RUTGERS L.J. 347, 381–82 (1993) (discussing the “dualist nature” of French author's rights regime).

10. Ginsburg, *supra* note 5, at 993.

11. Roberta Rosenthal Kwall, *Copyright and the Moral Right: Is an American Marriage Possible?*, 38 VAND. L. REV. 1, 12–16 (1985) (comparing artistic protection in the United States and Europe).

12. This development also had an impact on the level of originality required in a work, which has always been held to a higher standard under *droit d'auteur* systems than in copyright systems. As the discussions, *infra* Section III.A, show, these differences between the copyright system and *droit d'auteur* system within the EU proved to be a major hurdle in the EU's harmonization efforts in general and with regard to database protection in particular. The different level of originality required under the two systems, for example, had a major impact on the development of the dual copyright/*sui generis* approach adopted in the Database Directive.

The importance of copyright grew at the national and European level as well as at the international level as a result of a few factors: first, the differences between copyright and author's rights system. Second, information became more important to states' economies. At the international level, this, in turn, led in the late nineteenth and twentieth centuries to the adoption of a few copyright treaties.<sup>13</sup> The most notable change at the international arena, however, occurred during the late 1980s and 1990s when international trade agreements such as the TRIPS Agreement<sup>14</sup> and the North American Free Trade Agreement<sup>15</sup> were adopted and explicitly incorporated copyright regulations as well as other forms of intellectual property. These agreements introduced, *inter alia*, minimum standards of copyright protection and as a result reduced the differences between author's rights and copyright systems.

## **B. Development of the Information Society**

The 1990s brought significant developments in the fields of computers, telecommunications, and information technology. These developments, in turn, stimulated the creation of a new global market for electronic information services and products, a market that is occupied substantially by electronic databases. The emergence of these new technological developments and the global information market challenged many traditional branches of the law, including intellectual property law. A particularly

---

13. In 1886, the Berne Convention for the protection of literary works was adopted. *See* Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, 828 U.N.T.S. 221, as last revised at Paris, July 24, 1971, S. Treaty Doc. No. 99-27, 1161 U.N.T.S. 3 [hereinafter Berne Convention]. The Universal Copyright Convention was then adopted in 1952 to harmonize agreements between Berne Convention members and nonmembers. *See* The Universal Copyright Convention, Sept. 6, 1952, 6 U.S.T. 2731, 216 U.N.T.S. 134, as last revised at Paris, July 24, 1971, 25 U.S.T. 1341, 943 U.N.T.S. 178. Additionally, different international organizations concluded additional agreements during the 1960s and 1970s in response to technological changes and growth in trade in informational works. *See, e.g.*, Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organisations, Oct. 26, 1961, 496 U.N.T.S. 43; Geneva Convention for the Protection of Producers of Phonograms Against Unauthorised Duplication of their Phonograms, Oct. 29, 1971, 25 U.S.T. 309, T.I.A.S. No. 7808 [hereinafter Geneva Phonograms Convention]; Brussels Convention Relating to the Distribution of Programme-Carrying Signals Transmitted by Satellite, May 21, 1974, 1144 U.N.T.S. 3.

14. Agreement on Trade-Related Aspects Of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, Legal Instruments—Results of the Uruguay Round of Multilateral Trade Negotiations, 33 I.L.M. 81 (1994) [hereinafter TRIPs Agreement].

15. North American Free Trade Agreement, U.S.-Can.-Mex., Dec. 17, 1992, 107 Stat. 2057, 32 I.L.M. 605 (1993).

prominent part of this debate centers on how the law should address the protection of electronic databases.<sup>16</sup>

The EU Commission introduced its initial information policies and agenda in a series of a few major documents.<sup>17</sup> The most crucial statements were reflected in the 1988 Green Paper<sup>18</sup> and the 1991 "Follow-Up" Paper,<sup>19</sup> which provided a working program concerning future information

---

16. The debate over database protection can be traced to the U.S. Supreme Court's seminal decision in *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991). Legal scholarship discussing *Feist* abounds. See, e.g., Miriam Bitton, *Trends in Protection for Informational Works Under Copyright Law During the 19th and 20th Centuries*, 13 MICH. TELECOMM. & TECH. L. REV. 115 (2006) (providing a historical analysis regarding the "industrious collection" doctrine as well as the "sweat of the brow" doctrine); J.H. Reichman, *Legal Hybrids Between the Patent and Copyright Paradigms*, 94 COLUM. L. REV. 2432, 2475-76, 2490-92, 2506-20 (1994) (discussing contraction of lead time due to ease of copying databases and other technologies); Miriam Bitton, *A New Outlook on the Economic Dimension of the Database Protection Debate*, 47 IDEA 93, 99-100 (2006) [hereinafter *Economic Dimension*] (discussing "free riding" problem with respect to databases and Reichman's analysis); ANSELM KAMPERMAN SANDERS, UNFAIR COMPETITION LAW: THE PROTECTION OF INTELLECTUAL AND INDUSTRIAL CREATIVITY 100 n.15 (1997) ("Competition is essentially a process of the formation of opinion: by spreading information . . . [i]t creates the views people have about what is best and cheapest.") (quoting F. A. HAYEK, *Economics and Knowledge*, in INDIVIDUALISM AND ECONOMIC ORDER 106 (1948)); see also *Collections of Information Antipiracy Act: Hearing on H.R. 354 Before the Subcomm. on Courts and Intellectual Property of the H. Comm. on the Judiciary*, 106th Cong. 189 (1999) [hereinafter *The 1999 CIAA Hearing*] (statement of Joshua Lederberg, Nobel laureate, on behalf of NAS, NAE, IOM, and the AAAS).

17. International developments, mainly in the United States, also served as a catalyst for action in the EU concerning the creation of the so-called "information society" through the expansion of telecommunication and information services. In the early 1990s, governments in Japan and North America acknowledged the importance of new communication and information technologies for socio-economic growth. In the United States, the new Clinton administration established the Information Infrastructure Task Force ("IITF") in 1993. The IITF introduced the National Information Infrastructure ("NII") agenda, emphasizing information policy in U.S. industrial policy. See *The National Information Infrastructure: Agenda for Action*, Administration Policy Statement, 58 Fed. Reg. 49,025 (Sept. 21, 1993). Similar organizations also emerged in Japan. See MINISTRY OF INT'L TRADE & INDUS., PROGRAM FOR ADVANCED INFORMATION INFRASTRUCTURE (1994), available at <http://www.ifla.org.sg/II/asiapac.htm>; INSTIT. OF INTELLECTUAL PROP. (IIP), MULTIMEDIA COMMITTEE, PREDICTED PROBLEMS AND POSSIBLE SOLUTIONS FOR ADMINISTERING INTELLECTUAL PROPERTY RIGHTS IN A MULTIMEDIA SOCIETY 62 (1995).

18. Comm'n of the European Communities, *Commission Green Paper on Copyright and the Challenge of Technology: Copyright Issues Requiring Immediate Action*, COM (1988) 172 final (June 7, 1988).

19. Comm'n of the European Communities, *Follow-up to the Green Paper: Working Programme of the Commission in the Field of Copyright and Neighbouring Rights*, COM (1990) 584 final.

policy directives. Additionally, the “Delors White Paper,”<sup>20</sup> which the Council later approved in its “Bangemann Report”<sup>21</sup> and follow-up action plan<sup>22</sup> also proved crucial because it was designed to create a European Information Society (“EIS”) and supported the “laissez-faire” approach as a means to achieve the creation of the information society.<sup>23</sup>

Alongside the 1988 White Paper,<sup>24</sup> the 1991 “follow-up” Paper<sup>25</sup> and the release of the Commission’s “Bangemann Report” and action plan, Directorate General XV (DGXV) of the Commission also considered possible reforms in the field of copyright law in light of the emergence of the “information superhighway.”<sup>26</sup> However, DGXV found that most Member States were not interested in major copyright law reform at the EU level.<sup>27</sup> DGXV’s next step was the consideration of the questions of whether and to what extent digitalization required harmonization of copyright law. It published a second Green Paper on Copyright in July 1995

---

20. *Commission White Paper on Growth, Competitiveness, Employment: The Challenges and Ways Forward into the 21st Century*, COM (1993) 700 final (Dec. 5, 1993).

21. Bangemann Group, *Europe and the Global Information Society: Recommendations of the High-Level Group on the Information Society to the Corfu European Council*, Bulletin of the European Union, Supplement No. 2/94 (May 26, 1994) [hereinafter Bangemann Report].

22. The Corfu European Council invited the Commission to prepare an Action Plan to implement the recommendations in its initial report. See European Comm’n, *Europe’s Way to the Information Society: An Action Plan*, COM (1994) 347 final (July 19, 1994) [hereinafter *Europe’s Way*].

23. See Frederick W. Pfister, *Net Neutrality: An International Policy for the United States*, 9 SAN DIEGO INT’L L.J. 167, 184–85 (2007). Interestingly, although the U.S. government is traditionally perceived as supportive of laissez-faire more than the EU, the government took a more active role in shaping the NII. However, despite the initial market zeal of the Bangemann report, there had been an increasing realization in a need for state involvement to ensure the successful development of the Information Society. See William H. Dutton et al., *The Politics of Information and Communication Policy: The Information Superhighway*, in INFORMATION AND COMMUNICATION TECHNOLOGIES: VISIONS AND REALITIES 400–03 (William H. Dutton ed., 1996).

24. Comm’n of the European Communities, *Commission Green Paper on Copyright and the Challenge of Technology: Copyright Issues Requiring Immediate Action*, COM (1988) 172 final (June 7, 1988).

25. Comm’n of the European Communities, *Follow-up to the Green Paper: Working Programme of the Commission in the Field of Copyright and Neighbouring Rights*, COM (1990) 584 final.

26. See *Europe’s Way*, *supra* note 22, at 5.

27. DGXV distributed a questionnaire and held a public hearing. See DGXV/E/4 Questionnaire on Copyright and Related Rights in the Information Society, June 2, 1994 (exploring the following six subjects: evolution of the superhighways; scope of the information infrastructure; identification and clearance of rights; choice of legal regime; review of existing regimes; other relevant issues).

that reflected its findings.<sup>28</sup> Similar to the first 1988 copyright Green Paper, the 1995 Green Paper overtly favored the economic interests of right holders over rights of authors and users of information products, relying on harmonization and the removal of barriers to the internal market as a justification.<sup>29</sup>

During the first half of 1996, the Commission held a public hearing and a conference and as a result decided to adopt its follow-up communication in November 1996, outlining four areas requiring legislative action: (1) the legal protection of anti-copying systems; (2) the distribution right; (3) the reproduction right; and (4) the communication to the public right of "on-demand" services.<sup>30</sup> Additionally, the Commission identified some areas for further studying: the broadcasting right; rights management; moral rights; and questions of jurisdiction, applicable law, and enforcement.<sup>31</sup> This follow-up communication aimed to remove obstacles to trade and competition in copyrighted products.

Last, during the 1990s, there had also been a consideration of the question whether digitalization required copyright law reform at the World Intellectual Property Organization (WIPO).<sup>32</sup> As early as 1989, WIPO had already started working on a protocol to the Berne Convention that would adapt it to the digital era. Additionally, and also in response to challenges posed by new technologies, WIPO introduced three new treaties in its 1996 diplomatic conference: a treaty on literary and artistic works (Copyright Treaty), a treaty on the rights of performers and phonogram producers (New Instrument), and a treaty on databases.<sup>33</sup> While WIPO, the Unit-

---

28. Comm'n of the European Communities, *Green Paper on Copyright and Related Rights in the Information Society*, at 66, COM (1995) 382 final (July 27, 1995) [hereinafter 1995 Green Paper].

29. See generally *id.*; Commission *Green Paper on Copyright and the Challenge of Technology: Copyright Issues Requiring Immediate Action*, COM (1988) 172 final (June 7, 1988) [hereinafter 1988 Green Paper].

30. Comm'n of the European Communities, *Follow-up to the Green Paper on Copyright and Related Rights in the Information Society*, at 15-27, COM (1996) 568 final (Nov. 20, 1996).

31. *Id.* at 20-28.

32. See WORLD INTELLECTUAL PROP. ORG. [WIPO], WIPO WORLDWIDE SYMPOSIUM ON THE INTELLECTUAL PROPERTY ASPECTS OF ARTIFICIAL INTELLIGENCE, STANFORD UNIVERSITY (MAR. 25-27, 1991) (1991). For an examination of WIPO's digital agenda, see Mihály Fiscor, *Towards a Global Solution: The Digital Agenda of the Berne Protocol and the New Instrument: The Rorschach Test of Digital Transmissions*, in THE FUTURE OF COPYRIGHT IN A DIGITAL ENVIRONMENT 111 (P. Bernt Hugenholtz ed., 1996).

33. Provisional documents (Aug. 30, 1996), which set out the basic proposals for the WIPO conference, promoted an agenda in favor of right holders. The late release date of

ed States, and the EU all supported these draft treaties,<sup>34</sup> only the first two treaties were eventually adopted.<sup>35</sup>

### III. THE FORMULATION OF THE DATABASE DIRECTIVE

Part III proceeds to discuss the formulation of the Database Directive in the EU. Section III.A begins with a discussion of protection mechanisms for compilations that pre-date the Database Directive. The discussion points to the very different ways in which EU Member States protected compilations. A discussion of these differences explains many of the challenges introduced during the formulation process of the Directive. Section III.B then discusses the reasons for the emergence of database protection as a policy issue, pointing to different factors that suggest EU copyright policy was not motivated only by harmonization aspirations and consideration of the subject in the international arena, but also by a desire to regulate the level of investment in database production in the community and internationally.

Section III.C discusses the formulation process of the Database Directive. It discusses different phases in the consideration of the Directive, shedding light on important developments that occurred during each phase while also critically discussing each phase. The discussion also highlights some of the characteristics of EU copyright law policy. Specifically, Section III.C discusses why the EU settled on the dual copyright/*sui generis* model, showing that it was probably an inevitable approach to take given the differences in the copyright regimes of the Member States. It then turns to discuss the different phases of the consideration of the Directive: pre-proposal discussions, the Commission Database Proposal, the Council Common Position, and the adopted Database Directive. Finally, Section III.D discusses the Database Directive in its final form. It focuses on

---

these documents also limited the time available for consultation and debate with users groups, among others. The provisional documents were as follows: Basic Proposal for the Substantive Provisions on the Protection of Literary and Artistic Works, WIPO Doc. CRNR/DC/4 (Aug. 30, 1996) (prepared by Director General) – this was supposed to be the new protocol to the Berne Convention; Basic Proposal for the Substantive Provisions on the Protection of the Rights of Performers and Producers of Phonograms, WIPO Doc. CRNR/DC/5 (Aug. 30, 1996) (prepared by the Chairman of the Committee of Experts) – this was supposed to be the new instruments; Basic Proposal for the Substantive Provisions of a Treaty on Databases, WIPO Doc. CRNR/DC/6 (Aug. 30, 1996) (prepared by the Chairman of the Committee of Experts).

34. Pamela Samuelson, *The US Digital Agenda at WIPO*, 37 VA. J. INT'L L. 369, 387–88 (1997).

35. The database proposal was dropped, but it is still on WIPO's agenda for future action.

treatment of databases and the actual implementation of the dual copyright/*sui generis* model.

The discussion below serves a few goals. First, it points to the issues that proved controversial during the formulation process as well as the main changes that ultimately occurred, exposing the depth of exploring the subject by the EU. Second, it reveals that there had not been a serious debate at all levels of the EU institutions regarding whether protection was actually needed. Third, it shows that the proposed Directive was actually strengthened during the formulation process. Fourth, and most importantly, it illustrates that during the consideration of the Directive the database industry was over-represented whereas users and consumer groups were under-represented, a fact that contributed to the adoption of an imbalanced property right in data. This last point clearly shows that the lobby within the EU could effectively mobilize, in part due to concentration interests in the database industry (led mainly by United Kingdom firms). In turn, efforts to organize user interest groups were presumably more complex or cumbersome.<sup>36</sup>

#### A. The State of Protection for Compilations Prior to the Database Directive

Copyright protection for databases or compilations in the EU Member States in the years preceding the Directive's enactment could be divided into three main categories: (1) Member States in which the threshold of originality for compilation copyright was quite low;<sup>37</sup> (2) Member States in which the threshold of originality was a fairly high threshold in the selection and arrangement of the compilation;<sup>38</sup> and (3) Member States that provided protection outside the copyright law framework.<sup>39</sup> Many Member States fell into the second category.<sup>40</sup> Examples of countries belonging to the first, very limited, category include the United Kingdom and Ire-

---

36. In the United States, however, database protection beneficiaries were less concentrated than in Europe. There actually existed a split in the database industry concerning database protection. See *Economic Dimension*, *supra* note 16, at 106–107.

37. 1988 Green Paper, *supra* note 29, at 212 ¶¶ 6.4.3-6.4.4; see Jaap H. Spoor, *Copyright Protection and Reverse Engineering of Software: Implementation and Effects of the EC Directive*, 19 U. DAYTON L. REV. 1063, 1065–66 (1994) (reviewing the originality requirements for copyrights in general prior to enactment of any EU Directive).

38. 1988 Green Paper, *supra* note 29, at 211-212 ¶¶ 6.4.1-6.4.3.

39. *Id.* at 213 ¶¶ 6.4.5-6.4.6.

40. See, e.g., MARK J. DAVISON, *THE LEGAL PROTECTION OF DATABASES* 16 (2003) (suggesting that “[t]he standard of originality in many European countries also requires an element of intellectual creativity”).

and,<sup>41</sup> because Anglo-Irish common law incorporated a “sweat of the brow” doctrine for extending copyright protection to unoriginal compilations of factual information, finding labor a basis for copyright protection.<sup>42</sup>

In keeping with the “author’s right” approach that prevails throughout most of Continental Europe and defines originality as an expression of the author’s personality, countries in the second category included most *droit d’auteur* regimes.<sup>43</sup> For example, in Germany, apart from the protection offered under “Kleine Münze” (meaning small change) to certain works that exhibit very limited creativity such as simple maps,<sup>44</sup> compilations were eligible for copyright protection only when they exhibited sufficient intellectual creativity in their selection and arrangement. However, these states’ information compilers were not left entirely unprotected because many states provided a remedy against wholesale copying through unfair competition laws regardless of the originality of the work.<sup>45</sup>

Finally, since the early 60s Nordic countries and The Netherlands provided protection under copyright law.<sup>46</sup> The copyright acts of all Nordic countries—Denmark, Finland, Iceland, Norway and Sweden—contain a “catalogue rule” provision.<sup>47</sup> Such “catalogue rule” provisions expressly protect non-original compilations of data, such as catalogues, tables, and similar compilations, provided they contain many items.<sup>48</sup> For example,

---

41. *Id.* at 126–29 (discussing Ireland), 143–58 (discussing the U.K.).

42. *Id.*; see also Spoor, *supra* note 37, at 1066. In a 1959 case concerning the use by a pools promoter of the football fixtures list, the Judge concluded that because the Football League through its employees had expended “skill, labour, times, judgment and ingenuity” in preparing the fixture list, it was entitled to copyright protection. *Football League Ltd. v. Littlewoods Pools Ltd.*, (1959) Ch. 637, 654.

43. DAVISON, *supra* note 40, at 16–17; see also Silke von Lewinski, *Copyright in Central and Eastern Europe: An Intellectual Property Metamorphosis*, 8 FORDHAM INTEL. PROP. MEDIA & ENT. L.J. 39 (1997).

44. Silke von Lewinsky, *Protection of and vis-à-vis Databases: Germany*, in COPYRIGHT IN CYBERSPACE 480–481 (Marcel Dellebeke ed., 1996); see also P. Bernt Hugenholtz, *Protection of Compilations of Facts in Germany and the Netherlands*, in PROTECTING WORKS OF FACTS: COPYRIGHT, FREEDOM OF EXPRESSION AND INFORMATION LAW 59, 62–63 (Egbert J. Dommering & P. Bernt Hugenholtz eds., 1991) [hereinafter Hugenholtz, *Compilations of Facts*].

45. See DAVISON, *supra* note 40, at 103. For discussion of specific countries see *id.* at 115–16 (France), 123–24 (Germany).

46. *Id.* at 59 n.41 (Nordic countries), 133 (The Netherlands).

47. Gunnar W.G. Karnell, *The Nordic Catalogue Rule*, in PROTECTING WORKS OF FACTS: COPYRIGHT, FREEDOM OF EXPRESSION AND INFORMATION LAW 67 (Egbert J. Dommering & P. Bernt Hugenholtz eds., 1991); see also Hugenholtz, *Compilations of Facts*, *supra* note 44, at 59, 63–64.

48. Karnell, *supra* note 47, at 67.

the Danish Copyright Act allowed protection for tables, catalogues, and other works that compile information.<sup>49</sup> The information compiled need not exhibit originality and the protection is only against reproduction, which lasts for only ten years or under certain circumstances for fifteen years.<sup>50</sup>

These differences between Member States are not surprising given the different traditions within the EU of author's rights and copyright. Only by the mid-1980s, did debate emerge as to copyright law protection for electronic databases.<sup>51</sup> However, the EU started introducing initiatives regarding database protection only once the economic importance of such products in the market for information products increased.<sup>52</sup> Providing incentives for the creation of databases was part of the EU's general agenda of removing obstacles to the creation of a European information services market.<sup>53</sup>

## B. The Emergence of Database Protection As a Policy Matter

The database protection debate is one of the most disputed issues in intellectual property law and thus serves as a good case study for examining the EU's copyright policy. By analyzing how the Database Directive first emerged as a policy issue and then was formulated, as well as considering its passage in light of EU copyright harmonization efforts, we can see the EU's intellectual property law policy-making process.

A number of factors contributed to the emergence of database protection as a policy matter and provided a catalyst for harmonization in this field. First, the rapid expansion of the Internet raised the EU's awareness of the growth in the magnitude of information created and processed every

---

49. 1988 Green Paper, *supra* note 29, at 213.

50. *See* Karnell, *supra* note 47, at 68–69.

51. *See, e.g.*, Comm'n of the European Communities, *The Establishment at Community Level of a Policy and a Plan of Priority Actions for the Development of an Information Services Market*, at 4–5 ¶ 3, COM (1987) 360 final (Sept. 2, 1987) [hereinafter *Establishment of Policy*]; 1988 Green Paper, *supra* note 29, at 207–217 ¶ 6; David P. Lewis, *Copyright Aspects of Databases*, 4 *COMPUTER L. & PRACTICE* 2 (1987).

52. *Establishment of Policy*, *supra* note 51, at 4–5 ¶ 3.

53. *See* 1988 Green Paper, *supra* note 29, at 207; Comm'n of the European Communities, *Proposal for a Council Directive on the Legal Protection of Databases*, at 16, COM (1992) 24 final (May 13, 1992) [hereinafter *Proposal for a Council Directive*] (“In view of the uncertainty and possible divergence of interpretation which surround the protection of databases at present, there is clearly a need to establish at least a basic harmonized framework.”).

year in the different fields of industry and commerce, and the important role of databases for the evolution of a market for information products.<sup>54</sup>

The debate gained worldwide prominence due to a number of initiatives that extended some protection to databases and contemplated even more extensive protection. Notably, the TRIPs Agreement<sup>55</sup> introduced minimum standards regarding copyright protection for databases.<sup>56</sup> Discussion in WIPO considered the provision of significantly broader intellectual property rights in databases than in the United States under *Feist*.<sup>57</sup> In August 1996, the WIPO distributed a draft text entitled “Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in respect of Databases” that introduced a *sui generis* right in databases.<sup>58</sup> Generally, this proposal protects the “substantial investment in the collection, assembly, verification, organization or presentation of the contents” of a database.<sup>59</sup> Additionally, it grants a “right to prohibit or authorize the utilization or extraction of” a substantial part of database contents.<sup>60</sup>

Second, the EU also expressed concerns about the “great imbalance in the level of investment in database creation both as between the Member States themselves, and between the Community and the world’s largest database-producing countries.”<sup>61</sup> The Commission was mainly concerned over the fragmented nature of the EU information market, which was the result of divergent policies within Member States.<sup>62</sup>

As the discussion, *infra*, illustrates, the relative weakness of the European information market stemmed from legal, technological, and linguistic

---

54. *Proposal for a Council Directive*, *supra* note 53, at Recitals 8–9.

55. TRIPs Agreement, *supra* note 14.

56. The TRIPs Agreement stated that “[c]ompilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations, shall be protected as such.” *Id.* at art. 10. Similar language is contained in the WIPO Copyright Treaty, art. 5, Dec. 20, 1996, S. Treaty Doc. No. 105-17, 36 I.L.M. 65 (1997).

57. WIPO, *Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases*, WIPO Doc. CRNR/DC/6, notes 1.08-1.10 (Aug. 30, 1996) [hereinafter *Draft Database Treaty*]; *supra* note 16 and accompanying text. The subject of database protection remains on the agenda of the WIPO Standing Committee on Copyright and Related Rights. In practice, however, WIPO’s proposal has been on hold, awaiting U.S. action. See WIPO, Standing Comm. on Copyright & Related Rights, *Report from 11th Session ¶ 9-23* (June 7–9, 2004), available at [http://www.wipo.int/edocs/mdocs/copyright/en/sccr\\_11/sccr\\_11\\_4.pdf](http://www.wipo.int/edocs/mdocs/copyright/en/sccr_11/sccr_11_4.pdf).

58. *Draft Database Treaty*, *supra* note 57, at art. 3.

59. *Id.* at art. 1.

60. *Id.* at art. 3.

61. *Proposal for a Council Directive*, *supra* note 53, at Recital 11.

62. *Id.* at Recital 1-4.

barriers between the Member States themselves and between Member States and the leading players in the international information market, especially the United States and Japan.

The different database protection regimes employed by Member States resulted in substantial legal barriers.<sup>63</sup> Technological barriers were the result of a few factors. First, the digital revolution originated in the United States rather than in Europe. The United States created the Internet, which quickly fostered a prospering Internet economy.<sup>64</sup> The United States also led the field of electronic commerce.<sup>65</sup> Therefore, the EU aspired to pass legislation that would close, or at least narrow, such gaps in legal protection.<sup>66</sup>

By the early 1990s, a debate emerged over the scope of protection available to databases. The Commission viewed the case law developing in this area as evidence that copyright was insufficient to protect electronic databases.<sup>67</sup> One notable decision, *Van Daele v. Romme*<sup>68</sup> (“*Van Daele*”)

---

63. See discussion *supra* Section III.A discussing the variance of Member State regimes.

64. Charles R. McManis, *Database Protection in the Digital Information Age*, 7 ROGER WILLIAMS U. L. REV. 7, 29–30 (2001). For an interesting discussion of the origin of the Internet, see generally TIM BERNERS-LEE & MARK FISCHETTI, *WEAVING THE WEB: THE ORIGINAL DESIGN AND ULTIMATE DESTINY OF THE WORLD WIDE WEB BY ITS INVENTOR* (1999).

65. See McManis, *supra* note 64, at 31.

66. See *id.*, at 30–31, 29–30 (2001); see also Mark Powell, *The EC Database Directive: A Revolutionary Means of Protecting Databases*, 2 INT’L COMPUTER LAW. ADVISOR 11 (1994) (“The relative weakness of the European electronics information market is as much due to linguistically fragmented markets and structural deficiencies (low installed base of CD-ROM drives and prohibitively expensive telecommunications services in particular) as to any legislative inadequacy.”); see generally U.N. Conference on Trade & Dev., Trade and Dev. Bd., *Can Electronic Commerce Be An Engine for Global Growth? Electronic Commerce and the Integration of Developing Countries and Countries with Economies in International Transition in International Trade*, U.N. Doc. TD/B/COM.3/23 (June 1, 1999) (providing figures showing the massive dominance of the United States in the field of e-commerce and business-to-business transactions). Interestingly, given the large number of languages within the EU, markets were less competitive with regard to database products because of elimination of economies of scale pertaining to databases’ production that is language-based. There is greater demand for English-based databases, most notably in the field of business transactions, because English is the preferred language in many contexts. Since software is vital for database creation, the U.S. gains an additional lead over the EU. Moreover, nearly 50% of Americans rely on Internet-based data, resulting in high demand for data in the English language. See McManis, *supra* note 64, at 31–32.

67. The Commission pointed explicitly to the *Feist* decision in its Explanatory Memorandum accompanying its proposal, expressing concern that under *Feist* “it may well be that electronic databases as well as collections in paper form, which do not meet the

in the Netherlands, raised concern regarding the adequacy of copyright protection for databases within the EU.<sup>69</sup>

In *Van Daele*, the plaintiff, a publisher of a leading Dutch language dictionary, brought a copyright infringement suit against the defendant, Mr. Romme, who had copied the 230,000 keyword entries in the plaintiff's dictionary into his electronic database and offered the database to crossword enthusiasts. The words were not stored alphabetically in the defendant's electronic database.<sup>70</sup>

The Utrecht District Court and the Amsterdam Court of Appeals both granted an injunction against the defendant, holding that the plaintiff's selection of keywords was original under Dutch compilation copyright law.<sup>71</sup> However, in January 1991 the Hoge Raad (Dutch Supreme Court)

---

test of originality, will be excluded from copyright protection regardless of the skill, labour, effort, or financial investment expended in their creation." *Proposal for a Council Directive*, *supra* note 53, at 17.

68. Romme/*Van Dale* Lexicografie, B.V., Hoge Raad der Nederlanden [HR] [Supreme Court of The Netherlands], 4 January 1991, *translated in* PROTECTING WORKS OF FACTS: COPYRIGHT, FREEDOM OF EXPRESSION AND INFORMATION LAW, app. I 93–96 (Egbert J. Dommering & P. Bernt Hugenholtz eds., 1991) [hereinafter Romme, PROTECTING WORKS OF FACTS].

69. Soon after *Van Daele*, the U.S. Supreme Court issued its seminal decision in *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991). The Court decided that a work must be original to the author and must possess at least some minimal degree of creativity. *Id.* at 345. The Court also decided that originality is constitutionally mandated, stemming from the Constitution's Copyright and Patent Clause that aims to promote the "Progress of Science and the useful Arts." *Id.* at 346. The Court's decision also clarified that its holding means that the "copyright in a factual compilation is thin" and that "subsequent compiler remains free to use the facts contained in another's publication in preparing a competing work." *Id.* at 349. For further analysis of *Feist* and its influence on international copyright protection, see *Proposal for a Council Directive*, *supra* note 53, at 17; W. Matthew Wayman, *International Database Protection: A Multilateral Treaty Solution to the United States' Database Dilemma*, 37 SANTA CLARA L. REV. 427, 439 (1997); Clive D. Thorne, *The Infringement of Database Compilations: A Case for Reform?*, 9 EUR. INTEL. PROP. REV. 331, 332 (1991) (arguing that although these judgments show a high level of protection potentially available to protect databases in the UK, "recent developments outside the United Kingdom show a preferable pattern for the protection of compilations," including the *Feist* judgment). *Feist* also contributed to the emergence of database protection as a policy matter at the international level in both the WIPO that considered the inclusion of database protection in a protocol to the Berne Convention, and in the WTO that discussed the harmonization of database protection during the negotiations of the TRIPs Agreement. The discussion of the matter internationally had probably also indirectly contributed to the emergence of database protection as a policy matter in the EU. See Wayman, *supra*, at 462–465.

70. Romme, PROTECTING WORKS OF FACTS, *supra* note 68, at 94.

71. *Id.* at 93–94.

overturned this decision, finding that a compilation of factual data does not meet the originality threshold and as a result cannot be protected by copyright:

[S]uch a collection is in itself no more than a number of factual data which do not in themselves qualify for copyright protection. This would be otherwise only if that collection were the result of a selection process expressing its *maker's personal views*.<sup>72</sup>

The court allowed the plaintiff to bring an appeal on its decision to the Court of Appeals in The Hague.<sup>73</sup> The Appellate Court confirmed the original Amsterdam Court of Appeals decision, finding that the plaintiff's selection of keywords for the dictionary reflected the personal view of the author regarding the nature of the current Dutch language.<sup>74</sup> Therefore, the dictionary was found eligible for copyright protection.<sup>75</sup>

Notwithstanding the reversal of the Supreme Court decision and the finding that the plaintiff's work was eligible for copyright protection, the case was significant in the EU because it raised awareness and signaled the limits of the protection available to factual compilations. The decision suggests that compilations of factual data do not meet the originality requirement.<sup>76</sup> Although this case does not necessarily provide evidence that there exists a problem that needs to be remedied, which I argued in a different Article,<sup>77</sup> the EU institutions failed to grasp and assess the limited importance of the decision. Rather than exploring whether such decisions proved the existence of market failure in the database sector, asking what other legal, technological, and economic mechanisms database producers employ in the information market, or determining whether such mechanisms provide sufficient incentives to create databases, the EU accepted as given the underlying economic assumptions of proponents of database protection and assumed that with no legal protection producers will have no incentive to produce databases.<sup>78</sup>

---

72. *Id.* at 95 (emphasis added).

73. See P. Bernt Hugenholtz, *Chronicle of The Netherlands: Dutch Copyright Law, 1990-1995*, 187 REVUE INTERNATIONALE DU DROIT D'AUTEUR 110 (2001), available at <http://www.ivir.nl/publications/hugenholtz/PBH-RIDA2000.doc>.

74. *See id.*

75. *See id.*

76. *Id.*

77. See generally *Economic Dimension*, *supra* note 16, at 139-141.

78. Council Directive 96/9, Legal Protection of Databases, art. 9, 1996 O.J. (L 77/20) (EC) [hereinafter Database Directive] ("The making of databases requires the investment of considerable human, technical and financial resources while such databases

In summary, database protection emerged as a policy matter due to a few discernible policy factors: the expansion of the internet that in turn raised awareness to the role of databases in the development of the information market; the imbalance in the level of investment in database creation amongst Member States and between the community and the major international players in the global database market; concerns over the existence of legal, technological, and linguistic barriers between Member States; the issuance of the The Netherlands' courts decisions in *Van Daele*; and the attention given to the subject in the international arena.

The discussion that follows explores the action taken by the EU regarding database protection in response to these factors by closely looking at the formulation process of the Database Directive.

### C. The Formulation Process: From Proposal to Directive

#### 1. *Settling on Sui Generis Protection*

By 1986 the Commission began considering the subject of database protection.<sup>79</sup> The Commission first focused on copyright law as a harmonization's path for databases protection. This emphasis on copyright law was partly due to the use of copyright law as a means for protecting computer software<sup>80</sup> as well as the international acceptance of the need to rethink the applicability of copyright doctrines to the computer environment.<sup>81</sup> However, it is possible that the Commission eventually pursued a dual copyright/*sui generis* approach in its database proposal, probably due to a lack of clarity in the Commission and the Member States regarding copyright protection for electronic databases.<sup>82</sup>

---

can be copied or accessed at a fraction of the cost needed to design them independently.”).

79. Comm'n of the European Communities, Questionnaire Relating to Provisions of National Copyright Laws of Specific Importance for the Operations of Computerised Information Systems, submitted to the DGXIII Legal Advisory Board (1986).

80. See discussion *supra* Section III.B.

81. See Intergovernmental Comm. of the Universal Copyright Convention, *Report of the Second Committee of Governmental Experts on Copyright Problems Arising from the Use of Computers for Access to or the Creation of Works*, UNESCO/WIPO/CEGO/II/7 (Aug. 13, 1982).

82. 1988 Green Paper, *supra* note 29, at 211-214 ¶ 6.4 (“The protection accorded to databases relates under existing national legislation and international conventions to the characteristics of the works stored therein, rather than to the database itself as a collection of information.”); see also Jean Paul Triaille, *Can You Copy Maps and the Facts They Contain?*, *MANAGING INTELL. PROP.* 31, 39-40 (Dec. 1991) (discussing the scope of database protection as well as the protection of the individual components of a database).

As a result, the Commission's 1988 Copyright Green Paper<sup>83</sup> was inconclusive regarding copyright protection for electronic databases.<sup>84</sup> Also, it is possible that inaction concerning database protection can be attributed to the discussion of the Software Directive proposal as well as lack of industry interest.<sup>85</sup> Discussion of the subject resumed only during an April 1990 public hearing.<sup>86</sup> Most interested parties supported only copyright solutions to the database protection problem.<sup>87</sup>

The Commission disclosed its plan to introduce a copyright solution in a 1990 follow-up document to the Green Paper.<sup>88</sup> There are a few ways to explain this initial anti-*sui generis* approach.<sup>89</sup> First, perhaps most interested parties believed that electronic databases already enjoyed meaningful copyright protection as well as other forms of legal protection similar to their old-fashioned print predecessors. For example, in the United Kingdom, the "sweat of the brow" doctrine was applicable to databases.<sup>90</sup> Second, the EU may have wished to avoid potential conflicts with non-EU states and the international community in general; a *sui generis* regime would have constituted a departure from accepted forms of protection un-

---

83. 1988 Green Paper, *supra* note 29.

84. *Id.* at 208 ¶ 6.3.1, 216 ¶ 6.7.1.

85. Comm'n of the European Communities, *Follow-up to the Green Paper: Working Programme of the Commission in the Field of Copyright and Neighbouring Rights*, COM (1990) 584 final, 18 [hereinafter *Working Programme*] ("The hearing confirmed that there was overwhelming support from right holders for protection of databases by means of copyright. No support was expressed for a '*sui generis*' approach.").

86. Prior to this public hearing, the Commission sent out a questionnaire to interested parties regarding database protection. See P. Gibbons, *EEC Hearing on Copyright and Databases*, NEWSIDIC No. 102 (European Association of Information Services, Amersfoort, Neth.), Aug. 1990, at 5.

87. Written Interview with Dr. Jens Gaster, Principal Administrator, Industrial Property Unit, DG Internal Market Directorate D2, European Commission (Dec. 8, 2008) [hereinafter *Gaster Written Interview*] (suggesting that "Anglo-Saxon EU interest supported this approach since they wished to maintain (optionally?) sweat of the brow copyright (which never existed in Continental Europe)" and also suggesting that "at the time the focus was on electronic on-line databases, thus mainly on UK industry"); Comm'n of the European Communities, *Follow-up to the Green Paper: Working Programme of the Commission in the Field of Copyright and Neighbouring Rights*, COM (1990) 584 final, 18 [hereinafter *Working Programme*] ("The hearing confirmed that there was overwhelming support from right holders for protection of databases by means of copyright. No support was expressed for a '*sui generis*' approach.").

88. *Working Programme*, *supra* note 85.

89. *Id.* at 18.

90. See Paula Baron, *Back to the Future: Learning from the Past in the Database Debate*, 62 OHIO ST. L.J. 879, 895 (2001) ("'Sweat of the brow theory' is the basis for copyright protection of compilations in the United Kingdom" and applies to databases via case law.).

der international intellectual property law. Last, it is possible that some of the Member States may not have understood the problems caused by the application of copyright law to a database's underlying content or raw data.<sup>91</sup>

Despite this initial anti-*sui generis* solution approach of both the Commission and interested parties, the Commission realized the possible economic contribution of the information sector to the economies of Member States.<sup>92</sup> In light of U.S. companies' continued dominance in this sector, the Commission wanted to improve the EU market share in providing electronic information services both in European markets and worldwide.<sup>93</sup> In order to achieve this goal, the Commission believed that guaranteeing adequate protection for electronic databases was essential.<sup>94</sup> Furthermore, the Commission realized that the marked differences between the different copyright traditions within the EU would yield significant differences in protection between Member States.<sup>95</sup> The dominance of author's rights countries in the EU prevented the Commission from adopting a common law "sweat of the brow" solution approach for database protection because such a solution would not meet the high originality threshold in author's rights countries.<sup>96</sup> However, introduction of a *sui generis* right

---

91. George Metaxas, *Protection of Databases: Quietly Steering in the Wrong Direction?*, 12 EUR. INTELL. PROP. REV. 227, 227 (1990) ("As a result, the first comments on behalf of the interested circles have, in some cases, simply missed the question (which only relates to the mode of compilation), wrongly confusing it with the issue of the protection of existing copyright works which are incorporated into a database.").

92. *Proposal for a Council Directive*, supra note 53, at ¶¶ 1.2, 2.1.1.

93. *Id.* at ¶ 1.1.

94. *Id.* at ¶¶ 1.1, 1.4, 2.2.11.

95. *Id.* at 15.

The legislation of the Member States probably serves to protect collections or compilations of works or other material by copyright either as works under Article 2(1) or as collections under Article 2(5) of the Berne Convention but it is unclear whether in all cases such protection extends to 'databases' and to electronic databases in particular . . . it is certainly the case that different results will be obtained in practice by the application of the legislation of the Member States to a given database.

*Id.* at 15.

96. See Opinion on the Proposal for a Council Directive on the Legal Protection of Databases, 1993 O.J. (C19/02), ¶ 2.6.2 [hereinafter Committee Opinion].

It would be wrong to compromise on the question of whether or not something should be protected by allowing a measure of short-term intellectual property protection with a compulsory license. It is preferable to take a decision on whether something qualifies for protection and, if so, then to grant intellectual property protection of a high standard.

bypassed this problem. In addition, the hurdles faced in reaching an agreement concerning the originality level during negotiations for the Software Directive also contributed to the realization of the Commission that a *sui generis* solution approach would avoid potential conflicts.<sup>97</sup> Finally, the significant differences between old-fashioned print compilations and electronic databases, and the resulting inability to accommodate the latter through copyright law, motivated the Commission to promote a *sui generis* approach.<sup>98</sup>

## 2. *Pre-Proposal Discussions*

The Commission began its work on preparing a proposal pertaining to database protection by May 1991, realizing that additional measures were necessary and considering different forms of protection such as contracts, unfair competition, neighboring rights, and *sui generis* right.<sup>99</sup> In August 1991, the Commission released a draft proposal for a directive for internal consultation within the Commission. This initial consultation involved a few Commission Directorates General (DG).<sup>100</sup> Naturally, different Directorates were involved to different degrees. First, Directorates General 1 (DG1) (external economic relations), Directorates General X (DGX) (audio-visual, information, communication, and culture), and Directorates General XXIII (DGIII) (enterprise policy, distributive trade, tourism, and

---

*Id.*

97. International copyright agreements, like most national copyright laws, lack explicit definitions of originality. *See, e.g.*, Berne Convention, *supra* note 13. These agreements also provide no definition of what constitutes a work either in terms of quantity or quality. The originality approach adopted in the Software Directive was whether the work was the “author’s own intellectual creation.” Council Directive 91/250/EEC, On the Legal Protection of Computer Programs, 1991 O.J. (L 122/42) (EC) [hereinafter Software Directive]. For a thorough analysis of the Software Directive see BRIDGET CZARNOTA & ROBERT J. HART, LEGAL PROTECTION OF COMPUTER PROGRAMS IN EUROPE: A GUIDE TO THE EC DIRECTIVE 43–45 (1991).

98. Metaxas, *supra* note 91, at 234.

One can only hope that, eventually, the inevitability of a *sui generis* solution for databases will gradually be appreciated after all and the tide will be reversed. We may then come to terms with the unpalatable but inevitable truth: copyright provisions cannot be stretched infinitely in order to reach the parts other intellectual property rights cannot reach.

*Id.*

99. Written Interview with Mr. Jean-Paul Triallie, Partner, De Wolf & Partners, Brussels, Belgium, (Dec. 10, 2008).

100. Telephone Interview with Dr. Jens Gaster, Principal Administrator, Industrial Property Unit, DG Internal Market Directorate D2, European Commission (Apr. 18, 2008) [hereinafter Gaster Interview]; *see also Proposal for a Council Directive, supra* note 53.

cooperation) supported the draft of the proposal.<sup>101</sup> However, DG1 also expressed concerns regarding some of the proposal's elements regarding the unfair extraction right and the reciprocity provision because they constitute departure from acceptable international norms under the Berne Convention, namely the reciprocity provision and the unfair extraction right.<sup>102</sup> DGX and DGXXIII also found the text of the draft proposal complex.<sup>103</sup> Additionally, Directorates General IV (DGIV) (competition) and DGXIII expressed concerns regarding the potential anti-competitive effects of the draft proposal, and continued to be involved on this issue throughout the formulation process.<sup>104</sup>

Last, DGXIII was very involved in the different phases of the formulation process and was mainly disturbed by the vagueness of certain definitions employed in the proposal.<sup>105</sup>

As a result of these consultations, the draft proposal was amended and the Commission adopted the draft in January 1992<sup>106</sup> and in May 1992 it was presented to the Council.<sup>107</sup> It is unclear whether there was any involvement of external entities such as Member States, other EU institutions, or others in drafting the proposal.<sup>108</sup> After 1992, however, it is clear that different interest groups, such as the information industry and other industries, became more involved.<sup>109</sup> It is possible that because the draft proposal pertained to a specialized market, interest groups' involvement was minimal at first.<sup>110</sup> Furthermore, the EU information industry was dominated by the United Kingdom market, and was also in its infancy and relatively small.<sup>111</sup> Most interested parties were therefore inexperienced

---

101. Gaster Interview, *supra* note 100.

102. *Id.*

103. *Id.*

104. *Id.* DGIV had commissioned its own study on copyright and information. See VINCENT PORTER, COPYRIGHT AND INFORMATION: LIMITS TO THE PROTECTION OF LITERARY AND PSEUDO-LITERARY WORKS IN THE MEMBER STATES OF THE EUROPEAN COMMUNITIES, A REPORT PREPARED FOR THE COMMISSION OF THE EUROPEAN COMMUNITIES (DG IV) (1992).

105. Gaster Interview, *supra* note 100.

106. *Proposal for a Council Directive*, *supra* note 53; see also Commission Information Note, *Fighting International Piracy of Databases: European Commission Proposes to Harmonize Legal Protection in the Community*, Jan. 29, 1992; Andrew Hill, *EC Agrees Legal Safeguards for Electronic Databases*, FIN. TIMES, Jan. 30, 1992, at 2.

107. *Proposal for a Council Directive*, *supra* note 53.

108. Gaster Interview, *supra* note 100.

109. *Id.*

110. *Id.*

111. *Id.*; DAVISON, *supra* note 40, at 61 (suggesting that at the time, the UK controlled 60% of the European community database market share).

and could not fully appreciate what was at stake. Last, the progress achieved in the international arena at both WIPO and the WTO regarding copyright protection for databases in the Berne Convention and the TRIPs Agreement, respectively, also probably contributed to the feeling that the proposal dealt inadequately with a major problem.<sup>112</sup>

### 3. *The Commission Database Proposal*

The proposal aimed to provide harmonized and stable legal protection to electronic databases as well as encourage investment in the European information industry.<sup>113</sup> Specifically, the introduction of the most innovative *sui generis* right was meant to “create a climate in which investment in data processing can be stimulated and protected against misappropriation.”<sup>114</sup> This proposal, however, was not the product of a real need or a response to a problem of piracy in the database sector; neither a need nor a piracy problem were shown, and the issues were not even raised during the formulation process.

While not stating so explicitly, the proposal aimed to provide a comparative advantage to database producers from Europe and to exert pressure on competitive rivals, especially the U.S. producers.<sup>115</sup> The methods the proposal employed reflect these aims. The most significant features of the proposal were the definition of the term “database,” the innovative copyright/*sui generis* dual approach to protection, the compulsory license provision, and the provision of protection on reciprocal basis.

Article 1(1) of the proposal defined the term “database” as “a collection of works or materials arranged, stored and accessed by electronic means, and the electronic materials necessary for the operation of the database.”<sup>116</sup> The definition did not cover non-electronic databases<sup>117</sup> or “any computer programme used in the making or operation of the data-

---

112. Gaster Interview, *supra* note 100.

113. This Section provides a summary of the main features of the Commission’s proposal that are required for the purposes of this Article. For a more comprehensive legal analysis of the proposal’s text, see generally Peter F. Kunzlik, *Proposed EC Council Directive on the Legal Protection of Databases*, 8 COMPUTER L. & SECURITY REPORT 116 (1992); Michael Pattison, *The European Commission’s Proposal on the Protection of Computer Databases*, 14 EUR. INTELL. PROP. REV. 113 (1992); Charles Oppenheim, *Copyright, Controversy and Compulsory Licenses*, INFO. WORLD REV., Mar. 6, 1992.

114. *Proposal for a Council Directive*, *supra* note 53, at 25 ¶ 3.2.7–8.

115. *Id.* at 2 ¶ 1.1 (suggesting that about “one quarter of the world’s accessible on-line databases [were] of European origin compared with the U.S. share of the world market of 56%”).

116. *Id.* at 66 art. 1(1).

117. *Id.* at 61 Recital 19, 67 art. 2(2).

base.”<sup>118</sup> As the discussions, *infra*, concerning the development of the text of the Directive illustrate, this definition proved problematic as a result of the exclusion of non-electronic databases and the breadth of this definition.<sup>119</sup>

Article 2(1) pertained to the copyright element of the dual approach, providing protection to “collections within the meaning of Article 2(5) of the Berne Convention.”<sup>120</sup> The level of originality was adopted from the Software Directive,<sup>121</sup> and provided that the selection and arrangement of the collection “constitutes the author’s own intellectual creation.”<sup>122</sup> This protection does not extend to the underlying content of the database.<sup>123</sup> This originality threshold introduced problems,<sup>124</sup> mainly because it was not in line with the originality threshold in the U.K., for example, that endorsed a “sweat of the brow” regime.<sup>125</sup>

Article 2(5) introduced the *sui generis* element of the dual approach as a “right for the maker of a database to prevent the unauthorised extraction or re-utilisation, from that database, of its contents, in whole or in substantial part, for commercial purposes.”<sup>126</sup> It also provided that the only databases that are ineligible for copyright protection can be protected by the *sui generis* right.<sup>127</sup>

Article 3 provides that the database’s author is either the person who created it or, where recognized by the Member State’s legislation “the legal person designated as the rightholder by that legislation.”<sup>128</sup>

Article 5 lists the exclusive rights of the author and Article 6 provides the exceptions to copyright.<sup>129</sup> Specifically, Article 6(2) deals with the

---

118. *Id.* at 66 art. 1(1).

119. *See* Pattison, *supra* note 113, at 114–15; *Proposal for a Council Directive, supra* note 53, at Recital 17–18; *see also* Powell, *supra* note 66, at 12.

120. *Proposal for a Council Directive, supra* note 53, at art. 2(1).

121. Software Directive, *supra* note 97, at art. 1(3).

122. *Id.*

123. *Proposal for a Council Directive, supra* note 53, at 61 Recitals 20–21.

124. *See* Jean Hughes & Elizabeth Weightman, *EC Database Protection: Fine Tuning the Commission’s Proposal*, 5 EUR. INTELL. PROP. REV. 147, 148–49 (1992).

125. *Id.* at 149.

126. *Proposal for a Council Directive, supra* note 53, at art. 2(5).

127. *Id.* at Recitals 28–29. This provision proved problematic for some interested parties who argued that the *sui generis* protection should be cumulative. *See, e.g.*, British Computer Society, *Intellectual Property Committee Comments on the EC Proposal for a Council Directive on the Legal Protection of Databases*, 9 COMPUTER L. & SEC. REP. 4 (1993).

128. *Proposal for a Council Directive, supra* note 53, at art. 3(1).

129. *Id.* at 47 ¶ 6.1; *see also id.* at 24–25.

ability to contractually deny exceptions.<sup>130</sup> Article 6(3) provides that these exceptions “are without prejudice to any rights subsisting in the works or materials contained in the database.”<sup>131</sup> Member States are also required by Article 7 to apply the same exceptions that they provide in their domestic copyright legislation with regard to illustrations and brief quotations for teaching purposes.<sup>132</sup> The term of copyright protection under Article 9(1), would be identical to that provided for literary works.<sup>133</sup>

Article 8 addressed the potential anti-competitive effects of the *sui generis* right and provided a compulsory license of database content when the database’s underlying materials or works contained in a database “cannot be independently created, collected or obtained from any other source,”<sup>134</sup> and when “the database is made publicly available by a public body which is either established to assemble or disclose important information pursuant to legislation, or is under a general duty to do so.”<sup>135</sup>

Due to the centrality of competition law in the EU the Commission proposal’s compulsory license provisions were the most controversial aspects of the proposal.<sup>136</sup> The major directorate generals who were involved in crafting the EU’s copyright policy were especially concerned regarding the Commission’s proposals’ anti-competitive effects.<sup>137</sup> Thus, competition law played a significant role in setting the boundaries of copyright law in Europe.<sup>138</sup>

---

130. *Id.* at art. 6(2).

131. *Id.* at art. 6(3).

132. *Id.* at Recital 26, art. 7.

133. Following the adoption of the Duration Directive, the duration of this protection became life of the author plus 70 years. *See* Council Directive 93/98, Harmonising the Term of Protection of Copyright and Certain Related Rights, 1993 O.J. (L 290/9) (EU) ¶ 11.

134. *Proposal for a Council Directive*, *supra* note 53, at art. 8(1).

135. *Id.* at art. 8(2).

136. *See* Pattison, *supra* note 113, at 118. Some clarification is provided in *Proposal for a Council Directive*, *supra* note 53, at Recital 33. However, the Explanatory Memorandum, using the example of the Stock market, states, “If the Stock Market refused to supply the figures to more than one applicant, remedies under competition rules might have to be sought to deal with that issue.” *Proposal for a Council Directive*, *supra* note 53, at 51 ¶ 8.1; *see generally* Kunzlik, *supra* note 113, at 118.

137. Thomas C. Vinje, *Symposium on U.S.-E.C. Legal Relations: Recent Developments In European Intellectual Property Law: How Will They Affect You and When?*, 13 J.L. & COM. 301, 302 (1994) (“Trends show that DGXV . . . tends to promote the broadest possible intellectual property protection without a great deal of consideration for the potential anti-competitive consequences of its proposals. DGIII, DGIV, and DGXIII . . . tend to be more concerned with the impact of competition on Commission proposals.”).

138. This is illustrated by *Radio Telefís Eireann v. Comm’n* (“Magill”), 1995 4 C.M.L.R. 718. In *Magill*, the ECJ imposed a compulsory license on three broadcasters

The remainder of Article 8 deals with exceptions to the *sui generis* right. These exceptions allow a lawful user to use “insubstantial parts”<sup>139</sup> of a database without the authorization of the rightholder.<sup>140</sup> The use of qualitative and quantitative criteria in the definition of the term “insubstantial part” creates uncertainty regarding possible liability. Commercial use of “insubstantial part” also requires acknowledgement of the source.<sup>141</sup> These exceptions apply only to the extent that they do not conflict with prior obligations or rights.<sup>142</sup>

As for the term of protection, Article 9(3) states that the *sui generis* right “shall expire at the end of the period of 10 years from the date when the database is first lawfully made available to the public.”<sup>143</sup> Additionally, “[i]nsubstantial changes to the contents of the database shall not extend the original period of protection of that database by the right to prevent unfair extraction.”<sup>144</sup>

Article 11 introduces the reciprocity provision and provides that the *sui generis* right will apply only to databases whose producers are either nationals of the Member States or who have habitual residence in the territory of the European community.<sup>145</sup> The EU Council, however, may conclude agreements to extend the application of the *sui generis* right to third countries,<sup>146</sup> but only “if such third countries offer comparable protection to databases produced by nationals of the Member States . . .”<sup>147</sup> Database producers, mainly from the United States, strongly criticized this principle of reciprocity as opposed to national treatment in relation to the *sui gene-*

---

after their refusal to grant the Irish publisher, Magill TV Guide Ltd., licenses to include their program information in its weekly program guide. The ECJ held that the broadcasters were abusing their dominant position under Article 86 of the EC Treaty. *See* Thomas C. Vinje, *Harmonising Intellectual Property Laws in the European Union, Past, Present and Future*, 17 EUR. INTELL. PROP. REV. 361, 374–76 (1995); *see also* Thomas C. Vinje, *The Final Word on Magill*, 17 EUR. INTELL. PROP. REV. 297, 303 (1995) (arguing that even after *Magill*, refusal to license intellectual property rights will, in the vast majority of cases, remain immune to attack under Article 86, but that the ECJ has preserved the flexibility to apply Article 86 to special circumstances such as those sometimes found in information technology cases).

139. *Proposal for a Council Directive*, *supra* note 53, at art. 1(3).

140. *Id.* at art. 8(4)–(5).

141. *Id.* at art. 8(4).

142. *Id.* at art. 8(6).

143. *Id.* at art. 9(3).

144. *Id.* at art. 9(4).

145. *Id.* at art. 11(1).

146. *Proposal for a Council Directive*, *supra* note 53, at art. 11(3).

147. *Id.* at 65 Recital 38.

*ris* right.<sup>148</sup> The EU had used such a reciprocity clause only once before in the directive proposal on data protection.<sup>149</sup> However, the United States participated in the same behavior when it enacted *sui generis* protection for semiconductor chip designs and conditioned protection of foreign chip designs on the grant of equivalent protection for U.S.-originated semiconductor chips.<sup>150</sup> The use of such a clause may have reflected the Commission's desire to provide a comparative advantage to European database producers over their U.S. competitors. Finally, Article 12(1) provides that the directive provisions are without prejudice to other legal provisions.<sup>151</sup>

In summary, this first Commission Proposal attempted to provide a modest solution to the lack of harmonization regarding database protection. It restricted the definition of a "database" to only electronic databases. The *sui generis* right given was limited in that it only affected extraction or re-utilization for commercial purposes. It did not apply if the database content was copyrightable. Further, it was subject to exceptions and compulsory licensing under certain circumstances. Such limited protection, however, eventually expanded to create a stronger property right in data.

#### 4. *The Proposal and Its Amended Text*

The Economic and Social Committee ("ECOSOC") prepared an opinion on the proposal in November 1992.<sup>152</sup> ECOSOC took the view that the Council should have the goal of having a strong database industry<sup>153</sup>

---

148. See McManis, *supra* note 64, at 45 ("The EU Database Directive seems designed to favor European database producers at the expense of their customers and non-EU competitors, and to pressure the rest of the world to create comparable protection."); see also U.S. PATENT AND TRADEMARK OFFICE, REPORT ON RECOMMENDATIONS FROM APRIL 1998 CONFERENCE ON DATABASE PROTECTION AND ACCESS ISSUES 22 (1998):

An American firm that does not enjoy protection under the EU Directive faces several possible competitive disadvantages. First and most obviously, its noncopyrightable database may be duplicated and remarketed by others. Second, European data sources looking for a firm to 'process' and market raw data will be more likely to enter into a contract with a European company that can guarantee protection of the database versus an American company that cannot.

149. European Communities, *Draft Proposal for a Council Directive Concerning the Protection of Individuals in Relation to the Processing of Personal Data*, at 41, COM (1990) 287 draft (Sept. 13, 1990).

150. See Semiconductor Chip Protection Act of 1984, Pub. L. No. 98-620, § 902(a)(2), 98 Stat. 3347 (1989) (codified as amended at 17 U.S.C. § 908(a) (2005)).

151. *Proposal for a Council Directive*, *supra* note 53, at 73, art. 12(1).

152. ECOSOC, *Information Note*, IND/451-322/192, Nov. 24, 1992.

153. *Id.* at ¶ 2.1.

and providing legal protection that can accomplish that goal.<sup>154</sup> It supported harmonization of the law concerning copyright protection for databases and recommended the introduction of a uniform standard of originality based on the “sweat of the brow” doctrine, mirroring the view of the UK information industry.<sup>155</sup> This opinion, however, did not significantly impact the formulation process in general or the proposal in particular because of the limited powers of the Economic and Social Committee and also because it is considered an unimportant institution in the EU.<sup>156</sup>

Additionally, the Commission and its officials engaged in discussions with the Legal Advisory Board (“LAB”), interested parties,<sup>157</sup> and Member States representatives.<sup>158</sup> However, it should be noted that the limited representation of information users continued throughout the entire formulation process.<sup>159</sup>

The European Parliament examined the Commission’s proposal upon the Council’s request and referred it to a committee in July 1992.<sup>160</sup> The

---

154. *Id.*

155. *Id.* at ¶ 2.2.- ¶ 2.7.

156. NEILL NUGENT, *THE GOVERNMENT AND POLITICS OF THE EUROPEAN UNION* 310 (1994) (arguing that sources of weaknesses of the ECOSOC “include the part-time capacity of its members, the personal rather than representational nature of much of its membership, and the perception by many interests that advisory committees and direct forms of lobbying are more effective channels of influence”). Simultaneously, several Member States discussed the proposal, but the extent of these discussions was dictated by each Member States’ share of the EU information market and their traditions regarding public consultations. Thus, the subject was naturally discussed extensively in the United Kingdom because of its significant share in the information market. Members of the U.S. database industry also debated and strongly criticized the proposal because its reciprocity provision imposed constrictive requirements on U.S. producers given the legal landscape after *Feist* and the constitutionalization of the originality requirement. Gail J. Hupper, Summary of Proceedings of the Forum on the European Community Database Directive (Dec. 9, 1992) (unpublished manuscript, on file with author).

157. Gaster Written Interview, *supra* note 87.

158. *Id.*

159. LEGAL ADVISORY BOARD, *LAB REPLY TO THE GREEN PAPER ON COPYRIGHT AND RELATED RIGHTS IN THE INFORMATION SOCIETY* (1995), available at <http://ec.europa.eu/archives/ISPO/legal/en/ipr/reply/reply.html>. *But see* Gaster Written Interview, *supra* note 87 (suggesting that “[l]obbying of all stakeholders interests (including anti-IP interests) was and is as intensive in Brussels as it is in Washington”).

160. The Energy Committee provided LEGA with a short opinion approving the proposal on May 26, 1993 with no suggested amendments. The Economic and Monetary Affairs and Industrial Policy Committee (“ECON”) started its work on Jan. 27/28, 1993, and adopted its opinion on June 2, 1993, which it passed on to LEGA. ECON proposed 3 amendments: clarifying the compulsory license provision (Article 8(1)), extending the *sui generis* term to 15 years with further protection possible after substantial change (Article

Committee on Legal Affairs and Citizens' Rights ("LEGA") was responsible for examining the proposal. In a March 1993 LEGA hearing,<sup>161</sup> experts and interested parties testified and they generally supported the proposal.<sup>162</sup> A few witnesses in the hearing, however, believed that rather than adopting the *sui generis* approach, the "sweat of the brow" doctrine should be extended to the rest of the EU.<sup>163</sup>

Ultimately, the 1993 Amendments failed to significantly affect the final version of the Directive, which rejected some of the key recommendations expressed in LEGA's opinion. The only key recommendation ultimately adopted was the proposal to extend the term of the *sui generis* right's term to fifteen years. This also signaled the Committee's intent to adopt a modest approach concerning database protection. However, the next substantive draft of the Directive went in a very different direction.

As a response, the Commission prepared a draft of the amended proposal.<sup>164</sup> The amended proposal accepted most of the amendments suggested by the Parliament, and strengthened the protection by extending the *sui generis* right to fifteen years<sup>165</sup> and introducing a "burden of proof" requirement on users so that any extraction and re-utilization of insubstantial parts of a database did not prejudice the owners' exclusive rights.<sup>166</sup> Most of the rejected amendments dealt with means that would have enhanced users' rights and changed definitions based on discussions at both the TRIPs Agreement and the WIPO negotiations.<sup>167</sup>

Within the Commission, however, there still remained concerns regarding the strengthening of protection by the two newly introduced provisions which placed the burden of proof on the user of insubstantial parts<sup>168</sup> and retroactively applied the protection offered by the proposal.<sup>169</sup>

---

9(3)), and a proposal for a new provision to review the implementation of the directive after the first 5 years and every 2 years thereafter (Article 13(3)).

161. Formal presentations with accompanying reports were made by many organizations, including the following organizations: The Spanish National Library, Madrid; Reed-Elsevier; Federation of European Publishers; European Space Agency; EUSIDIC; Dun & Bradstreet; and more.

162. *Id.*

163. Gaster Interview, *supra* note 100.

164. Comm'n of European Communities, Amended Proposal on the Legal Protection of Databases, 1993 O.J. 308/1 (Nov. 15, 1993) [hereinafter 1993 Amended Proposal].

165. *Id.* at art. 12(1).

166. *Id.* at art. 11(b).

167. European Commission, Monitoring of the Decision-Making Process Between Institutions, [http://ec.europa.eu/prelex/detail\\_dossier\\_real.cfm?CL=en&DosId=20478](http://ec.europa.eu/prelex/detail_dossier_real.cfm?CL=en&DosId=20478) (last visited Dec. 15, 2008) [hereinafter Procedure File] (summarizing the procedural developments regarding the legal protection of databases in the EU).

168. 1993 Amended Proposal, *supra* note 164, at art. 11(b).

Notwithstanding these concerns, the Commission adopted the amended proposal in October 1993<sup>170</sup> and presented it to the Council.<sup>171</sup>

5. *The Council Common Position.*

The most significant shift in the amended text, from a liability rule regime to an exclusive property rule regime, occurred during its consideration in the Council.<sup>172</sup> The Commission's proposal was presented to the Council in April 1992 and the copyright group of the Council was assigned to handle the issue. Interestingly, probably as a result of the hostility of the U.K. database industry towards the proposal, during the U.K. Council Presidency, the proposal had not been discussed.<sup>173</sup> During the Danish Council Presidency that followed, however, the Council group held several meetings to familiarize Member States' representatives with the proposal.<sup>174</sup>

Developments in Europe and the United States changed the policy environment of database protection during late 1993. Most notably, the TRIPs Agreement, which extended protection to electronic and non-electronic databases,<sup>175</sup> was completed in December 1993. After the Commission introduced the amended text to the Council in October 1993, discussions resumed.

---

169. *Id.* at art. 15(2).

170. Procedure File, *supra* note 167.

171. It should be noted that the database proposal had not generated much controversy or interest throughout the formal policy process described thus far. Although the United States and the U.K. information industries were interested due to their significant share in the global information market, most Member States showed very minimal interest in the proposal. Particularly in this first phase of the directive proposal, the information industry in other Member States was aware that the U.K. information industry was active in the proposal discussions. Therefore, they adopted the position that if the proposal was good enough for U.K. operators like Reuters, it would be good enough for them and so they did not feel any great necessity to become active themselves. *See* Responses to the Proposal from EC Committee of the American Chamber of Commerce [AMC-HAM], Jan. 1993. However, this indifferent approach of Member States, authors, and users' groups would subsequently be replaced with a greater awareness of the issue. This change in interest was due, in part, to legal scholars' involvement, which was triggered by the strengthening of the proposal. *See* William R. Cornish, *Protection of and vis-à-vis Databases*, in *COPYRIGHT IN CYBERSPACE* 435, 435-42 (Marcel Dellebeke ed., 1997).

172. J.H. Reichman & Pamela Samuelson, *Intellectual Property Rights in Data?*, 50 *VAND. L. REV.* 51, 84 (1997).

173. Gaster Interview, *supra* note 100.

174. *Id.*

175. TRIPs Agreement, *supra* note 14, at art. 10(2).

During the January 1994 discussion of the Council's first consolidated text<sup>176</sup> two key issues were heavily debated: the applicability of the directive to non-electronic databases and the desirability of adopting provisions pertaining to employer's economic rights over the work of their employees.<sup>177</sup>

The Greek Council Presidency referred the question of extension of the directive to non-electronic databases to the Committee of Permanent Representatives ("COREPER"), the highest ranking permanent organ of the Council of Ministers that prepares the agenda for the ministerial council for the EU meetings. The Council ultimately supported the extension in its February 1994 meeting. As a result, the Commission also decided that it would also support the extension of copyright protection for the original structure of a database to non-electronic databases.<sup>178</sup>

The Council prepared the second consolidated text focusing on the *sui generis* right after another meeting of its working group during March 1994.<sup>179</sup> At this stage, most Member States focused on specific provisions in the *sui generis* chapter, specifically supporting the removal of the 15-year term of protection and the user's burden of proof. The U.K., Ireland, and Germany, however, supported, a 50-year term for all databases.<sup>180</sup> The U.K., Ireland, Denmark, and Italy also opposed the Commission position that the *sui generis* right extends to databases' underlying raw data.<sup>181</sup> Unlike most Member States, however, Germany and France questioned the need for a *sui generis* right at all.<sup>182</sup> Additionally, France further opposed the introduction of compulsory licenses on public bodies, while the U.K. suggested that such licenses should be available only in situations where information arises from activities other than direct procurement.<sup>183</sup>

In April 1994, the Council held additional discussions on the second consolidated text, where many issues remained disputed. First, while

---

176. Council Document (EC) No. 5693/94 from Council Secretariat to Council Working Group on Copyright (Mar. 18, 1994) [hereinafter Council Document 5693/94].

177. Gaster Interview, *supra* note 100.

178. Gaster Interview, *supra* note 100; *see also* JENS-LIENHARD GASTER, DER RECHTSSCHUTZ VON DATENBANKEN: KOMMENTAR ZUR RICHTLINIE 96/9/EG MIT ERLÄUTERUNGEN ZUR UMSETZUNG IN DAS DEUTSCHE UND ÖSTEREICHISCHE RECHT 27-28 ¶¶ 17-19 (1999) [hereinafter DER RECHTSSCHUTZ].

179. Council Document 5693/94, *supra* note 176.

180. Gaster Interview, *supra* note 100; *see also* DER RECHTSSCHUTZ, *supra* note 178, at 156-57 ¶¶ 630-636.

181. Gaster Interview, *supra* note 100.

182. *Id.*

183. *Id.*

France, holding to their author's rights tradition,<sup>184</sup> supported the adoption of the high "author's own personality" threshold,<sup>185</sup> most Member States supported the adoption of TRIPs' lower level of originality. Second, the scope and coverage of the term "database" was still disputed as Portugal, Belgium, and France, specifically, opposed to the inclusion of non-electronic databases.<sup>186</sup> These Council group negotiations reflected the emergence of tensions between the Commission and Member States as well as the transition in Member States' activities from merely text clarifications to promotion of their own national agenda.

These disagreements brought the discussions in the Council working group to a stalemate in June 1994.<sup>187</sup> However, shortly thereafter new events advanced the debate. First, the commercialization of the Internet enhanced the centrality of the database protection debate among many Member States.<sup>188</sup> Second, the Commission started discussing general changes to copyright law to deal with challenges posed by new digital technologies.<sup>189</sup> Finally, the Bangemann group introduced its report and action plan to the Council in May 1994, emphasizing the urgent need to complete the Database Directive.<sup>190</sup>

Indeed, during the second half of 1994, the German Presidency made significant progress on these issues. Some agreements regarding the copyright chapter of the Directive were achieved during late 1994 as a result of the Council working group's consideration of issues identified by the working group. More meaningful progress was made concerning the *sui*

---

184. See *supra* Section III.A.

185. Gaster Interview, *supra* note 100; see also DER RECHTSSCHUTZ, *supra* note 178, at 51-53 ¶¶ 130-141, 119-121.

186. Gaster Interview, *supra* note 100; see also DER RECHTSSCHUTZ, *supra* note 178, at 37 ¶¶ 63-64.

187. An attempt to include the draft directive on the agenda for the COREPER and Internal Market Council in June 1994 was strongly opposed by some delegations.

188. Laurence Kaye, *The Proposed EU Directive for the Legal Protection of Databases: A Cornerstone of the Information Society?*, 17 EUR. INTELL. PROP. REV. 583 (1995).

189. In early June 1994, DGXV/E/4 released a questionnaire to interested parties on Copyright and Related Rights in the Information Society. In July a public hearing was organized and attended by many representatives of European companies and associations, EFTA representatives, other non-EU European countries, and most of the members of the Council copyright working group. The Commission later published the replies of these interested parties to the questionnaire. The issues that were raised during these discussions were similar to those discussed during the database protection debate.

190. See Bangemann Report, *supra* note 21, at 21.

*generis* chapter, however.<sup>191</sup> In order to advance the *sui generis* right debate, the Commission made a few proposals: (1) it would apply to databases whose content is already protected by copyright law;<sup>192</sup> and (2) its application to non-electronic databases would be conditioned upon showing investment in obtaining, verifying, or presenting their content. These proposals clearly strengthened the *sui generis* right, but disagreements over its application to non-electronic databases persisted.<sup>193</sup> For example, the German Council Presidency preferred to adopt a database protection regime resembling German law, which provides copyright protection together with some unfair competition rules.<sup>194</sup>

Despite of these difficulties, the German Presidency placed the proposal on the agenda of the COREPER December meeting at which it was agreed that the Council working group needed to more carefully define the scope and content of both the copyright and *sui generis* rights. During the following French Presidency the working group managed to make progress, revising the Council's fourth consolidated text during April 1995 in a manner addressing most of the open disputed issues.<sup>195</sup>

The most important changes to the directive text occurred in the beginning of May 1995 when most Member States agreed to the changes regarding the *sui generis* chapter. Ultimately, the majority of Member States supported the following changes:

- application of the directive to all types of databases;
- removal of the employer/employee economic rights provision;
- extension of the *sui generis* right to databases whose underlying materials are copyrightable;

---

191. These concerns included: adoption of the TRIPs definition for "database" and whether such protection would be applicable to the copyright and *sui generis* right; whether to include a provision regarding employer/employee economic rights; whether to employ the Berne Convention's provision on exceptions to copyright; whether to provide a compulsory license; what the term of protection should be; how to assess "substantial change" that justifies additional protection and whether there was a need to adopt a "data stamping" regime to enable differentiation of protectable elements in a database; whether to apply a reciprocity or national treatment regime to the *sui generis* right; and whether to provide retroactive protection to databases created before the directive. Gaster Interview, *supra* note 100; *see also* DER RECHTSSCHUTZ, *supra* note 178, at 121-22 ¶¶ 475-761.

192. Gaster Interview, *supra* note 100; *see also* DER RECHTSSCHUTZ, *supra* note 178, at 118 ¶¶ 458-459, 136, ¶ 547.

193. Gaster Interview, *supra* note 100.

194. *Id.*; *see also* DER RECHTSSCHUTZ, *supra* note 178, at 118-121 ¶¶ 460-461, 471-472.

195. Gaster Interview, *supra* note 100; *see also* DER RECHTSSCHUTZ, *supra* note 178, at 28 ¶¶ 18-19.

a 15-year term of protection for the *sui generis* right;  
 prohibition of reproduction of electronic databases for private purposes;  
 provision of exception to the *sui generis* right allowing extraction and/or re-utilization for private purposes of a substantial part of a non-electronic database;  
 provision of exceptions to the *sui generis* right allowing extraction and/or re-utilization for scientific research or educational purposes of a substantial part of the content of a database;  
 provision of protection on reciprocal basis;  
 and providing that contractual provisions contrary to articles that guarantee users' rights would be null and void.<sup>196</sup>

In May 1995, the COREPER discussed these agreements, and the remaining disagreements, and suggested the removal of the compulsory license provisions.<sup>197</sup> The Council, after making minor modifications to the text, approved this compromise and finally adopted the common position in July 1995.<sup>198</sup>

As the above discussion shows, the most significant changes to the original Commission proposal occurred during the Council negotiations under the German and French presidencies. A new property right in data therefore replaced the original modest approach to database protection. That new right was inherently imbalanced as it failed to consider adequately the interests of both database producers and users.<sup>199</sup>

#### 6. *The Adopted Database Directive*

The Commission accepted the Council's common position in September 1995 and forwarded its decision to the Parliament, beginning its second reading. The Commission strongly supported the common posi-

196. Gaster Interview, *supra* note 100; see also DER RECHTSSCHUTZ, *supra* note 178, at 28 ¶ 19; Council Common Position (EC) No. 20/95 of 30 Oct. 1995, 1995 O.J. (C 288/14) [hereinafter Common Position 20/95].

197. Gaster Interview, *supra* note 100.

198. Common Position 20/95, *supra* note 196.

199. Jens L. Gaster, *The New EU Directive Concerning the Legal Protection of Databases*, 20 FORDHAM INT'L L.J. 1129, 1142 (1997) [hereinafter *EU Directive*]:

In the course of controversial negotiations at the Council ongoing since the summer of 1994, the business law-like approach was strengthened and finally a right protecting substantial investments in databases was established. Hence a protection of the 'sweat of the brow' by a *sui generis* right was finally established and the dogmatic conflict between copyright and *droit d'auteur* in the area of databases was replaced by a dualistic concept.

tion, but expressed its preference that the compulsory license provision be reintroduced.<sup>200</sup> The Parliament's Committee on Legal Affairs and Citizen's Rights quickly supported the Council's common position, subject to a few minor amendments. In December 1995, the Parliament adopted the report and its amendments. In response, the Commission prepared a new text integrating these amendments<sup>201</sup> that was adopted unanimously in February 1996.<sup>202</sup> The Database Directive was formally enacted on March 11, 1996, with January 1, 1998 as an implementation date in Member States.<sup>203</sup>

In general, the discussion of the formulation process supports the contention that the Database Directive was not necessarily the product of an efficient and market-based understanding of the database industry, informed by public participation in the formulation process, because users and authors were under-represented in the process. Instead, it was a product of political influence, where misguided protectionist desires were implicitly aimed at gaining some competitive advantage over other international players in the information market, mainly the United States. Further, the resulting Directive completely failed to reflect consensus on either the need for, or the appropriate scope of, protection for a new *sui generis* form of intellectual property in data contained in a database. This is specifically highlighted by the fact that the EU itself has never satisfactorily explained the economic case for even creating such right.

---

200. COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, SEC (95) 1430 final, at 5 (Sep. 15, 1995):

This compromise text is of great importance in the context of the information society since most of the new products and services will operate from databases. The harmonized system as established by the eventual Directive will enable the doctrine of copyright to be brought closer to that of *droit d'auteur* in this crucial sector. This in itself will undoubtedly have a non negligible effect on the work of the international bodies responsible for harmonizing intellectual property law at the global level.

201. DAVISON, *supra* note 40, at 68. The most significant amendments were the addition of a requirement to indicate the source where a database is used for the purpose of illustration for teaching or scientific research (Article 6(2)(b)) and the replacement of the term "successors in title" with the term "rightholder" regarding the *sui generis* right's beneficiaries (Article 11(1)).

202. This took place during a meeting of the Agriculture Council. Press Release, Council of the European Union, Agriculture & Fisheries Council, Press: 36, NR: 5300/96, Other Decisions: Intellectual Property: Legal Protection of Databases (Feb. 26, 1996), available at [http://www.consilium.europa.eu/ueDocs/cms\\_Data/docs/pressData/en/agricult/001a0013.htm](http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/agricult/001a0013.htm).

203. Database Directive, *supra* note 78, at Art 1.6.

#### D. Treatment of Databases under the Database Directive

The European Union adopted the Database Directive on March 11, 1996.<sup>204</sup> The Directive protects both electronic and non-electronic databases.<sup>205</sup> It provides copyright protection for database structure<sup>206</sup> and a *sui generis* “data right” in database content. Member States were obligated to implement the Directive by January 1, 1998, and all twenty-five EU Member States ultimately adopted national measures implementing the Directive.<sup>207</sup>

The copyright portion of the Directive attempts to harmonize the scope of EU copyright protection, specifically standards of originality and exclusive rights (i.e., “restricted acts”) and exceptions to them. The standard adopted, stemming from the EU’s 1991 directive on the protection of computer programs,<sup>208</sup> provides copyright protection only to databases that, “by reason of the selection or arrangement of [its] contents, constitute the author’s own intellectual creation.”<sup>209</sup> Under the Directive, the copyright owner is entitled to certain exclusive rights or “restricted acts:” reproduction, adaptation, distribution, and display or performance of the database content to the public.<sup>210</sup>

---

204. *Id.*

205. *See id.* at art. 1(1), Recital 14. The term “database” is defined in the Directive as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.” *Id.* at art. 1(2). Explicitly excluded from protection under the Directive are “computer programs used in the making or operation of databases accessible by electronic means.” *Id.* at art. 1(3).

206. *See id.* at art. 3.

207. Germany, Sweden and the United Kingdom met the implementation deadline (1 Jan. 1998); Austria and France adopted in 1998 laws whose provisions apply retroactively as of 1 Jan. 1998; Belgium, Denmark, Finland, and Spain implemented in 1998; Italy and The Netherlands implemented in 1999; Greece and Portugal implemented in 2000; Ireland and Luxembourg implemented in 2001; the 10 new Member States (Cyprus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia and Slovenia) implemented between 1999 and 2003. They had an obligation to implement before the date of accession, i.e., 1 May 2004. *See* COMM’N OF THE EUROPEAN COMMUNITIES, DG INTERNAL MARKET AND SERVICES WORKING PAPER, FIRST EVALUATION OF DIRECTIVE 96/9/EC ON THE LEGAL PROTECTION OF DATABASES 11 (2005) [hereinafter WORKING PAPER], available at [http://ec.europa.eu/internal\\_market/copyright/docs/databases/evaluation\\_report\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/databases/evaluation_report_en.pdf).

208. Software Directive, *supra* note 97, at art. 1(3).

209. Database Directive, *supra* note 78, at art. 3(1). This standard reflects an attempt to require a level of originality that is neither too low, like the standard in the United Kingdom, The Netherlands, and Ireland, nor too high, such as the standard under German law. *EU Directive*, *supra* note 199, at 1136, 1141.

210. Database Directive, *supra* note 78, at art. 5. The Directive only covers economic rights under copyright; moral rights are beyond the scope of the Directive. *Id.* at Recital

The *sui generis* right provides protection for qualitatively and/or quantitatively “substantial investment in either the obtaining, verification or presentation of the contents”<sup>211</sup> of databases regardless of whether or not they can be copyrighted.<sup>212</sup> Acts prevented by the right are “extraction”<sup>213</sup> and/or re-utilisation<sup>214</sup> of the whole or of a substantial part . . . of the contents of that database.”<sup>215</sup> The right does not extend to extraction and/or re-utilisation of “insubstantial parts of its contents . . . for any purposes whatsoever,”<sup>216</sup> and any contractual provision to the contrary is null and void.<sup>217</sup>

The Directive allows Member States to adopt exceptions for certain non-commercial uses. It allows extractions “for private purposes of the contents of a non-electronic database,” “for the purposes of illustration for teaching or scientific research,” and for “the purposes of public security or an administrative or judicial procedure.”<sup>218</sup> Unlike the copyright portion of the Directive, however, the *sui generis* right does not permit Member States to adopt other exceptions.

---

28. A lawful user does not need authorization to engage in any restricted act “which is necessary for the purposes of access to the contents of the databases and normal use of the contents.” *Id.* at art. 6(1). The Directive permits Member States under certain circumstances to impose some limits on the restricted acts—reproduction for private purposes of non-electronic databases, use for purposes of illustration for teaching or scientific research, use for purposes of public security or an administrative or judicial procedure, and other exceptions to copyright traditionally authorized under national law as long as they do not “unreasonably prejudice[] the rightholder’s legitimate interests or conflict[] with normal exploitation of the database.” *Id.* at art. 6(2)-(3). This language is patterned after similar language in the Berne Convention, *supra* note 13, at art. 10, and the TRIPs Agreement, *supra* note 14, at art. 13. For an analysis of the “fair use” implications of article 6(2), see Reichman & Samuelson, *supra* note 172, at 78-9.

211. Berne Convention, *supra* note 13, at art. 7(1).

212. *Id.* at Recitals 6-7. The Directive does not define the term “substantial investment,” but does provide that “any investment” “may consist in the deployment of financial resources and/or the expending of time, effort and energy[.]” *Id.* at Recital 40.

213. *Id.* at art. 7(2)(a) (“[E]xtraction shall mean the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form.”) (internal quotations omitted).

214. *Id.* at art. 7(2)(b) (“Re-utilisation shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission.”) (internal quotations omitted).

215. *Id.* at art. 7(1).

216. *Id.* at art. 8(1).

217. *Id.* at art. 15.

218. *Id.* at art. 9(a)-(c)

Additional provisions of the Directive impose obligations on lawful users that limit the scope of these exceptions.<sup>219</sup> The *sui generis* term of protection is fifteen years.<sup>220</sup> Qualitative or quantitative “substantial change” “which would result in the database being considered to be a substantial new investment,” entitles the database to an additional fifteen-year term of protection.<sup>221</sup> The *sui generis* right is available only to database producers who are either EU nationals or habitual residents,<sup>222</sup> which includes business entities that have a business presence in the EU.<sup>223</sup> The right is only available to non-EU nationals who do not meet these conditions on the basis of reciprocity. Thus, the EU can conclude agreements to extend the right to databases made in third countries<sup>224</sup> only where the third country offers “comparable protection” to EU databases.<sup>225</sup>

#### IV. QUESTIONING THE WISDOM OF EU POLICY CONCERNING DATABASE PROTECTION

The discussion so far has explored the EU copyright law agendas and followed the formulation process of the Database Directive, shedding light on EU copyright policy in general and key changes in the Database Directive formulation in particular.<sup>226</sup>

Part IV will analyze the wisdom of EU copyright policy concerning database protection. Section IV.A discusses the actual effect of the Database Directive, which ultimately shows that many of the EU’s underlying assumptions were flawed. Section IV.B then discusses the possible invalidity, or at least vulnerability, of the Directive, namely, potential European and international challenges to the Directive that were overlooked in its development process. Together, these inquiries shed light on additional problems with EU copyright law policy.

---

219. *See, e.g., id.* at art. 8(2), 8(3).

220. *Id.* at art. 10(1).

221. *Id.* at art. 10(3).

222. *Id.* at art. 11(1).

223. Defined as a company or firm “formed in accordance with the law of a Member State and having their registered office, central administration or principal place of business within the Community[,]” or if formed elsewhere but with a registered office in the Community, its operations “must be genuinely linked on an ongoing basis with the economy of a Member State.” *Id.* at art. 11(2).

224. *Id.* at art. 11(3).

225. *Id.* at Recital 56; *see also EU Directive, supra* note 199, at 1148.

226. *See supra* Part III.

### A. Does the Database Directive Really Matter?

EU policy makers argue that the Database Directive matters because it created certainty among the EU Member States, removed obstacles to the free movement of services and goods, aspired to abolish discriminating legislation (such as the Nordic states' protection provided only to catalogues first published in Nordic states), and removed distortions to competition.<sup>227</sup> Whether the Database Directive really matters is difficult to determine because empirical data on the European experiment is almost non-existent. Additionally, available data, including consideration of the lobbying efforts involved in the process, suggests that the consideration of the subject was biased. However, the wisdom of the Directive may be challenged on several different fronts.

#### 1. *The Availability of Market-Based, Technological, and Legal Mechanisms for Protection of Databases*

The EU failed to acknowledge the availability of market-based, technological, and legal mechanisms that provide databases with sufficient protection to make additional statutory protections unnecessary. As discussed elsewhere, databases may enjoy sufficient protection already; further legal protection for the underlying content will provide, if anything, only short-lived advantages.<sup>228</sup> Review of the economic dimension of database protection demonstrates the errors of blind acceptance of the argument that the database industry would experience market failure if the law does not protect databases' raw data directly.<sup>229</sup> The empirical data cited both in support and in contradiction to this anticipated market failure are inconclusive.<sup>230</sup>

Furthermore, the discussion overlooks important alternative mechanisms that can be used to protect databases. First, databases benefit from *de facto* "protectability."<sup>231</sup> Such protection stems from the inherent features and characteristics of databases,<sup>232</sup> most notably, bundling services with raw data constitutes a barrier to market entry to competitors.<sup>233</sup> Much of the value of databases, in fact, lies in the value added to the raw data.<sup>234</sup>

---

227. Gaster Written Interview, *supra* note 87.

228. *See generally Economic Dimension, supra* note 16.

229. *Id.*

230. *See* discussion *infra* Section IV.A.2.

231. *See Economic Dimension, supra* note 16, at 126-30.

232. *Id.* at 126.

233. *Id.* at 127.

234. *Id.* For example, differentiation of factual compilations by adding to them a sophisticated search software. The LEXIS and WESTLAW databases contain the same

Therefore, a market failure argument based on the theory of public goods is not especially applicable here, and it is at best a weak defense.<sup>235</sup> Second, database producers employ many private market and technological mechanisms to prevent unauthorized extraction of data and to differentiate their databases.<sup>236</sup> Finally, various legal mechanisms protect databases.<sup>237</sup> Prominent among them are indirect forms of protection, such as computer crime and privacy laws, as well as direct forms of protection, including legal protection for technological measures, trade secrecy, trademark law, contracts, unfair competition, and tort law.<sup>238</sup> These market-based, technological, and legal mechanisms are available to database producers in the different Member States.<sup>239</sup> Therefore, the EU's failure to discuss the economic justification or market demand for legislation like the Directive is a clear oversight in its rationale.

## 2. *Evaluating the Empirical Evidence to Date: No Evidence of Market Failure*

Evaluation of the empirical evidence to date suggests that no market failure existed in the database sector. In fact, the evidence suggests that the additional protections did not result in any increase in database production in the years following the implementation of the Directive in the Member States.<sup>240</sup> In fact, existing economic evidence is inconclusive.<sup>241</sup> The discussion that follows will therefore discuss and synthesize existing evidence.

One major report prepared for the European Commission assessed the economic importance of copyright industries to the economy of its individual nations for the year 2000.<sup>242</sup> According to the report, the database,

---

public-domain materials. However, each offers different search software that provides the end-user with sophisticated search capabilities.

235. *See id.*

236. *Id.* at 131-44; *see* discussion *infra* Sections IV.B.1 and IV.B.2.

237. *See Economic Dimension*, *supra* note 16, at 147-68.

238. *Id.*

239. ESTELLE DERCLAYE, THE LEGAL PROTECTION OF DATABASES: A COMPARATIVE ANALYSIS 150-73 (discussing protection by unfair competition laws in Europe), 174-90 (discussing protection by contracts in Europe), 191-221 (discussing protection by technological measures and anti-circumvention provisions in Europe) (2008).

240. *See id.* at 109-17.

241. *Id.*

242. MEDIA GROUP, TURKU SCH. OF ECON. & BUS. ADMIN. BUS. RES. & DEV. CTR., THE CONTRIBUTION OF COPYRIGHT AND RELATED RIGHTS TO THE EUROPEAN ECONOMY (2003), *available at* [http://ec.europa.eu/internal\\_market/copyright/docs/studies/etd2002653001e34\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/studies/etd2002653001e34_en.pdf).

software, and print media industries<sup>243</sup> contribute each 1% of the EU gross domestic product, and as such make the largest contribution to the European economy among core copyright industries.<sup>244</sup> In 2000, on average the database and software industries contributed 1.35% to the GDP of European nations, with the highest relative production rate in the U.K. followed by Sweden, France, The Netherlands, Italy, and Germany.<sup>245</sup> In all copyright industries, the U.K. database industry makes the highest contribution to GDP, reflecting its relative maturity,<sup>246</sup> the international prominence of the English language, and the cultural importance of the industry.<sup>247</sup> It is hard to determine, however, whether the Database Directive served some role in the market growth attributable to databases.

An economic study by Stephen Maurer found that the number of new companies entering four database markets (France, U.K., Germany, and the United States) showed a sharp, one-time growth spurt in the three EU countries after their respective governments implemented the Database Directive in 1998,<sup>248</sup> while those in the United States experienced no such growth.<sup>249</sup> Maurer's study found, however, that the one-time growth spurt was not sustained. In fact, he shows that the number of companies entering the database market returned to the pre-Directive levels.<sup>250</sup> Thus, the Database Directive does not appear to have had any long-term effects.<sup>251</sup> The EU's own study, explored *infra*, examined only one isolated year, and so does not, and cannot, provide an accurate picture of the long-term effects of the Directive.<sup>252</sup> Moreover, also it has been suggested that the Database

---

243. Software and databases industries are industries that engage in the maintenance, creation, and sales of computer software. *Id.* at 23.

244. "Core" copyright industries are defined as industries that involve the production, and usually consumption and distribution of copyright works as well as other subject matter. *Id.* at 20.

245. *Id.* at 4.

246. See MEDIA GROUP, *supra* note 242, at 6.

247. *Id.*

248. Stephen M. Maurer et al., *Europe's Database Experiment*, 294 SCI. 789, 789 fig.1 (2001).

249. *Id.*

250. *Id.* The figure only tracks business entries, not the number of firms doing businesses. If all the companies that wanted to enter the market did in 1998, and had continued successfully, then the one time growth may have had a huge impact on the database market in the EU. Maurer's data does not address this possibility.

251. See *id.*

252. See WORKING PAPER, *supra* note 207, at 5.

Directive did not result in a significant shift of the database industry to Europe.<sup>253</sup>

A recent EU report assessing the effects of the Database Directive is inconclusive in its findings, but is nevertheless suggestive with regard to the impact database protection had on database production. The EU began reviewing the Database Directive in the summer of 2005 to assess the impact it had on the database market. In December 2005, the EU released the long awaited results of its study on the effects of the Database Directive. The Working Paper contains two parts, discussing empirical evidence as well as the operative side of such evidence, and suggesting possible avenues of action the EU could take.

a) Empirical Part of the Working Paper

The report relied upon two sources in assessing the Directive. First, an online survey of 500 database producers in Western Europe generated 101 responses, 65% of which came from private companies, most of which are based in the United Kingdom (30%), Italy, Germany, France and Belgium (46% together).<sup>254</sup> Second, information from the world's largest database directory, the Gale Directory of Databases ("GDD"), provided statistics indicating the growth of the global database industry since the 1970s.<sup>255</sup> Given the nature of the sources of the EU study, namely objective GDD data and biased survey results, it is not surprising that the results from the two sources were not similar.<sup>256</sup>

The survey indicated great support within the database industry for the *sui generis* right.<sup>257</sup> The Working Paper summarizes the survey's results:

[T]he European publishing and database industries claim that "sui generis" protection is crucial to the continued success of their activities. 75% of respondents to the Commission services' on-line survey are aware of the existence of the "sui generis" right; among these, 80% feel "protected" or "well protected" by such right. 90% believe that database protection at EU level, as opposed to national level, is

---

253. See e.g., Symposium, *Panel I: Database Protection*, 11 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 275, 300 (2001) (comments made by Mr. Alain Strowel, Law Firm NautaDutilh, Brussels, Belgium).

254. WORKING PAPER, *supra* note 207, at 5 n.5.

255. WORKING PAPER, *supra* note 207, at 5.

256. See generally *id.*

257. *Id.* at 20, 22, 25.

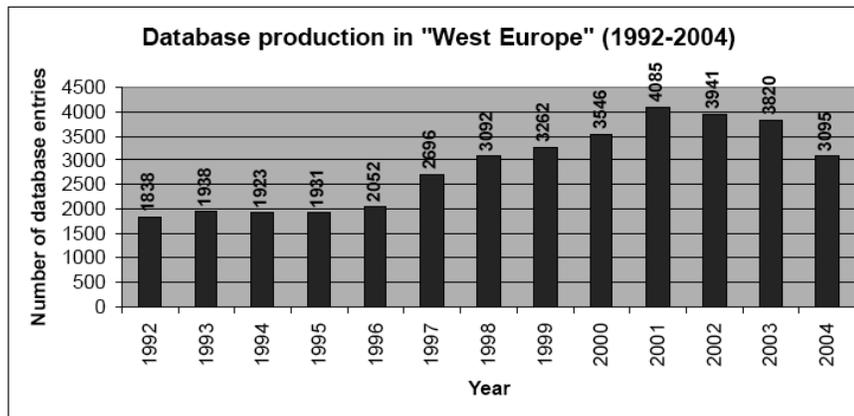
important and 65% believe that today the legal protection of databases is higher than before harmonization.<sup>258</sup>

The survey respondents also believe that the *sui generis* right brought about many positive changes. First, it facilitated the marketing of databases. Second, it created more business opportunities. Third, it reduced the costs of databases protection. Last, it also created certainty.<sup>259</sup>

Regarding the GDD data, while they were “the only empirical figures available at [that] stage to measure the evolution of the database markets[,] these figures [were] subject to considerable uncertainty.”<sup>260</sup> The figures below, reprinted from the Working Paper, display the GDD data:

Figure 1

Figure 5 - Trend of the database sector in “West Europe” (1992-2004)



Source: The Gale Directory of Databases 2005, Vol. 1, Part 2.

Figure 1 reflects the number of “entries” into the GDD from “Western Europe.”<sup>261</sup> The number of such entries since the Directive’s implementation into national laws in 1998 has been stable.<sup>262</sup> As is shown in Figure 1, in 2004 the amount of database “entries” of Western European origin was

258. *Id.* at 20.

259. *Id.* at 25.

260. *Id.* at 19.

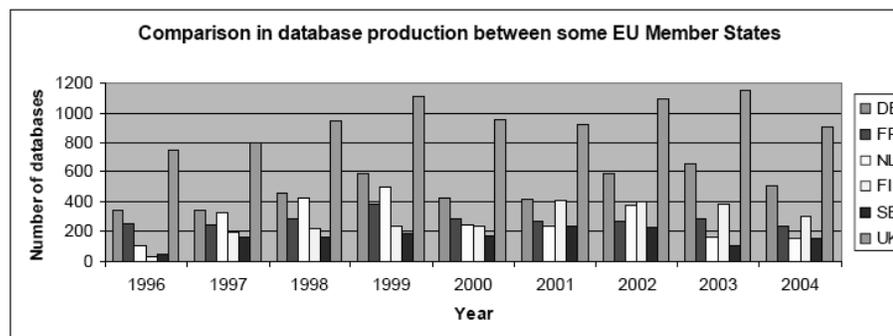
261. *Id.* at 18 n.36 (“The GDD does not define the ‘Western Europe’ market, but reports that the UK should be included in such market. Other EU countries’ markets for which the GDD reports significant figures are Germany, France, the Netherlands, Finland, and Sweden.”).

262. *Id.* at 18.

3095 compared to 3092 entries in 1998.<sup>263</sup> Concerning the decline of database “entries” from 2001, it was argued that as a result of a shift toward the provision of information online, the number of database “entries” decreased.<sup>264</sup> Additionally, it was pointed out that other technological changes occurred that might have impacted the results. Specifically, a shift in database delivery methods from “stand-alone database products, such as CD-ROM, and dedicated on-line access to specific databases, to ‘portal’-based applications” that allow “a single point of access to many databases.”<sup>265</sup> It is alleged that such trends are not reflected in the GDD.<sup>266</sup> As the Working Paper states, “these figures are subject to considerable uncertainty.”<sup>267</sup>

**Figure 2**

**Figure 6 - Comparison in database production between some EU Member States (1996-2004)**



Source: The Gale Directory of Databases (1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005 editions)

Figure 2 illustrates that in the UK, which endorsed the “sweat of the brow” doctrine pre-directive, database production has shown a net increase from 1996 through 2004, and it has remained the largest database producer in Europe.<sup>268</sup> The Working Paper suggested that other reasons might account for this long-time success such as “relative maturity of the UK database industry and the success of databases that are produced in English.”<sup>269</sup>

263. *Id.*

264. *Id.* at 19.

265. WORKING PAPER, *supra* note 207, at 19.

266. *Id.*

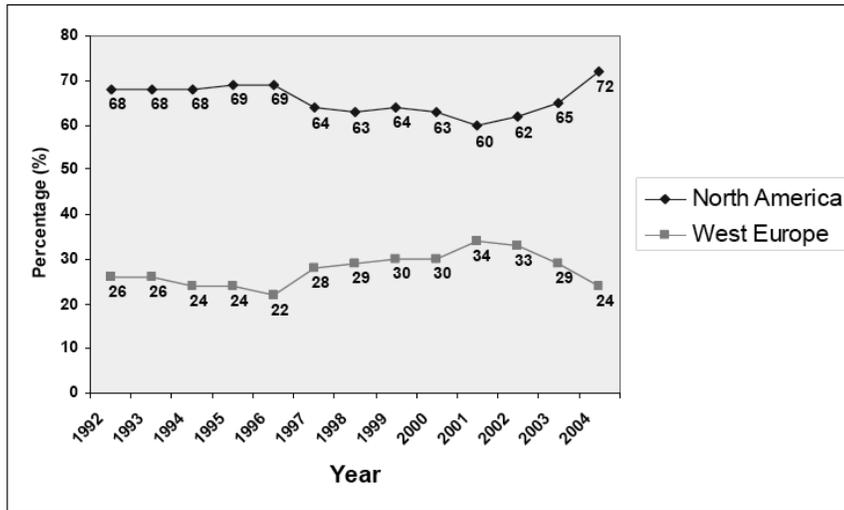
267. *Id.*

268. *Id.* at 20.

269. *Id.*; see also discussion *supra* note 246-247 and accompanying text.

Figure 3

Figure 7 - Database production in North America and West Europe (1992-2004)



Source: The Gale Directory of Databases 2005, Vol. 1, Part 2.

The data in Figure 3 compares U.S. and EU database production. “During the period of 1996-2001, Western Europe’s share in global database production increased from 22% to 34% while the ‘North American’ share decreased from 69% to 60% during the same period.”<sup>270</sup> Additionally, “[b]etween 2002 and 2004, the European share decreased from 33% to 24% while the U.S. share increased from 62% to 72%.”<sup>271</sup> Importantly, “[t]he ratio of European/US database production, which was nearly 1:2 in 1996, [became] 1:3 in 2004.”<sup>272</sup>

One could argue, however, that measuring productivity of the database sector based on the number of databases produced is problematic for many reasons. First, such data does not take into account major technological changes over the last two decades that significantly changed the way information is produced and provided. Additionally, such data does not provide any information regarding the nature of the databases produced (for example, commercial, scientific, etc.) so there is no indication whether certain databases are under or over produced, for example.

However, due to lack of other empirical data, measuring the number of databases produced was the only way to explore the evolution of databases sales since the introduction of the Database Directive. The GDD thus pro-

270. *Id.* at 22.

271. *Id.*

272. *Id.*

vides the available data.<sup>273</sup> Specifically, for the purposes of the GDD, measuring the size of the database industry is conducted by looking at changes in database “entries” into the GDD.<sup>274</sup>

When the Working Paper was published, the EU also invited stakeholder submissions as an additional source of data for reassessment of the Directive. Overall, there were 55 contributors that included eight users, 13 academic associations, and 31 database producers.<sup>275</sup> Most of the contributors did not wish to repeal the *sui generis* right, and they were evenly split as to whether to amend the Database Directive.<sup>276</sup> Not surprisingly, the submissions included letters from opponents of the *sui generis* right, such as the Consumer Project on Technology and an alliance of scientific and academic advocacy organizations.<sup>277</sup> These letters criticized the Directive for its over-breadth and also attacked the survey discussed in the Working Paper for a lack of objectivity, due to its reliance upon data provided by the database industry.<sup>278</sup> The submissions, however, also included letters from supporters, such as the U.K. Newspaper Society and the Brussels-based International Federation of Reproductive Rights Organizations, who pointed out that the Working Paper’s GDD data was inconclusive and had little probative value.<sup>279</sup>

In summary, the Working Paper and the later submissions all provide data sources that are either subjective in nature (the survey results and the

---

273. *Id.* at 18.

274. Database “entry” in the GDD represents different kinds of databases such as databases on one or more media (CD-ROM, diskette, on-line, etc.). *Id.* at 5 n.6.

275. EUROPA, Protection of Databases, [http://ec.europa.eu/internal\\_market/copyright/prot-databases/prot-databases\\_en.htm#20060427](http://ec.europa.eu/internal_market/copyright/prot-databases/prot-databases_en.htm#20060427) (last updated Jan. 9, 2008).

276. *Id.*

277. See Communication & Information Resource Centre Administrator [CIRCA], Database Consultation, [http://circa.europa.eu/Public/irc/market/market\\_consultations/library?l=/copyright\\_neighbouring/database\\_consultation&vm=detailed&sb=Title](http://circa.europa.eu/Public/irc/market/market_consultations/library?l=/copyright_neighbouring/database_consultation&vm=detailed&sb=Title) (last visited Oct. 8, 2008).

278. See MICHELLE CHILDS, CPTECH COMMENTS ON DG INTERNAL MARKET AND SERVICES WORKING PAPER (2006), available at [http://circa.europa.eu/Public/irc/market/market\\_consultations/library?l=/copyright\\_neighbouring/database\\_consultation/cptech\\_enpdf/\\_EN\\_1.0\\_&a=d](http://circa.europa.eu/Public/irc/market/market_consultations/library?l=/copyright_neighbouring/database_consultation/cptech_enpdf/_EN_1.0_&a=d); ALL European Academics (ALLEA), *Response to the Working Paper Evaluating Directive 96/9/EC on the Legal Protection of Databases*, available at [http://circa.europa.eu/Public/irc/market/market\\_consultations/library?l=/copyright\\_neighbouring/database\\_consultation/allea\\_enpdf/\\_EN\\_1.0\\_&a=d](http://circa.europa.eu/Public/irc/market/market_consultations/library?l=/copyright_neighbouring/database_consultation/allea_enpdf/_EN_1.0_&a=d) (last visited Nov. 3, 2008).

279. See Letter from Catherine Courtney, Legal Adviser, The Newspaper Society, to Tilman Lueder, Head of Unit D1 Copyright and Knowledge-based Economy, European Commission (Mar. 12, 2006), available at [http://circa.europa.eu/Public/irc/market/market\\_consultations/library?l=/copyright\\_neighbouring/database\\_consultation/newspaper\\_society/\\_EN\\_1.0\\_&a=d](http://circa.europa.eu/Public/irc/market/market_consultations/library?l=/copyright_neighbouring/database_consultation/newspaper_society/_EN_1.0_&a=d).

later stakeholders submissions) or merely inconclusive (the GDD data), and thus fail to measure and assess the impact of the Directive.

b) Policy Options of the Working Paper

The Working Paper also provides conclusions with regard to the Directive and then provides four policy options.<sup>280</sup> It concludes, based on European Court of Justice case law, that the *sui generis* right is not easy to understand and that the protection it provides is close to providing a property right in data.<sup>281</sup> Furthermore, based on the economic analysis discussed above, it concludes that “the economic impact of the *sui generis* right is unproven.”<sup>282</sup> The Working Paper then moves on to discuss four policy options: repealing the Directive, withdrawal of the *sui generis* right provisions, amending the *sui generis* provisions, and maintaining the status quo.<sup>283</sup>

In conclusion, the survey indicates great support within the database industry for the *sui generis* right whereas the empirical data shows that the *sui generis* right has had no significant economic impact on the production of databases. It is still unclear, however, what course of action the EU will take. However, given the unproven effect of the Directive, it seems unlikely that the EU will remain inactive concerning the Directive. It will probably try to reassess the impact of the Directive in the near future and decide whether to take further action.

3. *The Futility of Reciprocity*

The Database Directive’s *sui generis* right is available to database producers who are either EU nationals or habitual residents.<sup>284</sup> However, business entities with a business presence in the EU are included. Business presence is defined as a central administration, principal place of business, or a registered office in the EU, coupled with a genuine, ongoing operational link with the economy of a Member State.<sup>285</sup> Although no data regarding foreign affiliates of non-EU companies operating in the EU database market is readily available, it is possible that non-EU companies seeking to benefit from the *sui generis* right could use this loophole either because they already have some presence in the EU, have moved some of

---

280. WORKING PAPER, *supra* note 207, at 23-27.

281. *Id.* at 23-24.

282. *Id.* at 24.

283. *Id.* at 25-27.

284. Database Directive, *supra* note 78, at art. 11(1).

285. *Id.* at art. 11(2).

their operations to Europe, or found a local partner.<sup>286</sup> Even if none of these conditions are met, protection may still be available to non-EU residents via the reciprocity provision. If their own countries' provision of database protection is substantially similar to that afforded by the EU, then they may enjoy such protection within the EU.<sup>287</sup> It should be noted, however, that no "massive move" of non-EU companies transferring their operations to the EU exists.<sup>288</sup> Therefore, if the goal behind the reciprocity requirement was to provide a competitive advantage to European companies and to incentivize foreign companies to move their operations to the EU so they can benefit from the protection, this goal has not been accomplished.<sup>289</sup>

It should be noted, however, that the Database Directive negotiations show that some of the major players in the EU database industry, such as the United Kingdom, opposed the reciprocity provision and supported the provision of the *sui generis* right on a purely national basis.<sup>290</sup> This opposition is in line with the requirement of national treatment under international intellectual property treaties.<sup>291</sup> However, since granting *sui generis* rights in subject matter, which is otherwise unprotectable under the international intellectual property law treaties, then one could argue that reciprocity is a legitimate tool at a country's disposal.

#### 4. *Non-Identical Implementation of the Directive*

It is possible that the disparate implementation of the Database Directive in Member States, as well as the courts' struggles to apply the Directive's ill-defined standards, created uncertainty and unpredictability in this field, thus reducing database producers' reliance on the Directive as a meaningful source of protection.<sup>292</sup>

---

286. Maurer, *supra* note 248, at 790 ("Furthermore, the Council Directive contains a loophole: If a U.S. company wants database rights, it can get them by moving some of its operations to Europe or else by finding a local partner."); see Malla Pollack, *The Right to Know?: Delimiting Database Protection at the Juncture of the Commerce Clause, the Intellectual Property Clause and the First Amendment*, 17 CARDOZO ARTS & ENT. L.J. 47, 112 (1999) ("American' producers of databases can, moreover, obtain protection inside the European Union by supplying their databases through European affiliates or offices.").

287. Database Directive, *supra* note 78, at art. 11 (3); Recital 56.

288. Symposium, *supra* note 253, at 300.

289. *Id.*

290. Gaster Interview, *supra* note 100.

291. See, e.g., TRIPs Agreement, *supra* note 14, at art. 3.

292. See generally DAVISON, *supra* note 40, at 152-59 (providing an overview regarding the transposition of the Directive, pointing to the different levels of harmonization in different contexts).

As explained above, the Member States experienced significant difficulties throughout the formulation process with the implications of the Database Directive and its terminology. Indeed, the Database Directive includes many undefined terms that are still awaiting court interpretation.<sup>293</sup> For example, while the Directive requires “substantial investment” by the database producer, measured either “qualitatively” or “quantitatively,” there exists little guidance as to what minimal amount of investment is actually required.<sup>294</sup> Similarly, while the Directive prohibits re-utilization of a “substantial” part of a database, no definition or quantification of this term is provided.<sup>295</sup> The problems with such a standard are self-evident in the context of databases. Some numerical scale, percentage, or even some other quantitative method could have provided clearer guidance.

The European Commission’s Internal Market directorate general commissioned a study that shed light on the Member States’ legislation implementing the Database Directive.<sup>296</sup> Although all twenty-five Member States have now implemented the Database Directive, the study only addresses the implementation of the fifteen states that were EU members at the time.<sup>297</sup>

Not surprisingly, the study revealed that Member States implemented the Directive’s fundamental provisions and text with varying levels of faithfulness.<sup>298</sup> The study introduced a scale regarding Member States implementation, with “consistent” implementation reflecting the highest form of loyalty to the Directive and “extremely sketchily” reflecting the lowest form of loyalty. For example, the study showed that Italy and Spain were the only countries that “very dedicatedly” implemented the Direc-

---

293. See generally P. Bernt Hugenholtz, *Implementing the European Database Directive*, in INTELLECTUAL PROPERTY AND INFORMATION LAW, ESSAYS IN HONOUR OF HERMAN COHEN JEHOAM 183 (Jan J.C. Kabel & Gerard J.H.M. Mom eds., 1998) [hereinafter *Implementing the Directive*]; Estelle Derclaye, *What is a Database?—A Critical Analysis of the Definition of a Database in the European Database Directive and Suggestions for an International Definition*, 5 J. WORLD INTELL. PROP. 981 (2002); Estelle Derclaye, *Databases Sui Generis Right: What is a Substantial Investment? A Tentative Definition*, 36 INT’L REV. INDUS. PROP. & COMPETITION L. 2 (2005).

294. DAVISON, *supra* note 40, at 189-90.

295. *Id.*

296. See NAUTADUTILH, THE IMPLEMENTATION AND APPLICATION OF DIRECTIVE 96/9/EC ON THE LEGAL PROTECTION OF DATABASES, STUDY CONTRACT ETD/2001/B5-3001/E/72 (2002), available at [http://ec.europa.eu/internal\\_market/copyright/docs/databases/etd2001b53001e72\\_en.pdf](http://ec.europa.eu/internal_market/copyright/docs/databases/etd2001b53001e72_en.pdf); see also *Implementing the Directive*, *supra* note 293, at 183 (providing a status report on the implementation of the Directive in the Member States).

297. NAUTADUTILH, *supra* note 296, at 8, 11-12.

298. *Id.* at 371.

tive.<sup>299</sup> Belgium and Greece implemented the Directive “quite dedicatedly”.<sup>300</sup> Austria, France, Ireland, The Netherlands, Portugal, and the United Kingdom implemented it “mostly satisfactorily, though not flawlessly” while Germany and Luxembourg implemented it “more or less satisfactorily.”<sup>301</sup> Denmark and Finland implemented it “overly sketchily” and Sweden implemented it “extremely sketchily.”<sup>302</sup>

While these “grades” do not shimmer with clarity, it is clear that the Directive has not fully harmonized the legal protection for databases. A closer look at these Member States’ legislation reveals that their legislation departs significantly from the Directive in relation to fundamental provisions, undermining the Directive’s goal to achieve harmonization. States had implemented the definition of database, the *sui generis* right, the reciprocity provision and exceptions to the *sui generis* right, differently from each other and the Directive.

Different states implemented the definition of a “database” differently. Some, such as Denmark, Finland, and Sweden, failed to provide one, overlooking its binding nature.<sup>303</sup> Other Member States, such as Luxembourg, created their own definition,<sup>304</sup> whereas Germany, France, and Spain overlooked the applicability of this definition for databases protected under both copyright and the *sui generis* regimes.<sup>305</sup>

The *sui generis* right had also been implemented differently, possibly because Member States did not wish to destabilize their existing legal regimes. For instance, Denmark, Sweden, and Finland, countries in which the “catalogue rule” prevailed prior to the Database Directive,<sup>306</sup> incorporated the “substantial investment” requirement.<sup>307</sup> Nevertheless, these countries’ legislations suggested implicitly that a large number of information items is enough to entitle a database to protection even when not the result of substantial investment.<sup>308</sup>

Even the reciprocity provision of the Directive was implemented differently by Member States.<sup>309</sup> Luxembourg, for example, had not incorpo-

---

299. *Id.*

300. *Id.*

301. NAUTADUTILH, *supra* note 296, at 371.

302. *Id.*

303. *Id.*

304. *Id.* at 371, 229, 232-33.

305. *Id.*

306. *See* discussion *supra* note 46-50 and accompanying text.

307. NAUTADUTILH, *supra* note 296, at 374.

308. *See id.*

309. *See id.* at 374-75.

rated the provision.<sup>310</sup> Germany did not adopt the genuine-link requirement on persons with a registered office in its territory.<sup>311</sup> Finland had adopted only the “registered office” part, omitting the “central administration or principal place of business options” language.<sup>312</sup> These changes are fundamental as they allow greater practical leeway to foreign legal entities.

Member States also differed in their implementation of the mandatory exception to the *sui generis* right. Although the Directive does not define the term “lawful user,” it is arguably clear, based on the Explanatory Memorandum of the Proposal and the Commission report on the implementation and effects of the Software Directive.<sup>313</sup> However, Belgium, Ireland, and the U.K. had expanded the definition to others as well.<sup>314</sup> Additionally, Denmark, Finland, and Sweden had vested the lawful user with a right to extract and re-utilize the whole catalogue, rather than only insubstantial parts of it. These changes clearly depart from the Directive’s allowable exceptions.<sup>315</sup>

Last, Denmark, Finland, and Sweden had not adopted the Directive provision regarding substantial new investment that entitles the database producer to a new 15-year term of protection.<sup>316</sup>

##### 5. *Confusion in the Courts*

Uncertainty concerning the Directive’s language, scope and application has led to confusion in the European courts that limited its signific-

---

310. *Id.* at 375.

311. *Id.*

312. *Id.*

313. *Id.* at 376.

314. *Id.* at 376. Specifically, Belgium defined lawful user in Article 2(4) of the Legal Protection of Database Act 1998 as “a person who effects acts of extraction or re-utilisation authorised by the database maker or permitted by law.” Consequently, the definition is not restricted to those in a contractual relationship with the maker, and includes those who are accessing a database in order to avail themselves of one of the exceptions to the right of extraction or re-utilisation, including the right to use insubstantial part. The Irish Copyright and Related Rights Act 2000 in Article 327 permits lawful users to extract and re-utilize insubstantial parts of the contents of a database. The Act defined “lawful user” in Article 320 as “any person who, under a license to undertake the acts restricted by any database right in the database, or otherwise, has a right to use the databases.” Regulation 12 of The Copyright and Rights in Databases Regulations 1997, SI 1997, No. 3032 (Eng.), available at <http://www.uk-legislation.hmso.gov.uk/si/si1997/19973032.htm>, defined lawful user as “any person who (whether under a licence to do any of the acts restricted by any database right in the database or otherwise) has a right to use the database.”

315. NAUTADUTILH, *supra* note 296, at 376.

316. *Id.* at 377.

ance. Moreover, a closer look at some of the case law applying these instruments in the Member States as well as emerging European Court of Justice (“ECJ”) case law provides even more evidence of the uncertainty regarding database protection law in the EU.<sup>317</sup> It should be noted, however, that more guidance is provided by the courts in the EU, including the ECJ, so it is believed that with the passage of time the strength of this argument will be weakened.<sup>318</sup>

Courts have been struggling with the application of different key concepts such as the notion of “database,” the “substantial investment” requirement, the status of database “maker,” among others.<sup>319</sup> Discussion of the court’s decisions regarding the phrase “substantial investment” is illustrative.

The Database Directive does not offer guidance in interpreting the term “substantial investment.” It is unclear how much database producer needs to invest in creating a database in order to qualify for *sui generis* protection. Additionally, it is unclear which “investment” is “eligible” for recoupment. This uncertainty is especially problematic when dealing with databases generated as by-products (“spin-offs”) of other principal activities that are already incentivized. For example, telephone directories and TV program listings are offered to the public as a by-product of TV and broadcasting services.

Are the labor and cost attributable to the production of a database the only investment that should be considered with regard to the *sui generis* right or are the labor and cost spent in generally organizing these services should also be considered? In November 2004, the ECJ handed down four

---

317. P. Bernt Hugenholtz, Paper Presented at the Fordham Law School Eleventh Annual Conference on International IP Law & Policy: Program Schedules, Event Data and Telephone Subscriber Listings under the Database Directive: The ‘Spin-Off’ Doctrine in the Netherlands and Elsewhere in Europe (Apr. 14-25, 2003), available at <http://www.ivir.nl/publications/hugenholtz/spinofffordham.html> (based on article in 26 A.M.I. 161 (2002)); P. Bernt Hugenholtz, *Abuse of Database Right Sole-Source Information Banks under the EU Database Directive*, in ANTITRUST, PATENTS AND COPYRIGHT: EU AND US PERSPECTIVES 203 (François Lévêque & Howard A. Shelanski eds., 2005); Mark J. Davison & P. Bernt Hugenholtz, *Football Fixtures, Horse Races and Spin Offs: The ECJ Domesticates the Database Right*, 27 EUR. INTEL. PROP. REV. 113 (2005).

318. See C304/07, *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg*, 2008 WLR (D) 312 (ECJ).

319. See P. Bernt Hugenholtz, Presentation at the Ninth Annual Conference on International IP Law & Policy: The New Database Right: Early Case Law from Europe (Apr. 19-20, 2001), available at <http://www.ivir.nl/publications/hugenholtz/fordham2001.html> (discussing early case law in Member States, mostly from Germany, The Netherlands, and France); see also F.W. Grosheide, *Database Protection—The European Way*, 8 WASH. U. J.L. & POL’Y 39, 57-77 (2002).

decisions concerning the Database Directive that offer guidance for answering such questions.<sup>320</sup> Three of the cases were brought by Fixtures Marketing Ltd., and dealt with the use of lists of football fixtures by betting companies.<sup>321</sup> The other case, *British Horse-racing Board. Ltd. v. William Hill Organization Ltd.* (“BHB”) focused on the use by betting companies in Britain of data provided by BHB to one of its licensees (specifically, dates, times, and places of horse races as well as the names and numbers of horses taking part in each race).<sup>322</sup> The most important element of the ECJ’s decisions was the separation of the investment made to create the data from the investment needed to obtain, verify, or present it. In the decision concerning the *Oy Veikkans* case the ECJ stated in that connection that

[f]inding and collecting the data which make up a football fixture list do not require any particular effort on the part of the professional leagues. Those activities are indivisibly linked to the creation of those data, in which the leagues participate directly as those responsible for the organisation of football league fixtures. Obtaining the contents of a football fixture list thus does not require any investment independent of that required for the creation of the data contained in that list.<sup>323</sup>

Essentially, the ECJ found that the investment necessary to benefit from the *sui generis* right must be in obtaining, presenting, or verifying preexisting data.<sup>324</sup> The ECJ in effect implicitly embraced the “spin-off” doctrine.<sup>325</sup> According to this doctrine, a database producer can benefit from the *sui generis* right only when investment is attributable directly to database production.<sup>326</sup> The doctrine is premised on the incentive rationale of the *sui generis* right.

Equally important is the ECJ’s interpretation of the scope of database protection. The *sui generis* right protects producers of a database against

---

320. Case C-338/02, *Fixtures Mktg. Ltd. v. Svenska Spel AB*, 2004 E.C.R. I-10479; Case C-444/02, *Fixtures Mktg. Ltd. v. Organismos Prognostikon Agonon Podosfairou*, 2004 E.C.R. I-10549; Case C-46/02, *Fixtures Mktg. Ltd. v. Oy Veikkaus AB*, 2004 E.C.R. I-10365; Case C-203/02, *British Horseracing Bd. Ltd. v. William Hill Org. Ltd.*, 2004 E.C.R. I-10415.

321. *Svenska*, 2004 E.C.R. I-10479; *Organismos*, 2004 E.C.R. I-10549; *Oy Veikkaus*, 2004 E.C.R. I-10365.

322. *British Horseracing*, 2004 E.C.R. I-10415.

323. *Oy Veikkaus*, 2004 E.C.J. I-10365, at ¶ 44.

324. Davison & Hugenholtz, *supra* note 317, at 114.

325. *Id.*

326. *Id.*

the unauthorized extraction or re-utilization of a “substantial part” of the database.<sup>327</sup> The ECJ provides some guidance regarding how to assess what constitutes a “substantial part.” The court explained:

[S]ubstantial part, evaluated qualitatively, of the contents of a database refers to the scale of the investment in the obtaining, verification or presentation of the contents of the subject of the act of extraction and/or re-utilisation, regardless of whether that subject represents a quantitatively substantial part of the general contents of the protected database. A quantitatively negligible part of the contents of a database may in fact represent, in terms of obtaining, verification or presentation, significant human, technical or financial investment.<sup>328</sup>

The discussion of these cases illustrates how higher-level courts in Europe stepped in to further interpret and refine key concepts of the Directive. This line of cases also reflects an emerging line of cases that supports restrictive reading of the *sui generis* right to an extent that reduces many of the initial dangers associated with such legislation and significantly weakens the right. Of course, there are many open questions pertaining to the Database Directive that are awaiting interpretation. However, extensive discussion of these issues is beyond the scope of this Article.

#### 6. *Persistence of Other Barriers*

Last, as discussed above, the EU information market suffered from not only fragmentation by legal barriers but also technological and linguistic ones.<sup>329</sup> It is possible that the persistence of some or all of these barriers in the EU information market has also contributed to the ineffectiveness of the Directive and the EU’s inability to increase its share in that market

### **B. Possible Challenges to the Database Directive**

Section IV.B will explore possible challenges to the Directive and will suggest that the EU did not consider, and in fact overlooked, such challenges. The following discussion suggests that there are possibly two routes for challenging the Directive. The first route suggests bringing constitutional challenges to the Directive at both the member state and EU level. Such challenges can arguably restore the traditional balance between publishers, authors and users that the Directive disrupted and can signal to the EU institutions that their imbalanced approach is inappropriate. The

---

327. Database Directive, *supra* note 78, at art. 7.

328. C-203/02, *British Horseracing Board Ltd. v. William Hill Org. Ltd.*, 2004 E.C.R. I-10415, at ¶ 71 (Judgment of Nov. 9, 2004).

329. See discussion *supra* notes 63–66 and accompanying text.

second route discusses the possibility of challenging the Directive at the international level via the different intellectual property conventions.

1. *National and EU Constitutional Validity*

In his work on copyright and freedom of expression in Europe,<sup>330</sup> Professor Bernt Hugenholtz provides an overview of the state of European law concerning the conflict between copyright and freedom of expression. As Professor Hugenholtz describes, the European Court confirmed that speech relating to commercial interests enjoys only limited protection in Europe.<sup>331</sup> The European Court allows Member States latitude in restricting speech to serve the interests of commercial law and the law of unfair competition.<sup>332</sup> This line of cases suggests that Article 10 of the European Convention on Human Rights (“ECHR”), which provides that “[e]veryone has the right to freedom of expression” and that “[t]his right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers,”<sup>333</sup> could be invoked and would probably allow unauthorized use of copyrighted works for predominantly commercial purposes only in exceptional cases.<sup>334</sup>

Professor Hugenholtz demonstrates how the European Court does not treat all content-related speech restrictions equally and that, in cases involving political speech rather than “ordinary” expression, the European Commission and the European Court consistently granted a higher level of protection to the former than the latter, recognizing the democracy-enabling function of the freedoms Article 10 protects.<sup>335</sup>

Freedom of expression arguments are likely to succeed against copyright claims aimed at preventing political discourse, curtailing journalistic or artistic freedoms, suppressing publication of government-produced information, or impeding other forms of “public speech.”<sup>336</sup> In practice, this approach might imply that the European Court would be willing to find violations of Article 10 if national courts fail to interpret broadly or

---

330. P. Bernt Hugenholtz, *Copyright and Freedom of Expression in Europe*, in EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY, INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY 343 (Rochelle C. Dreyfuss et al. eds., 2001) [hereinafter *Freedom of Expression*].

331. *Id.* at 361.

332. *Id.*

333. European Convention for the Protection of Human Rights and Fundamental Freedoms, Nov. 4, 1950, 213 U.N.T.S. 221, art. 10 [hereinafter ECHR].

334. *See Freedom of Expression*, *supra* note 330, at 361.

335. *See Freedom of Expression*, *supra* note 330, at 361.

336. *Id.* at 362.

“stretch” existing copyright limitations to permit quotation, news reporting, artistic use, or reutilization of government information.<sup>337</sup> The Court might also be willing to find national copyright laws in direct contravention of Article 10 if they fail to provide exceptions for specific uses.<sup>338</sup> However, some factors must be taken into account while examining whether speech regulations are constitutional under Article 10.<sup>339</sup> Such regulation needs to go through the test of necessity “in a democratic society,” such as a proportionality test. The following factors must be taken into account: the degree of public interest in the speech; the substantiality of the restrictions; and the purpose of the regulation.<sup>340</sup> European case law also suggests that speech restrictions in line with European consensus will more readily be accepted than those reflecting national peculiarities.<sup>341</sup>

The Database Directive provides copyright-like property rights in intellectual creations, namely databases. While both copyrights and database protection are not explicitly listed in the ECHR as human rights, commentators and courts suggested that Article 1 of the First Additional Protocol of the convention that recognizes the property right as a human right also encompasses intangible creations and not only tangible goods.<sup>342</sup> Copyrights and database protection under the Database Directive are both property rights that are protectable human rights under the ECHR.<sup>343</sup> As such they are also considered “rights of others” under Article 10(2) of the ECHR, which can restrict the freedom of expression and the right to information of the public under Article 10.<sup>344</sup> Therefore, the analysis of the possible tension between copyright law and freedom of expression and information is also applicable with regard to the Database Directive. If the Database Directive has upset the “delicate balance” between copyright or the *sui generis* rights and the public interest, providing property rights in data where no economic need in such legislation has been demonstrated, the European Court might be convinced to step in and apply Article 10 to restore the balance. Additionally, it is also questionable whether such an all-encompassing right will be enforceable when one considers the user’s freedom of expression and information.

---

337. *Id.*

338. *Id.*

339. *Id.*

340. *Id.*

341. *Id.*

342. CHRISTOPHE GEIGER, *DROIT D’AUTEUR ET DROIT DU PUBLIC À L’INFORMATION—APPROCHE DE DROIT COMPARÉ* (2004).

343. *Id.* at 167.

344. *Id.* at 167-68.

In general, many writers have been critical not only of the Directive's two-tier copyright/*sui generis* approach. They have also been critical of the text's lack of clarity, the fact that many of its provisions were optional, and the frequency with which contentious issues were placed in the Directive's recitals.<sup>345</sup> Some argue that the initial fears of over-protection in the Directive were realized in the adopted text as a result of the imbalance of rights evident in the provisions on the *sui generis* right.<sup>346</sup> Aside from these general arguments, some also argue that the Directive enables right holders to acquire greater protection under the *sui generis* right than copyright has ever offered.<sup>347</sup>

Examination of the Database Directive provides some potential provisions for constitutional challenge under ECHR's Article 10. For example, the Directive allows for all exemptions traditionally found in the Member States' copyright laws.<sup>348</sup> However, unauthorized copying for private purposes from electronic databases is not permitted.<sup>349</sup>

Furthermore, the Directive allows only limited statutory exemptions with respect to the *sui generis* right in comparison to those allowed under the Directive's copyright chapter.<sup>350</sup> Article 9 of the Directive leaves no room for many traditional limitations, such as journalistic freedoms, quotation rights, library privileges or reuse of government information.<sup>351</sup> Apparently, the users' freedom to extract and re-utilize insubstantial parts of the database was considered sufficient.

Moreover, the makers of the Directive seem to have overlooked that extracting or re-utilizing even substantial parts of a database may constitute perfectly legitimate use. All but one exception available under the *sui generis* right—Article 8(1) that provides the right for lawful users to extract or re-utilize insubstantial parts of a database's contents—are optional and may be overridden by contract.<sup>352</sup> Again, it may well be that the narrow scope of exceptions to the *sui generis* right can also be challenged for not properly taking into account user's rights to an extent that upsets the

---

345. Georges Koumantos, *Les Bases de Données dans la Directive Communautaire [Databases in the Community Directive]*, 171 REVUE INTERNATIONALE DU DROIT D'AUTEUR (RIDA) 78 (1997).

346. For the strongest criticism of the EU and U.S. approaches to database protection, see Reichman & Samuelson, *supra* note 172, at 55.

347. *Id.* at 92.

348. *Cf.* Database Directive, *supra* note 78, at art. 9.

349. *Id.* at art. 9 (stating that extraction for private purposes of the contents of a non-electronic database are allowed).

350. *Compare id.* at art. 9 with art. 3.

351. *Id.* at art. 9.

352. *Id.* at art. 8.

delicate balance between property rights or specifically *sui generis* right and freedom of expression and information. Finally, one could argue that the *sui generis* right does not distinguish between non-copyrightable ideas and their copyrightable expression, thus potentially inhibiting the evolution of a “public domain substratum from which either research workers or second comers are progressively entitled to withdraw previously generated data without seeking licenses that may or may not be granted.”<sup>353</sup> By extension, database producers “obtain proprietary rights in data as such, a type of ownership that the copyright paradigm expressly precludes,”<sup>354</sup> exposing the Directive to an additional general basis for constitutional challenge.

Professor Hugenholtz also examined selected national constitutional laws in the Member States, showing that constitutional challenges might also be successful in some, or even many, national courts to an imbalanced copyright legislation.<sup>355</sup> Thus, challenges to the Database Directive in the national courts of the Member States may also be possible as the Directive grants a property right in data that is also in tension with freedom of expression and information.

To signal to the EU institutions that their imbalanced approach is inappropriate and unacceptable, the EU courts at all levels might use these constitutional challenges to restore the traditional balance between publishers, authors, and users. The EU’s imbalanced policies are evident, for example, in its later copyright harmonization initiative, the Copyright Directive.<sup>356</sup> This directive tried to harmonize certain aspects of copyright law in the EU, reflecting a clear departure from the previous piecemeal approach. The result was poor legislation that protects mainly the interests of “producers, broadcasters, and institutional users” and does not seriously take into account authors’ rights.<sup>357</sup>

## 2. *Invalidity under International IP Conventions*

The Directive’s reciprocity provision may be challenged at the international level under the different intellectual property conventions. As explained above, the EU adopted the reciprocity provision despite the trend toward national treatment because it is a useful negotiation chip in bilater-

---

353. Reichman & Samuelson, *supra* note 172, at 88.

354. *Id.* at 89.

355. *See Freedom of Expression, supra* note 330, at 354-58.

356. Council Directive 2001/29/EC, On the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society, 2001 O.J. (L 167/10).

357. P. Bernt Hugenholtz, *Why the Copyright Directive is Unimportant and Possibly Invalid*, 22 EUR. INTELL. PROP. REV. 499 (2000).

al negotiations with trading partners.<sup>358</sup> However, it may also be a potent tool for pressuring other countries to adopt similar legislation.<sup>359</sup>

A closer look at the different international instruments reveals that such a challenge probably has merit. The Berne Convention, the Copyright Treaty and the TRIPS Agreement all include national treatment obligation, and in the case of the TRIPS Agreement, the similar obligation of Most Favored Nation. The national treatment obligation requires a nation to extend the same copyright protection (as well as other kinds of IP rights) provided to its own nationals to the citizens of other nations that are parties to the relevant international agreement.<sup>360</sup> While TRIPS specifies certain minimum requirements of protection for intellectual property, the Agreement also provides that Members may provide more extensive protection to intellectual property than is required by the Agreement. Therefore, if a nation chooses to provide more extensive protection to intellectual property than is required under TRIPS, then it must provide that additional protection to nationals of other nations. Thus, it seems that a nation cannot provide that additional protection to such nationals only on a reciprocal basis. On its face, then, the Database Directive's reciprocity requirement is in direct violation of TRIPS, and arguably invalid.

The obligation of national treatment arguably extends only to the protection required by TRIPS so it can be argued that the Database Directive is not covered by TRIPS.<sup>361</sup> However, the Agreement seems to be clear in that obligations of national treatment and most favored nation treatment extend to all intellectual property. Another argument is that the Database Directive does not constitute intellectual property, and as such is not subject matter under TRIPS, and therefore is not subject to the national treatment and most favored nation provisions. However, examination of the Directive suggests that the *sui generis* right may well fall under the definition of "intellectual property" under Article 1(2) of the TRIPS Agreement. This Article specifically includes "copyrighted and related rights" as specific kind of intellectual property. On its face the *sui generis* right is similar to copyright as the scope and nature of the rights conferred on the pro-

---

358. See *supra* Section IV.A.3 (discussing the reciprocity requirement).

359. *Id.*

360. See, e.g., TRIPs Agreement, *supra* note 14, at art. 3(1); Berne Convention, *supra* note 13, at art. 5; see also TRIPs Agreement, *supra* note 14, at art. 4 (providing that "[w]ith regard to intellectual property, any advantage, favour, privilege or immunity granted by a Member to the nationals of any other country shall be accorded immediately and unconditionally to the nationals of all other Members").

361. See TRIPs Agreement, *supra* note 14, at art. 1(3) ("Members shall accord the treatment provided in this Agreement to the nationals of other Members.").

ducer are property rights in data for a possibly eternal term, even stronger than those granted under intellectual property laws that are usually limited in term and scope.

Finally, some have argued that the *sui generis* right might be characterized as a right to prevent a new species of “unfair competition,” within the meaning of Article 10.<sup>362</sup> Thus, the *sui generis* right might be subject to the national treatment obligations the Berne Convention and the TRIPs Agreement imposed.<sup>363</sup> However, the current version of the *sui generis* right appears to have acquired most of the traits of an exclusive property right—an economic right that “has nothing in common with unfair competition remedies because it does not sanction behaviour *a posteriori* and because it provides for a term of protection.”<sup>364</sup>

## V. CONCLUSION

This Article evaluates EU copyright policy through the lens of the debate over database protection. The database protection debate has challenged the intellectual property regimes not only of the EU but also of many other countries, raising questions as to the need in either extending copyright protection to unoriginal databases or, alternatively, crafting a form of *sui generis* protection.

An examination generally of the origins of EU copyright policy and specifically of the Directive’s formulation process, reveals the EU’s underlying policy goals as well as its copyright agenda. In particular, the analysis shows that while facilitating integration and trade between Member States were definitely major policy goals, other motivations also played a role in the design of the Directive, most notably the EU’s strong desire to provide its database producers with a competitive advantage in the global information market.

The Article also questions the EU’s copyright agenda, suggesting that the EU has failed to consider many issues in its formulation of the Directive. These include basic questions such as whether such a Directive is economically necessary and whether any such regime should not only ac-

---

362. *But see* William R. Cornish, *1996 European Community Directive on Database Protection*, 21 COLUM.-VLA J.L. & ARTS 1, 13 (1996) (arguing that the *sui generis* right in its nascent stage in the first proposal was more of an unfair-competition-type remedy than a true intellectual property right and as such can be subject to the national treatment requirement).

363. TRIPs Agreement, *supra* note 14, at art. 3(1); Berne Convention, *supra* note 13, at art. 5.

364. Jens L. Gaster, *The EU Council of Ministers’ Common Position Concerning the Legal Protection of Databases: A First Comment*, 6 ENT. L.R. 258 (1995).

count for publishers' interests but also offer a balanced approach to other important interests, such as those of end-users and authors. This Article also suggests that the EU has failed to consider the Directive's other potential vulnerabilities.

The EU's "experiments" with copyright law and *sui generis* protection raise many questions, some of which still await exploration. While this Article tried to explore EU copyright law policy critically through the lens of the Database Directive, additional research is required. The EU is a dominant player in the global communication and information market, and its policies can serve as good case studies for exploring the role of intellectual property more broadly in incentivizing research and development in different contexts. Therefore, future work on EU copyright policy will advance our understanding regarding intellectual property regimes in general and EU copyright policy in particular.

# PHYSIOLOGICAL STEPS DOCTRINE

By Andrew W. Torrance<sup>†</sup>

## TABLE OF CONTENTS

I. INTRODUCTION .....	1472
II. “NATURAL CONVERSION” PATENTS .....	1474
III. <i>IN VIVO</i> CONVERSION PATENTS.....	1478
A. INFRINGEMENT UNDER THE DOCTRINE OF EQUIVALENTS .....	1478
1. Ortho Pharmaceutical v. Smith.....	1478
B. PROBLEMS OF EVIDENCE .....	1479
1. Zenith Labs. v. Bristol-Myers Squibb.....	1479
C. INHERENCY .....	1482
1. Marion Merrell Dow, Inc. v. Geneva Pharms, Inc.....	1482
2. Schering Corp. v. Geneva Pharms., Inc. ....	1484
3. In re Omeprazole.....	1486
4. <i>Discussion</i> .....	1487
D. CLAIM CONSTRUCTION.....	1489
1. Marion Merrell Dow Inc v. Baker Norton Pharms., Inc. ....	1489
2. Hoechst-Roussel Pharms, Inc. v. Lehman .....	1491
3. Mylan Pharms, Inc. v. Thompson.....	1493
4. In re Buspirone.....	1495
5. Novartis Pharmaceuticals Corp. v. Eon Labs Manufacturing Inc.....	1496
6. <i>Discussion</i> .....	1498
IV. THE PHYSIOLOGICAL STEPS DOCTRINE.....	1500
A. THE EUROPEAN EQUIVALENT.....	1501
B. WITHIN THE UNITED STATES.....	1502

---

© 2008 Andrew W. Torrance

<sup>†</sup> Associate Professor of Law, University of Kansas School of Law. B.Sc. (Queen's), A.M., Ph.D., J.D. (Harvard). I would like to thank Profs. Mark Lemley and Dan Burk for valuable comments on an early version of this Article, Jonathan Singer for bringing the issue of *in vivo* conversion patent litigation to my attention, and for encouraging me to pursue research into the issue, and Profs. Cynthia Ho, Thomas Cotter, Michael Meurer, Jonathan Golden, and Adam Candeub for their suggestions at the 2008 MSU Intellectual Property and Communications Scholars Conference. In addition, I would like to thank Joseph Wright, Jonathan Grossman, and Zachary Lerner for their excellent research assistance with this Article.

V. CONCLUSION .....1505

I call on Congress to pass legislation that bans unethical practices such as the . . . patenting . . . of human life.

—George W. Bush<sup>1</sup>

I. INTRODUCTION

Most inventions exist in the world outside the human body. For these, the usual legal rules of the patent statute and interpreting case law apply. Inventions ranging from bicycles, barometric chambers, and bobby socks to tumbling mats, toasters, and toenail clippers all fall squarely within patentable subject matter. Moreover, since *Diamond v. Chakrabarty*, the potentially patentable subject matter has extended almost to the limits of the human imagination.<sup>2</sup> However, the bounds are not limitless. There exist specifically recognized exceptions to patentable subject matter; for instance, “[t]he laws of nature, physical phenomena, and abstract ideas have been held not patentable.”<sup>3</sup> However, there may also be exceptions yet unrecognized. This Article argues that the metabolic products of *in vivo* conversion by and within the human body fall within one of these unrecognized categories of unpatentable subject matter.

Patent claims whose elements involve the physiological functions inside a human being fit uneasily into patent law. The patent laws of the United States already include explicit limits on patent claims as they relate to humans and human bodies. Inventions carried out using human thought have long been subject to limitations such as the Mental Steps Doctrine.<sup>4</sup> Further restrictions include limits on inventions related to surgery and medicine, limiting liability for patent infringement by medical personnel and medical facilities, and the unpatentability of human-nonhuman genetic hybrids, or chimaeras.<sup>5</sup> Specifically, the “Weldon Amendment” rider, which has been renewed since 2004, states that “[n]one of the funds ap-

---

1. State of the Union 2008 (January 28, 2008), in 44 WEEKLY COMP. PRES. DOC. 117 at 120.

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

3. *Id.* at 309.

4. See, e.g., *Gottschalk v. Benson*, 409 U.S. 63, 67 (1972) (“Phenomena of nature, though just discovered, mental processes, and abstract intellectual concepts are not patentable, as they are the basic tools of scientific and technological work.”).

5. 35 U.S.C. § 287(c) (2000); Patent Application Is Disallowed as “Embracing” Human Being, 58 PAT. TRADEMARK & COPYRIGHT J. (BNA) 203 (June 17, 1999).

propriated or otherwise made available under this Act may be used to issue patents on claims directed to or encompassing a human organism.”<sup>6</sup> Similarly, it has been the stated policy of the USPTO since 1987 that “[a] claim directed to or including within its scope a human being will not be considered to be patentable subject matter.”<sup>7</sup> President William J. Clinton has publicly opposed patents on human clones,<sup>8</sup> and President George W. Bush has urged Congress “to pass legislation that bans unethical practices, such as the . . . patenting . . . of human life.”<sup>9</sup> Notwithstanding USPTO policy and the unambiguous exhortations of Presidents Clinton and Bush, there is no clear statutory prohibition against human beings, or properties thereof, being patentable subject matter. Though the law has yet to speak clearly on the matter, this Article suggests that inventions that include a human being as part of their structure or operation generally sit towards the unpatentable end of the spectrum.

Such classes of potentially unpatentable inventions involving significant human participation include products of *in vivo* conversion. *In vivo* conversion is a process, often metabolic in nature, wherein one substance, usually a chemical compound, is altered significantly by physiological pathways in the body into one or more different substances.<sup>10</sup> For example, when a patient ingests a therapeutic drug, that drug is often converted by the natural physiology of the digestive system into one or more chemically different metabolites. The end products of *in vivo* conversion sometimes possess therapeutic efficacy.

Many patent applications have claimed such therapeutic metabolites, either as compositions *per se* or as parts of methods of treatment. Although the USPTO has granted patent claims to such products generated by *in vivo* conversion of ingested drugs, and courts have noted the eligibility of such products as patentable subject matter, never has a United States court of final appeal upheld such a patent claim as valid, enforceable, and infringed.<sup>11</sup>

This Article reviews the judicial decisions that considered infringement by *in vivo* conversion, including the forerunners of *in vivo* conver-

---

6. Consolidated Appropriations Act, 2004, Pub. L. No. 108-199, § 634, 118 Stat. 3, 101 (2004).

7. Animals—Patentability, 1077 OFF. GAZ. PAT. OFFICE 24 (Apr. 21, 1987).

8. Eliot Marshall, *Clinton Urges Outlawing Human Cloning*, 276 SCI. 1640 (1997).

9. State of the Union 2008, (January 28, 2008), in 44 WEEKLY COMP. PRES. DOC. 117.

10. If the process of *in vivo* conversion transforms a precursor chemical into a second chemical that has therapeutic efficacy, the precursor is sometimes called a “prodrug,” and the resulting therapeutic chemical a “drug”.

11. See discussion *infra* Part III.

sion cases: those involving “natural conversion.” It then considers several possible explanations for these decisions’ unanimity in never finding *in vivo* conversion claims valid, enforceable, and infringed. Finally, it proposes a novel doctrinal framework that can explain the consistent outcome of *in vivo* conversion cases: the Physiological Steps Doctrine.

## II. “NATURAL CONVERSION” PATENTS

Long before they began to grapple with patentability and infringement issues surrounding *in vivo* conversion, courts considered an analogous phenomenon: “natural conversion.” Natural conversion involves the transformation of one or more starting substances into one or more product substances without direct human intervention to cause the transformation. The fermentation of alcohol provides a simple example. The carbohydrates in an initial mixture of grape juice and yeast, if maintained at an appropriate temperature and exposed to a minimal concentration of oxygen, will be chemically transformed into an ethanol mixture otherwise known as wine. Though wines are usually produced with careful human oversight, fermentation also commonly occurs in nature without human involvement.

When one or more noninfringing chemical reactants are converted into one or more different chemical products, and at least one of these products falls within a patent claim, the usual result is patent infringement. For example, in *Chemical Cleaning v. Dow Chemical*, Chemical Cleaning, Inc. (CCI) was found to have infringed a patent by selling a cleaning product, Sequestrol 60, that, upon decomposition, produced a chemical, thiourea, claimed in Dow Chemical’s patent.<sup>12</sup> The Fifth Circuit Court of Appeals affirmed the district court’s finding that CCI was liable for patent infringement under the doctrine of equivalents.<sup>13</sup> Similarly, in *Broadview Chemical v. Loctite*, Broadview Chemical Corporation tried to avoid infringement of a patent claim whose elements included the chemical quinone by substituting the chemical hydroquinone into its product instead.<sup>14</sup> The court, however, found that Broadview Chemical Corporation did infringe that claim because upon use of the product the hydroquinone un-

---

12. *Chem. Cleaning, Inc. v. Dow Chem. Co.*, 379 F.2d 294, 296-97 (5th Cir. 1967) (“Sequestrol 60 is prepared by compounding thiourea and formaldehyde under alkaline conditions. The process is reversible, and the trial court found that under the boiler treating conditions employed by CCI, Sequestrol 60 disassociates to produce about 90% by weight of thiourea in the free or uncombined form, and some formaldehyde.”).

13. *Id.* at 297.

14. *Broadview Chem. Corp. v. Loctite Corp.*, 159 U.S.P.Q. 80, 85 (D. Conn. 1968).

derwent a chemical reaction and released quinone.<sup>15</sup> Thus, in nonbiological chemical contexts, there is strict liability for patent infringement both in situations where the allegedly infringing chemical is initially present in a product and where the allegedly infringing chemical arises in the product by means of chemical reaction. However, involvement by an organism in the production of the infringing chemical appears to alter this result.

Prior to the first *in vivo* conversion case, *Feed Service Corp. v. Kent Feeds, Inc.*, involving a dispute between two agricultural feed companies, considered whether a claim to a specific mixture of ingredients could be infringed by a mixture initially lacking a claim element, but then subsequently generating that element by a natural process (in this case, fermentation) occurring within the initial mixture.<sup>16</sup> Although the district court found infringement, the Court of Appeals disagreed, ultimately laying the groundwork for later *in vivo* conversion cases.<sup>17</sup>

Feed Service Corporation, an agricultural feed company that made feed for livestock, owned U.S. Patent No. 2,808,332 ('332 patent).<sup>18</sup> The '332 patent had twenty-one claims for feeds of specified formulations and methods of using such feeds to improve rates of growth in cattle that according to the specifications was achieved by adding synthetic urea and ethyl alcohol (also known as ethanol) as ingredients.<sup>19</sup> The '332 patent disclosed that ethanol had been previously mentioned in association with feeding animals, but never to achieve improved growth in ruminants; in fact, ethanol was usually identified as an ingredient to avoid.<sup>20</sup> Feed Service marketed its feed under the trade name "Morea," and its product was a commercial success.<sup>21</sup>

Kent Feeds, Inc. developed and sold competing feeds, trade named "Bovino" and, later, "Bovino-Lac."<sup>22</sup> Feed Service sued Kent Feeds for infringing claims of the '332 patent, alleging that "Bovino-Lac" contained each and every claimed ingredient.<sup>23</sup> Kent Feeds disputed infringement on the ground that their product included fermented molasses and not etha-

---

15. *Id.*

16. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 528 F.2d 756, 763 (7th Cir.), *cert. denied*, 429 U.S. 870 (1976).

17. *Id.* at 763-64.

18. *Id.* at 757.

19. U.S. Patent No. 2,808,332 (filed Feb. 17, 1955); *Feed Serv. Corp.*, 528 F.2d at 758.

20. '332 Patent, col. 2 l.19-30.

21. *Feed Serv. Corp.*, 528 F.2d at 759.

22. *Id.* at 763.

23. *Id.* at 759.

nol.<sup>24</sup> In response, Feed Service argued that there was, in fact, infringement because “the fermentation process of the blackstrap molasses converts virtually all of the sugar in the molasses to alcohol. . . .”<sup>25</sup>

The district court agreed with Feed Services, finding that Kent Feeds’s product contained ethanol in the amounts specified by the claims,<sup>26</sup> and thus infringed all of the claims of the ’332 patent.<sup>27</sup> Just as in the three nonbiological chemical cases noted above, the district court appears to have been untroubled by the provenance of the patented chemical, drawing no distinction between the direct addition of ethanol to feed and ethanol produced by fermentation of molasses within the feed mixture itself.

Kent Feeds appealed, alleging that the claims of the ’332 patent claims were invalid, unenforceable, and not infringed by Feed Services’s feed product.<sup>28</sup> Unlike the district court, the Seventh Circuit Court of Appeals, relying on the prosecution history of the ’332 patent, considered the provenance of ethanol in the feed to be a decisive issue, and refused to construe the patent claims as covering any and all feed supplements containing ethanol, regardless of how the ethanol was obtained:

We read the claims to teach the use of alcohol in its liquid form and not the use of alcohol derived in a fermentation process of molasses or from other fermented sources. . . . We cannot say that its monopoly extends to the mere presence of alcohol resulting from a molasses fermentation process.<sup>29</sup>

Furthermore, the Court of Appeals distinguished between claimed ingredients added to feed by conscious human agency and claimed ingredients arising *in situ*:

Defendants do not *add* alcohol to their feed supplements and plaintiff does not charge them with that. The charge of infringement is based on the use by defendants of fermented molasses which provides the alcohol in question as a *natural occurring event*. . . . The fact that the defendants’ Bovino product may reach the same result as plaintiff’s Morea is not conclusive of the determination of infringement.<sup>30</sup>

---

24. *Id.* at 763.

25. *Id.*

26. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 185 U.S.P.Q. 745, 750 (N.D. Ill. 1975).

27. *Id.* at 753.

28. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 528 F.2d 756, 757 (7th Cir. 1976).

29. *Id.* at 763.

30. *Id.* at 764 (emphasis added).

Accordingly, the Court of Appeals reversed the district court's finding of infringement, though it did affirm the lower court's finding of validity.<sup>31</sup>

In his dissent, Judge Stevens disputed the majority's interpretation of "addition," noting "that the incorporation of fermented molasses is a method of adding ethanol."<sup>32</sup> In addition, some commentators have suggested that the Court of Appeals wrongly imported the claim element of "incorporating" from process claims into product claims.<sup>33</sup> However, as claims 11 and 16, both product claims to feed mixtures, are the only claims the Court of Appeals reproduces in its opinion,<sup>34</sup> the court appears to have been aware that its interpretation included product claims.

Although the Court of Appeals did not offer a clear rationale for its decision, it appears to have considered the "natural" origin of an ingredient in the feed mixture to be significant—perhaps even decisive.<sup>35</sup> By doing so, *Feed Service v. Kent Feeds* set the stage for later *in vivo* conversion cases by suggesting that claims to products generated by natural *biological* processes may be less patentable than claims to identical products made synthetically or artificially.

---

31. *Id.* Significantly, the court interpreted not just the process claims, but also the product claims covering the feed itself, to involve ethanol that had been added or incorporated as ethanol *per se*. *See id.* (noting that defendant's actions were "a far cry from plaintiff's overzealous charge of blatant infringement, literal piracy and outright duplication.").

32. *Id.* at 764.

33. *See, e.g.*, 3 MARTIN J. ADELMAN ET AL., PATENT LAW PERSPECTIVES § 3.2 (2d ed. 2006)

In *Feed Service Corp. v. Kent Feeds, Inc.*, the Seventh Circuit appears to have committed serious error in reversing the lower court's holding of infringement of composition of matter claims to a cattle feed supplement "comprising urea and ethanol." . . . One can find no warrant whatsoever, in fact or in law, for such a construction of these patent claims. The invention of these claims was a feed supplement comprising ethyl alcohol and urea—not how to make such a supplement. This is confirmed by the court's observation that "[T]he novelty of the patent in suit was the conception of the idea of incorporating ethyl alcohol and a synthetic nitrogen source in feed supplements. This led to the formulation of feed supplements containing ethyl alcohol and urea as the source of synthetic nitrogen." This opinion indicates either a failure on the part of the court adequately to comprehend patent law or an inability on the part of the court adequately to express its reasons for deciding as it did.

*Id.*

34. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 528 F.2d 756, 758 (7th Cir. 1976).

35. *Id.* at 764.

### III. *IN VIVO* CONVERSION PATENTS

Since the 1976 *Feed Service* opinion, there have been ten recorded judicial decisions of cases involving allegations of infringement of products generated by *in vivo* conversion of known drugs. These decisions were decided on different grounds: the doctrine of equivalents, evidentiary insufficiency, inherent anticipation, and claim construction. Though these cases display a variety of facts and rationales, their results agree in one significant respect: no appeals court ever found a claim on a product of *in vivo* conversion to be valid and infringed.<sup>36</sup>

#### A. Infringement Under the Doctrine of Equivalents

##### 1. *Ortho Pharmaceutical v. Smith*

Ortho Pharmaceutical Corporation (Ortho) sued for declaratory judgment against American Home Products, the exclusive licensee of U.S. Patent No. 3,959,322 ('322 patent), Dr. Herchel Smith, and Wyeth-Ayerst Laboratories (collectively American Home Products, or AHP), asking the district court for a declaration of invalidity of claims of the '322 patent.<sup>37</sup>

Ortho marketed norgestimate, a steroid oral contraceptive that it had developed by modifying norgestrel, a chemical covered by claims 5 and 19 of the '322 patent.<sup>38</sup> When ingested, norgestimate is transformed by *in vivo* conversion into norgestrel, one of two "products [that] are primarily responsible for the biological activity of norgestimate."<sup>39</sup> Thus, the district court found infringement under the doctrine of equivalents.<sup>40</sup> Although it did appeal on other grounds, including invalidity of the '322 patent, Ortho did not appeal the finding of patent infringement itself.<sup>41</sup>

Despite the fact that this case involved one compound, norgestimate, that is transformed by *in vivo* conversion into a different infringing compound, norgestrel, it is conceptually distinct from the other *in vivo* infringement cases because the district court based its decision on the doctrine of equivalents. Under the court's application of the doctrine of equivalents, both norgestimate and its *in vivo* product, norgestrel, were found

---

36. At first glance, *Ortho Pharm. Corp. v. Smith*, 959 F.2d 936, 939 (Fed. Cir. 1992), would seem to be an exception. However, here the court found infringement under the doctrine of equivalents.

37. *Ortho Pharm. Corp. v. Smith*, 959 F.2d 936, 937-38 (Fed. Cir. 1992).

38. *Ortho Pharm. Corp. v. Smith*, 18 U.S.P.Q.2d 1977, 31 (E.D. Pa. 1990).

39. *Ortho Pharm. Corp.*, 959 F.2d at 939.

40. *Id.*

41. *Id.* at 940.

independently to infringe claims 5 and 19 of the '332 patent. In other words, the court did not find that infringement was triggered by *in vivo* conversion. By contrast, all of the cases considered below involve the issue of whether infringement can be triggered by *in vivo* conversion.

## B. Problems of Evidence

### 1. *Zenith Labs. v. Bristol-Myers Squibb*

Bristol-Myers Squibb (BMS) developed an antibiotic, cefadroxil, and a novel, crystalline form of cefadroxil, named the "Bouzard monohydrate," that possessed significant advantages over other forms of cefadroxil in terms of its manufacture and therapeutic administration.<sup>42</sup> BMS owned U.S. Patent No. 4,504,657 ('657 patent), which contained a single claim to the Bouzard monohydrate specifying a chemical formula and thirty-seven specific x-ray diffraction properties.<sup>43</sup>

Zenith Laboratories planned to market a form of cefadroxil, Cefadroxil DC, that differed structurally from the Bouzard monohydrate, but converted into the Bouzard monohydrate through *in vivo* conversion.<sup>44</sup> After BMS alleged that Cefadroxil DC infringed the '657 patent claim, Zenith sued in district court for a declaratory judgment against BMS, alleging, among other things, that Cefadroxil DC did not infringe the claim of the '657 patent.<sup>45</sup> After agreeing that Cefadroxil DC did not literally infringe, BMS adjusted its theory of infringement to (1) infringement under the doctrine of equivalents and (2) infringement because "Zenith's product converted into the patented compound in the patient's stomach, and thus the sale of cefadroxil DC would induce infringement of the '657 patent under 35 U.S.C. § 271(b)(1988)."<sup>46</sup>

Initially, the court granted Zenith's motion for summary judgment of noninfringement.<sup>47</sup> But later, the court vacated its first decision, citing new evidence that "had demonstrated a genuine dispute on the *in vivo* conversion issue."<sup>48</sup> After a bench trial, the court found no infringement under the doctrine of equivalents, but "[s]ince an act of literal infringement thus occurs in the patient's stomach as a result of ingestion of cefadroxil DC, the court concluded that Zenith's sale of cefadroxil DC

---

42. *Zenith Labs., Inc. v. Bristol-Myers Squibb Co.*, 19 F.3d 1418, 1419 (Fed. Cir. 1994).

43. *Id.* at 1419-20.

44. *Id.* at 1420.

45. *Id.*

46. *Id.*

47. *Id.*

48. *Id.* at 1421.

would induce infringement of the 657 patent.”<sup>49</sup> Zenith subsequently appealed.<sup>50</sup>

The Court of Appeals for the Federal Circuit had never before considered a case involving infringement triggered by *in vivo* conversion. The Federal Circuit rejected Zenith’s proposed interpretation limiting the claim of the ’657 patent to a pre-ingested form of the Bouzard monohydrate.<sup>51</sup> Further, in Footnote 4, the Federal Circuit implied that a product of *in vivo* conversion could, in theory, trigger infringement: “The trial court apparently reached the same conclusion: ‘use of converted Bouzard monohydrate by a patient who ingests cefadroxil DC is an infringing use.’ (But see note 6 regarding the significance of the term ‘use.’)”<sup>52</sup> However, the parenthetical statement that ends Footnote 4 vitiates the possibility of finding infringement via *in vivo* conversion by employing the contrasting word “But” to signal that Footnote 6 (a recitation of arguments offered by Zenith, but not relied upon by the Federal Circuit) offers an interpretation of the word “use” that would require non-incidental *in vivo* conversion:

Zenith offered three other grounds on which the judgment of the trial court could be reversed: an incidental conversion to Bouzard crystals does not “use” the claimed compound; the reverse doctrine of equivalents forecloses literal infringement by conversion; and equitable estoppel.<sup>53</sup>

Because the Federal Circuit disposed of the case on evidentiary grounds, the court noted that “we need not address these other grounds for reversal.”<sup>54</sup> Nevertheless, the Federal Circuit’s decision to temper Footnote 4 by pointing out, but not disposing of, the interpretation of “use” in Footnote 6 suggests that *in vivo* conversion producing Bouzard crystals might not constitute infringement.

Indeed, the Federal Circuit reversed the district court’s finding of infringement under the doctrine of equivalents on evidentiary grounds.<sup>55</sup> First,

[the] district court, instead of requiring the comparison of the accused compound following conversion to be made with the lines

---

49. *Id.*

50. *Id.*

51. *Id.* at 1422.

52. *Id.* at 1422 n.4.

53. *Id.* at 1424 n.6.

54. *Id.*

55. *Id.* at 1426.

specified in the claim, allowed Bristol to make the comparison with the diffraction pattern exhibited by a sample (the reference pattern) of a material considered by Bristol to be the patented compound.<sup>56</sup>

In addition, the Federal Circuit considered the district court's infringement analysis insufficiently thorough:

15 of the lines recited in the claim (representing about 40% of the total) were not considered by the court in its comparison. Although the term "essentially" recited in the claim permits some leeway in the exactness of the comparison with the specified 37 lines of the claim, it does not permit ignoring a substantial number of lines altogether.<sup>57</sup>

Thus, "there was a failure of proof as to whether any crystals, assumed to form in the stomach from ingested cefadroxil DC, literally infringe the '657 claim."<sup>58</sup>

Given the Federal Circuit's finding of noninfringement by the product of *in vivo* conversion, as well as the ambiguity latent in Footnotes 4 and 6, it might seem odd that *dicta* in *Zenith v. Bristol-Myers Squibb* would be cited for the proposition that products of *in vivo* conversion can indeed trigger infringement. Despite this, *Zenith* is perhaps the most influential case to consider the issue of whether a product of *in vivo* conversion can trigger infringement. Every subsequent case addressing the issue of infringement by *in vivo* conversion, with the exception of *In re Buspirone* and *In re Omeprazole*, has cited *Zenith* for the proposition that such infringement can occur. Ironically, though this case is invoked as the poster child of infringement via *in vivo* conversion, the court in *Zenith* did not find infringement due to lack of evidence.<sup>59</sup> Consequently, its oft-cited statements of support of infringement via *in vivo* conversion are *dicta*. Nevertheless, a close look at these subsequent *in vivo* conversion cases dispels some of this incongruity: while cases do cite *Zenith v. Bristol-Myers Squibb* on the issue of infringement via *in vivo* conversion, no judicial opinion has found it solely controlling on the issue of infringement.

In *Zenith*, the Federal Circuit never squarely considered the principle of law for which it is usually cited. Instead, the court found that BMS had presented insufficient evidence that *Zenith*'s generic cefadroxil would

---

56. *Id.* at 1423.

57. *Id.* at 1424.

58. *Id.*

59. *Id.* at 1423-24.

meet each and every element of the claim of BMS's patent.<sup>60</sup> Of particular significance to the court was that fact that BMS had presented evidence of only thirty lines of x-ray diffraction relative intensities, of which the district court had compared only twenty-two lines, whereas the claim itself recited thirty-seven lines.<sup>61</sup>

Obtaining evidence of an *in vivo* conversion product is difficult. The challenges include obtaining a specific biological sample from a specific location at a specific time within a living human body. In fact, as the Federal Circuit explained, "[S]cientific fact appears to be that there is no known way to actually sample the contents of patients' stomachs at the precise moment and conduct the x-ray diffraction analyses required to ascertain if all 37 lines described in the patent are present."<sup>62</sup> As a result, the samples used as evidence in *Zenith* were created *in vitro*, were not biological in origin and did not come from a human who had ingested cefadroxil DC. Obviously, any process of gathering evidence that depends on so many contingencies, not to mention practical difficulties, is bound to yield a low rate of success. In addition, there are issues of informed consent and privacy that may prevent even an attempt at obtaining a sample. It is hard to imagine a court successfully ordering a patient who has ingested a drug to submit to such an invasive procedure in the civil context of a patent trial. Consequently, a lack of evidence is likely to remain a significant hurdle to proving infringement by *in vivo* conversion. Nevertheless, *Zenith* is the only *in vivo* conversion case thus far whose outcome regarding infringement can be attributed to evidentiary grounds.

### C. Inherency

#### 1. *Marion Merrell Dow, Inc. v. Geneva Pharms, Inc.*

Marion Merrell Dow, Inc. (MMD) developed terfenadine, an antihistamine drug with the advantageous property of not causing drowsiness, and marketed it under the trade name of Seldane®.<sup>63</sup> MMD owned U.S. Patent No. 3,878,217 ('217 patent), claiming administration of terfenadine and other piperidine derivatives.<sup>64</sup> MMD's subsequent research on ter-

---

60. Eitan Alexander Ogen, *Assembling a Theory of Infringement: Third Party Liability Based on In Vivo Production of Patented Pharmaceuticals*, 17 CARDOZO L. REV. 117, 139 (1995) ("The *Zenith* facts present an array of legal issues. . . . These issues were left unresolved because the CAFC decided the case on an evidentiary basis.").

61. *Zenith Labs.*, 19 F.3d at 1423-24.

62. *Id.* at 1422-23.

63. *Marion Merrell Dow, Inc. v. Geneva Pharms., Inc.*, 877 F. Supp. 531, 533 (D. Colo. 1994).

64. *Id.*

fenadine yielded terfenadine acid metabolite (TAM), a product produced by *in vivo* conversion after ingestion of terfenadine, and obtained U.S. Patent No. 4,254,129 ('129 patent) to cover both TAM itself and methods of administering a therapeutically effective amount of TAM.<sup>65</sup>

Geneva Pharmaceuticals, Inc. applied to the FDA for regulatory approval to market a generic version of terfenadine once the '217 patent expired, and stated, as part of the regulatory certification process, that its generic product would not infringe claims of the '129 patent.<sup>66</sup> In response, MMD sued Geneva, and Geneva requested summary judgment on the grounds that the asserted claims of the '129 patent were invalid as inherently anticipated.<sup>67</sup>

MMD alleged infringement based on a theory of *in vivo* conversion: "because the product to be marketed by Geneva converts after being ingested by a patient into a compound whose use, *inter alia*, is claimed in the 129 [sic] patent."<sup>68</sup> The district court pointed out "that infringement may result from the *in vivo* conversion of one product or compound into another," citing *Zenith* for support.<sup>69</sup>

Geneva argued that claims of the '129 patent were anticipated by the disclosure of the prior '217 patent and by a scientific article which had been issued and published more than a year prior to the priority date of the '129 patent.<sup>70</sup> Geneva acknowledged that both of these pieces of prior art disclosed preparation and administration of terfenadine, not TAM, but that, after ingestion, terfenadine was necessarily transformed via *in vivo* conversion into metabolic products, including TAM.<sup>71</sup> In other words, claims to the use of TAM in the '129 patent were allegedly inherently anticipated by the teachings of the prior art.<sup>72</sup>

The district court denied Geneva's motion for summary judgment because it was "unclear . . . whether, scientifically, all the elements regarding terfenadine and its administration, as claimed in the 217 [sic] patent, are identical to the elements regarding TAM and its administration as claimed in the 129 [sic] patent."<sup>73</sup>

---

65. *Id.*

66. *Id.* at 534.

67. *Id.* at 534-37.

68. *Id.* at 535 (quoting Pl.'s Resp., Introduction).

69. *Id.*

70. *Id.* at 536.

71. *Id.*

72. *Id.*

73. *Id.* at 537.

2. *Schering Corp. v. Geneva Pharms., Inc.*

Schering Corporation owned two patents with claims covering antihistamines: U.S. Patent No. 4,282,233 ('233 patent) claimed loratadine, an active ingredient of the brand-name antihistamine CLARITIN® marketed by Schering; and U.S. Patent No. 4,659,716 ('716 patent) claimed descarboethoxyloratadine (DCL), a metabolite resulting from *in vivo* conversion of loratadine.<sup>74</sup> Upon expiration of the '233 patent, Geneva Pharmaceuticals, Inc. and other generic drug manufacturers (*Geneva et al.*) sought to bring generic drugs containing loratadine to market.<sup>75</sup> As part of the process of applying for regulatory approval from the FDA, *Geneva et al.* certified that claims of the '716 patent were invalid.<sup>76</sup> In response, Schering filed a lawsuit against *Geneva et al.* alleging that these generic drugs containing loratadine infringed claims of the '716 patent that ostensibly covered DCL, but not loratadine.

The district court construed claims 1 and 3 of the '716 patent broadly, concluding that they covered all forms of DCL, including both synthetic DCL and DCL produced by *in vivo* conversion of loratadine.<sup>77</sup> Both Schering and *Geneva et al.* agreed to this interpretation.<sup>78</sup> Then, the court used the claim construction to find "that the '233 patent did not expressly disclose DCL."<sup>79</sup> However, because "DCL was necessarily formed as a metabolite by carrying out the process disclosed in the '233 patent," and the '233 patent had expired more than a year prior to earliest priority date of the '716 patent, the court granted summary judgment in favor of *Geneva et al.*, finding that the '233 patent "anticipated claims 1 and 3 of the '716 patent under 35 U.S.C. § 102(b)" by inherent anticipation.<sup>80</sup>

Schering appealed this grant of summary judgment.<sup>81</sup> The Federal Circuit recognized that this issue "may be a case of first impression, because the prior art supplies no express description of any part of the claimed subject matter."<sup>82</sup> In fact, the panel noted that

[i]n these prior cases, however, inherency was only necessary to supply a single missing limitation that was not expressly dis-

---

74. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1374-75 (Fed. Cir. 2003).

75. *Id.* at 1376.

76. *Id.*

77. *Id.*

78. *Id.*

79. *Id.*

80. *Id.*

81. *Id.*

82. *Id.* at 1378.

closed in the prior art. This case, as explained before, asks this court to find anticipation when the entire structure of the claimed subject matter is inherent in the prior art.<sup>83</sup>

Nevertheless, the panel affirmed the district court's decision regarding inherent anticipation, rejecting Schering's contention that DCL is formed accidentally: "The record shows that DCL necessarily and inevitably forms from loratadine under normal conditions. DCL is a necessary consequence of administering loratadine to patients."<sup>84</sup> Based on these findings, the panel concluded that human ingestion of loratadine would infringe claims 1 and 3 of the '716 patent because the loratadine would be transformed by *in vivo* conversion into the DCL metabolite.<sup>85</sup> Consequently, these same claims must be invalid in light of the '233 patent because "[a]n identical metabolite must then anticipate if earlier in time than the claimed compound."<sup>86</sup>

In *dicta*, the panel supported the proposition that a metabolite produced within the body by *in vivo* conversion can indeed trigger infringement of a claim covering that metabolite itself or its use: "This court has recognized that a person may infringe a claim to a metabolite if the person ingests a compound that metabolizes to form the metabolite."<sup>87</sup> The Federal Circuit later stressed that their finding of inherent anticipation in this case would not "preclude patent protection for metabolites of known drugs."<sup>88</sup> In fact, "[w]ith proper claiming, patent protection is available for metabolites of known drugs."<sup>89</sup>

The Federal Circuit, however, considered such "proper claiming" to be restricted to purified metabolites not found in nature in purified form, citing *In re Kratz*<sup>90</sup> and *In re Bergstrom*<sup>91</sup> as illustrations.<sup>92</sup> And such *in vivo* conversion metabolites, including those recited in claims 1 and 3 of the '716 patent, "may not receive protection via compound claims . . . [be-

---

83. *Id.* at 1379.

84. *Id.* at 1378.

85. *Id.* at 1380.

86. *Id.*

87. *Id.* (citing *Hoechst-Roussel Pharms., Inc. v. Lehman*, 109 F.3d 756, 759 (Fed. Cir. 1997) and *Zenith Lab., Inc. v. Bristol-Myers Squibb Co.*, 19 F.3d 1418, 1421-22 (Fed. Cir. 1994)).

88. *Id.* at 1381.

89. *Id.* (internal citations omitted).

90. *In re Kratz*, 592 F.2d 1169 (C.C.P.A. 1979) (claims to substantially pure compounds may be patentable).

91. *In re Bergstrom*, 427 F.2d 1394 (C.C.P.A. 1970) (claims to pure substances may be patentable).

92. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1381 (Fed. Cir. 2003).

cause such] bare compound claims include within their scope the recited compounds as chemical species in any surroundings, including within the human body as metabolites of a drug.”<sup>93</sup> The Federal Circuit stated a general rule as follows: “these broad compound claims are inherently anticipated by a prior art disclosure of a drug that metabolizes into the claimed compound.”<sup>94</sup> A patent applicant wishing to claim a metabolite would have to settle for claims reciting a pure and isolated metabolite, a pharmaceutical composition containing not only the metabolite but other ingredients as well, or a method of administering the metabolite or pharmaceutical composition thereof.<sup>95</sup>

### 3. *In re Omeprazole*

Astra Aktiebolag and related companies (Astra) marketed a gastric acid inhibiting drug brand-named PRILOSEC®, whose active ingredient was omeprazole.<sup>96</sup> Astra had listed two of its patents in the FDA Orange Book: U.S. Patent No. 4,255,431 ('431 patent), which includes claims to omeprazole as a compound and to the oral administration of omeprazole for gastric acid inhibition and U.S. Patent No. 4,636,499 ('499 patent), which includes compound claims to a class of metabolites of omeprazole, called sulphenamides, as well as method claims on administration of sulphenamides to treat gastroinflammatory diseases.<sup>97</sup> Because the '431 patent was close to expiration, several generic drug companies, Genpharm, Inc., Cheminor Drugs Ltd., Reddy-Cheminor, Inc., and Schein Pharmaceutical, Inc. (collectively Genpharm *et al.*) applied for an Abbreviated New Drug Application (ANDA) to market generic versions of omeprazole.<sup>98</sup> Astra sued Genpharm *et al.* for infringement of the '499 patent, which was not close to expiration, based on 35 U.S.C. § 271(e)(2)(A) because Genpharm *et al.*'s ANDAs included paragraph IV certifications specifically challenging the validity of the '499 patent.

Astra argued that “oral administration of omeprazole . . . [would] infringe the '499 patent because when a patient takes the Genpharm . . . products, sulphenamides will form in the patient's body,” citing *Zenith* in support of their position.<sup>99</sup> Genpharm *et al.* disputed Astra's interpretation

---

93. *Id.*

94. *Id.*

95. *Id.*

96. *In re Omeprazole Patent Litig.*, No. MDL 1291, 2001 WL 585534, at \*1 (S.D.N.Y. May 31, 2001).

97. *Id.* at 1-2.

98. *Id.*

99. *Id.* at 3.

of *Zenith*, and cited *Marion Merrell Dow, Inc. v. Baker Norton*,<sup>100</sup> for the proposition that *Zenith* did not articulate a *per se* rule that claims to compounds covered both those made synthetically and those produced by *in vivo* conversion.<sup>101</sup>

The district court decided this issue in favor of Genpharm *et al.* stating that: “It cannot be that a claim to a ‘compound’ covers the compound whether it is made synthetically or produced *in vivo*, regardless of whether such a construction is supported by the evidence intrinsic to the patent.”<sup>102</sup> It thus construed the claims as only covering synthetic sulphenamides, not metabolite sulphenamides resulting from the *in vivo* conversion of omeprazole.<sup>103</sup> The court then granted Genpharm *et al.* summary judgment of invalidity of the ’499 patent’s claims based on inherent anticipation by prior art teaching administration of omeprazole to inhibit gastric acid.<sup>104</sup>

#### 4. Discussion

A patent claim is invalid for anticipation under 35 U.S.C. § 102 if each and every element of that claim is disclosed by a single prior art reference.<sup>105</sup> Even if every element of a patent claim is not explicitly disclosed in a single prior art reference, a patent claim may still be anticipated if those claim elements not explicitly disclosed are disclosed inherently by the prior art reference.<sup>106</sup> Inherency includes both inherent anticipation and inherent obviousness.

Inherent anticipation has played a significant role in findings of noninfringement in three of the *in vivo* conversion cases discussed above.<sup>107</sup>

---

100. *Marion Merrell Dow Inc. v. Baker Norton Pharms., Inc.*, 948 F. Supp. 1050 (S.D. Fla. 1996).

101. *In re Omeprazole*, 2001 WL 585534, at \*3.

102. *Id.* at 4.

103. *Id.* at 7.

104. *Id.* at 12-13.

105. *See, e.g., Lewmar Marine, Inc. v. Bariant, Inc.*, 827 F.2d 744, 747 (Fed. Cir. 1987).

106. *See, e.g., Atlas Powder Co. v. Ireco, Inc.*, 190 F.3d 1342, 1347 (Fed. Cir. 1999) (“[A] prior art reference may anticipate when the claim limitation or limitations not expressly found in that reference are nonetheless inherent in it.”).

107. In a fourth case, *Mylan Pharms., Inc. v. Thompson*, 139 F. Supp. 2d 1 (D.D.C. 2001), inherent anticipation is mentioned in passing in discussion of an analogy. *See infra* note 153.

In a case of first impression,<sup>108</sup> the Federal Circuit in *Schering v. Geneva* ruled on whether there can be inherent anticipation “when the entire structure of the claimed subject matter is inherent in the prior art.”<sup>109</sup> *Schering* was the first reported Federal Circuit case that “considered invalidating a patent claim on the basis that the entire anticipatory disclosure was inherently disclosed in a prior-art reference.”<sup>110</sup> The court found that each and every element of claims 1 and 3 of the ’716 patent, covering a metabolite, DCL, produced by *in vivo* conversion of the antihistamine loratadine, were inherently disclosed by the prior art ’233 patent claiming loratadine.<sup>111</sup> This case provides strong support for the proposition that a metabolite necessarily produced by *in vivo* conversion after ingestion of known precursor drug is inherently anticipated by a prior art disclosure of that drug.

In *Marion Merrell v. Geneva*, Geneva moved for summary judgment, arguing that patent claims to the metabolite, TAM, were invalid due to inherent anticipation by a previous patent claiming therapeutic administration of terfenadine.<sup>112</sup> Geneva did not dispute that terfenadine was converted *in vivo* into TAM, or that such *in vivo* conversion would trigger infringement of MMD’s ’129 patent if that patent was valid.<sup>113</sup> In fact, Ge-

---

108. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1378 (Fed. Cir. 2003). This decision has attracted a large amount of commentary. See, e.g., Randy P. Boyer, *Schering Corporation v. Geneva Pharmaceuticals, Inc.: Requiem for the Recognition Requirement in the Law of Inherent Anticipation*, 14 FED. CIR. B.J. 677 (2005); Anne Brown & Mark Polyakov, *The Accidental and Inherent Anticipation Doctrines: Where Do We Stand and Where Are We Going?*, 4 J. MARSHALL. REV. INTELL. PROP. L. 63 (2004). Some commentary has suggested that the effect of this decision on the drug industry, at least, will be minimal. See, e.g., Cynthia Chen, *Schering Corp. v. Geneva Pharmaceuticals, Inc.: Clarification of the Inherent Anticipation Doctrine and Its Implications*, 20 BERKELEY TECH. L.J. 95, 121 (2005). Some have expressed support for the outcome of this case. See, e.g., Dan L. Burk & Mark A. Lemley, *Inherency*, 47 WM. & MARY L. REV. 371, 381 (2005) (“Schering and the rejection of *Elan* seem to have set the Federal Circuit firmly on the right course, recognizing that knowledge is not required for inherency.”). Others maintain that inherent anticipation should be interpreted as having a knowledge requirement. See, e.g., Peter D. Smith, Note, *Anticipating Too Much: Why the Court Should Avoid Expanding the Doctrine of Inherent Anticipation*, 61 N.Y.U. ANN. SURV. AM. L. 823, 853-54 (2006).

109. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1379 (Fed. Cir. 2003).

110. Robert A. Matthews, Jr., & Louis M. Troilo, *Schering Corp. v. Geneva Pharmaceuticals, Inc.: Just How Far Can Inherent Anticipation Extend?*, 20 SANTA CLARA COMPUTER & HIGH TECH L.J. 779, 781 (2004).

111. *Schering Corp.*, 339 F.3d at 1381.

112. *Marion Merrell, Inc. v. Geneva Pharms., Inc.*, 877 F. Supp. 531, 533, 535 (D. Colo. 1994).

113. *Id.* at 536.

neva argued that such conversion into TAM, in conjunction with the '217 patent and the Huther article, inherently anticipated the asserted '129 patent claims.<sup>114</sup> But the district court decided that the scientific issues and evidence underpinning the case were too uncertain to warrant summary judgment.<sup>115</sup>

The court in *In re Omeprazole* decided that any claim language in the patents that could be construed to cover metabolites generated *in vivo* would be invalid. Such a construction would be anticipated by the prior art patent containing claims covering the original product itself.<sup>116</sup> The court then construed the claims narrowly, avoiding their inherent anticipation, and consequently granted summary judgment of noninfringement by *in vivo* metabolites of omeprazole.<sup>117</sup>

In a sizable minority of *in vivo* conversion cases, courts have cited inherency as a ground for finding noninfringement. In fact, one, *Schering*, significantly expanded the scope of the inherency doctrine. Yet, as with evidentiary problems, inherency does not satisfactorily explain the outcomes in even a bare majority of *in vivo* conversion cases, let alone all of them. Another, more universal, rationale is required to explain the striking unanimity of results.

#### D. Claim Construction

##### 1. *Marion Merrell Dow Inc v. Baker Norton Pharms., Inc.*

Baker Norton Pharmaceuticals manufactured a generic version of terfenadine, a drug described and previously protected by claims in MMD's now expired '217 patent.<sup>118</sup> In response to an ANDA filed by Baker Norton to cover its generic terfenadine, MMD filed a lawsuit alleging that selling or manufacturing terfenadine would infringe the unexpired '129 patent owned by MMD and moved for summary judgment.<sup>119</sup> Because the '129 patent did not claim terfenadine, MMD's theory of infringement depended on the physiological transformation of terfenadine into TAM

---

114. *Id.*

115. *Id.* at 537.

116. *In re Omeprazole Patent Litig.*, No. MDL 1291, 2001 WL 585534, at \*12 (S.D.N.Y. May 31, 2001).

117. *Id.* at \*12-13.

118. *Marion Merrell Dow Inc. v. Baker Norton Pharms., Inc.*, 948 F. Supp. 1050, 1051 (S.D. Fla. 1996).

119. *Id.* at 1051-52.

within the body of a human who ingested generic terfenadine: infringement triggered by *in vivo* conversion.<sup>120</sup>

During claim construction, the court focused its analysis on the meaning of the claim element “compound.”<sup>121</sup> MMD argued for an expansive interpretation by which “compound” “refers to the compound TAM regardless of whether it is created by the liver’s metabolism of terfenadine (inter vivo conversion) or by synthetic means.”<sup>122</sup> Under this construction, TAM produced in the patient’s body by *in vivo* conversion would fall within claim 1 of the ’129 patent, and thus, infringement would lie. By contrast, Baker Norton urged the court to adopt a much narrower interpretation of “compound” that included “only synthetically produced TAM.”<sup>123</sup> Drawing on evidence from the organization of the claims themselves, discussion of TAM in the specification, and the prosecution history of the ’217 patent, the court sided with Baker Norton’s interpretation, construing the word “compound” to mean only synthetic TAM, and not TAM produced by *in vivo* conversion.<sup>124</sup>

Given the narrow interpretation of “compound,” the court found no literal infringement by TAM *naturally* produced in a patient’s body by *in vivo* conversion.<sup>125</sup> Furthermore, the court declined to find infringement under the doctrine of equivalents.<sup>126</sup> However, in its analysis under the doctrine of equivalents, the court provided only an opaque rationale for its decision not to find infringement, ostensibly relying on the discretion al-

---

120. *Id.* at 1053 (“MMD argues that Baker Norton's proposed practice of the expired '217 Patent covering terfenadine and its administration will literally infringe the '129 Patent because when a patient takes the Baker Norton product, his or her liver will necessarily produce TAM.”).

121. *Id.* at 1054.

122. *Id.*

123. *Id.*

124. *Id.* at 1053-56. Especially devastating to MMD’s chosen interpretation of “compound” was testimony from its former head of clinical pharmacology, Murray Weiner, M.D.:

[i]n my wildest dreams I wouldn't think of [contemplating that the claims of the '129 Patent application could cover the swallowing of terfenadine and the subsequent conversion to TAM] because I was aware that terfenadine has been swallowed for many, many years and that its action was known . . . There was nothing I could see invented of utility. . . . And for that reason *I didn't conceive that the well-known product terfenadine could come under a patent for something into which it is converted in the body.*

*Id.* at 1056.

125. *Id.* at 1056-57.

126. *Id.* at 1057.

lowed it by the equitable nature of that doctrine.<sup>127</sup> Consequently, the court granted Baker Norton's motion for summary judgment of noninfringement of claims of the '129 patent by generic terfenadine.<sup>128</sup>

Since the court construed the word "compound" as covering only synthetic TAM, it was unnecessary for the court to consider TAM produced naturally in the body. Therefore, nowhere did the court directly comment on the issue of whether or not a product of *in vivo* conversion could trigger infringement.<sup>129</sup>

## 2. *Hoechst-Roussel Pharms, Inc. v. Lehman*

After the FDA granted Warner-Lambert Company approval in 1993 to market its drug, COGNEX®, a treatment for Alzheimer's disease containing tacrine hydrochloride, Hoechst-Roussel Pharmaceuticals, Inc. (Hoechst), filed a lawsuit alleging that COGNEX® infringed their U.S. Patent No. 4,631,286 ('286 patent).<sup>130</sup> While the '286 patent included claims to 1-hydroxy-tacrine as a composition and methods of administering the compound to treat memory loss in a patient, it did not claim tacrine hydrochloride, the active ingredient in COGNEX®.<sup>131</sup> Litigation concluded with a consent judgment by the district court "in which Warner-Lambert admitted that tacrine hydrochloride infringe[d] certain claims of the '286 patent."<sup>132</sup>

Based upon the regulatory review period for FDA market approval of COGNEX®, Hoechst applied to the USPTO for a patent term extension for its '286 patent.<sup>133</sup> Under 35 U.S.C. § 156(a), a patent owner may request that "[the] term of a patent *which claims* a product, a method of using a product, or a method of manufacturing a product shall be ex-

---

127. *Id.*

128. *Id.*

129. The court, however, did address *Zenith*:

The Court declines to find a *per se* rule in *Zenith* which requires that all claims which describe a compound be construed as covering both the metabolically produced and synthetic produced forms of the compound, without regard to the language of the claims, the specification of the patent or the prosecution history. Instead, the Court will conduct the construction pursuant to *Markman*, and as will be demonstrated *infra*, the reference points for construction in this case are much more replete with evidence that a different interpretation of the term "compound" is appropriate here than was the case in *Zenith*.

*Id.* at 1054 n.4.

130. *Hoechst-Roussel Pharms., Inc. v. Lehman*, 109 F.3d 756, 757 (Fed. Cir. 1997).

131. *Id.*

132. *Id.*

133. *Id.*

tended . . . from the original expiration date of the patent if . . . the product has been subject to a regulatory review period before its commercial marketing or use.”<sup>134</sup> Hoechst contended that “a patent ‘claims’ an FDA-approved product, within the meaning of that term as employed in the statute, if the FDA-approved product would infringe a claim of that patent.”<sup>135</sup> Furthermore, “[b]ecause use of tacrine hydrochloride allegedly infringe[d] its claim to a method of using 1-hydroxy-tacrine, Hoechst contende[d] that the ’286 patent ‘claims’ a method of using tacrine hydrochloride.”<sup>136</sup> The USPTO denied this application for two reasons. First, Hoechst was an improper applicant and, second, the ’286 patent “does not claim tacrine hydrochloride, as required by the statute.”<sup>137</sup> The district court agreed with the USPTO’s findings and granted the USPTO’s motion for summary judgment.<sup>138</sup> Hoechst appealed to the Court of Appeals for the Federal Circuit.<sup>139</sup>

The Federal Circuit affirmed the district court’s judgment “on the basis that Hoechst’s patent does not claim either the drug product [tacrine hydrochloride] which received regulatory approval or its use.”<sup>140</sup>

However, in *dicta* the Federal Circuit discussed how infringement via *in vivo* conversion might occur:

Admittedly, Hoechst may be entitled to exclude others from administering tacrine hydrochloride to patients. But this right to exclude would not arise from the fact that Hoechst has claimed tacrine hydrochloride; nor would it arise from the fact that COGNEX® contains the product claimed by Hoechst, 1-hydroxy-tacrine. Instead, the right to exclude may arise from the fact that when administered, tacrine hydrochloride metabolizes into another product, 1-hydroxy-tacrine, which Hoechst has claimed.<sup>141</sup>

---

134. *Id.* at 758.

135. *Id.*

136. *Id.*

137. *Id.* at 757-58.

138. *Id.* at 758.

139. *Id.*

140. *Id.* at 757. For qualified support of the decision, see Matthew Hirsch, Hoechst-Roussel Pharmaceuticals, Inc. v. Lehman, 13 BERKELEY TECH. L.J. 163, 173 (1998) (“Unfortunately, the court in Hoechst came to the correct decision, but for the wrong reasons. Those wrong reasons are now precedent and undermine the Hatch/Waxman Act’s public policy goals of rewarding pharmaceutical innovators.”).

141. Hoechst-Roussel Pharms., 109 F.3d at 759.

The Federal Circuit cited *Zenith* for the proposition that “infringement may occur if the administered product is converted *in vivo* into the claimed product.”<sup>142</sup> However, it failed to point out that the basis for this statement from *Zenith* did not represent the rule of the case, but, instead, was *dicta*, because the Federal Circuit made no actual finding of infringement in *Zenith*. Though disagreeing on some issues, the concurrence by Judge Newman compounded the majority’s misinterpretation, agreeing that *Zenith* stands for the proposition that “*in vivo* conversion into the drug named in the claims is direct infringement.”<sup>143</sup>

### 3. *Mylan Pharms, Inc. v. Thompson*

In 1998 Mylan Pharmaceuticals, Inc. submitted an ANDA with the FDA, hoping to market a generic version of BuSpar®, BMS’ brand name for a drug containing the active ingredient buspirone hydrochloride (buspirone).<sup>144</sup> BMS owned U.S. Patent No. 4,182,763 (’763 patent), which claimed methods of using buspirone to treat patients with generalized anxiety disorder.<sup>145</sup> The ’763 patent expired on November 22, 2000, and Mylan planned to place its own generic version of buspirone on the market that same day.<sup>146</sup> However, the day before the ’763 patent expired, the USPTO issued BMS U.S. Patent No. 6,150,365 (’365 patent), which includes a claim to a method of treating anxiety in a patient by administering 6-hydroxy-buspirone, a metabolite of buspirone produced by *in vivo* conversion.<sup>147</sup>

In light of the issuance of the ’365 patent, and of a declaration from BMS asserting that the ’365 patent claimed the use of buspirone, the FDA declined to grant Mylan final approval for its ANDA covering generic buspirone.<sup>148</sup> Mylan and one other generic drug company filed lawsuits on November 30, 2000, seeking a preliminary injunction to have the ’365

---

142. *Id.* (citing *Zenith Labs., Inc. v. Bristol-Myers Squibb Co.*, 19 F.3d 1418, 1422 (Fed. Cir. 1994)).

143. *Id.* at 764 (citing *Zenith Labs., Inc. v. Bristol-Myers Squibb Co.*, 19 F.3d 1418, 1422 (Fed. Cir. 1994)).

144. *Mylan Pharms., Inc. v. Thompson*, 139 F. Supp. 2d 1 (D.D.C. 2001) (Claim construction by district court not addressed on the merits), *rev’d*, *Mylan Pharms., Inc. v. Thompson*, 268 F.3d 1323 (Fed. Cir. 2001).

145. *Id.* at 7-9.

146. *Id.* at 9.

147. *Id.*

148. *Id.* at 9-10.

patent delisted from the FDA's "Orange Book,"<sup>149</sup> freeing the path to approval of Mylan's ANDA.<sup>150</sup>

A threshold consideration for the court was whether the claim of the '365 patent covered use of buspirone in addition to use of its metabolite, 6-hydroxy-buspirone. If use of buspirone fell outside the claim, Bristol-Myers' declaration to the FDA would be inaccurate. The court interpreted the meaning of "claim a method of using [a drug]" in 21 U.S.C. § 355(c)(2) as equivalent to the meaning of "claims . . . a method using [a drug]" in 35 U.S.C. § 156(a).<sup>151</sup> Then, it concluded that, just as for the '286 patent in *Hoechst-Roussel*, "so too is the '365 patent limited to the use of the [6-hydroxy-buspirone] metabolite . . . , and therefore the '365 patent cannot claim the administration of buspirone."<sup>152</sup> This foreclosed the argument that administration of buspirone could trigger infringement by *in vivo* conversion into the 6-hydroxy-buspirone metabolite.<sup>153</sup> As a result, the court acceded to Mylan's request for a preliminary injunction, which required that (1) Bristol-Myers request the FDA to delist the '365

---

149. *Id.* at 4 ("Upon approval of the NDA [New Drug Application], the FDA publishes any claimed patents for the approved drug in 'Approved Drug Products with Therapeutic Equivalence Evaluations,' also known as the 'Orange Book.'" (citing 21 U.S.C. § 355(j)(7)(A)(iii) (2000))).

150. *Id.* at 9-10.

151. *Id.* at 19-21.

152. *Id.* at 21.

153. *Id.* at 24. To demonstrate that Bristol-Myers' '365 patent disclaims claim coverage of oral administration, Mylan offered the following analogy to illustrate *in vivo* conversion:

Let's assume that a Bristol scientist had found . . . that a particular chemical compound in an apple was metabolized in the human body into a compound we will call "Apple A" and that when you administer Apple A it improve[s] health. . . . They file a patent application and get a patent on the systemic administration of Apple A. . . . They make tablets with Apple A. They sell those tablets. They want to stop other people from making tablets with Apple-A in them. That is fine. That is a complicated case involving issues of inherency. This is not a complicated case because what they have done here is they have tried to use this patent to stop people from selling and eating apples by arguing that when you eat an apple, it is metabolized in the human body into the equivalent of the Bristol metabolite, the equivalent of Apple A.

*Id.* at 23 n.16. By making an analogy to eating apples, the court emphasized the implications of the natural character of the health benefits flowing from the apple: what human physiology does to the apple once ingested to produce those health benefits constitutes unpatentable subject matter. Given the result at which the court arrived—construing the claim of the '365 patent to exclude metabolites produced naturally by *in vivo* conversion—one can infer that the court approved of the reasoning in the analogy.

patent from the Orange Book and (2) the FDA grant immediate approval of Mylan's ANDA.<sup>154</sup>

#### 4. *In re Buspirone*

On the heels of *Mylan v. Thompson*, a growing number of patent lawsuits involving Bristol-Myers's '365 patent prompted the Judicial Panel on Multidistrict Litigation to consolidate four separate lawsuits into a single case in the District Court for the Southern District of New York.<sup>155</sup> There, litigation involving the '365 patent continued,<sup>156</sup> with the district court rejecting infringement by Mylan's generic buspirone of the '365 patent.<sup>157</sup> The court based its finding on three different lines of analysis. First, the court construed the claim language "systemic administration to the mammal of an effective but non-toxic anxiolytic dose of the 6-hydroxy-metabolite" to mean "the administration of an externally-measured quantity of the metabolite into the body, and not to the administration of a dose of buspirone into the body, which, in turn, produces variable and changing levels (not doses) of the metabolite in the bloodstream."<sup>158</sup> Next, after reviewing the prosecution history of the '365 patent, the court decisively rejected Bristol-Myers' assertion that the patent claim covered buspirone:

In sum, every time Bristol-Myers explicitly claimed a use of "buspirone" or a "prodrug" of the 6-hydroxy-metabolite, the application was rejected. Bristol-Myers only obtained the '365 Patent after omitting all references in the claim to "buspirone" and any "prodrug," and after making express declarations that the amendments acted to exclude uses of buspirone. . . . [A]ccordingly, Bristol-Myers cannot now reasonably assert a claim for

---

154. *Id.* at 29. On appeal, the Court of Appeals for the Federal Circuit reversed the grant of preliminary injunction on grounds unrelated to claim construction or patent infringement. *Mylan Pharms., Inc. v. Thompson*, 268 F.3d 1323, 1329-33 (Fed. Cir. 2001).

155. *In re Buspirone Patent Litig.*, 185 F. Supp. 2d 340 (S.D.N.Y. 2002), *judgment entered*, No. MDL 1413, 2003 U.S. Dist. LEXIS 26538 (S.D.N.Y. April 11, 2003).

156. The *In re Buspirone* litigation also involved significant questions of antitrust law centering around Bristol-Myers alleged attempts impermissibly to extend their patent monopoly by seeking patent protection for metabolites of buspirone. *See generally*, Tim Meade, *In re Buspirone Patent and Antitrust Litigation*, 9 RICH. J.L. & TECH. 1 (2002). For a general discussion of the antitrust issues implicated by patents claiming metabolites, see Christine S. Paine, Comment, *Brand-Name Drug Manufacturers Risk Antitrust Violations by Slowing Generic Production Through Patent Layering*, 33 SETON HALL L. REV. 479 (2003).

157. *In re Buspirone*, 185 F. Supp. 2d at 363.

158. *Id.* at 353 (citing U.S. Patent No. 6,150,365 col. 16 (filed June 6, 2000)).

the use of buspirone. Hence, the '365 Patent does not cover any uses of buspirone.<sup>159</sup>

Finally, the court held that if the '365 patent claim were construed to cover the use of busiprone, as Bristol-Myers urged, then the claim would be anticipated based on 35 U.S.C. § 102(b) because busiprone had been sold as a treatment for anxiety and its use for treating anxiety had been both published and public at least one year prior to the earliest priority date of the '365 patent.<sup>160</sup> Furthermore, the court rejected BMS's proposed claim construction because if 6-hydroxy-busiprone were reliably produced by administration of busiprone, then the claim of the '365 patent would be inherently anticipated by *in vivo* conversion of busiprone into its 6-hydroxy-busiprone metabolite.<sup>161</sup> Thus, the court concluded that a narrow claim construction, which included use of 6-hydroxy-busiprone *per se* but excluded use of busiprone, would be required to avoid invalidity of the claim.<sup>162</sup>

5. *Novartis Pharmaceuticals Corp. v. Eon Labs Manufacturing Inc.*

Novartis Pharmaceuticals Corporation and allied companies owned U.S. Patent No. 5,389,382 ('382 patent), which included claims directed to a hydrosol encapsulating the immunosuppressant drug cyclosporin.<sup>163</sup> Cyclosporin is difficult to administer to a patient because it is fairly insoluble in water, a problematic characteristic within the wet interior of the human digestive system.<sup>164</sup> The '382 patent disclosed and claimed increasing the effective solubility of cyclosporin by dissolving it "in a water-miscible solvent and then adding a comparatively large amount of water to that solution."<sup>165</sup> The result was a mixture of water and tiny particles containing cyclosporin that can be absorbed more easily from a patient's digestive system.<sup>166</sup>

Novartis sued Eon Labs Manufacturing, Inc. (Eon) in district court for infringing '382 patent claims, despite the fact that Eon's product was a capsule containing cyclosporin, ethanol, and no water.<sup>167</sup> Novartis ad-

---

159. *Id.* at 359 (internal citations omitted).

160. *Id.* at 359-63.

161. *Id.* at 362.

162. *Id.*

163. *Novartis Pharms. Corp. v. Eon Labs Mfg., Inc.*, 363 F.3d 1306, 1307 (Fed. Cir. 2004).

164. *Id.*

165. *Id.*

166. *Id.*

167. *Id.*

vanced an *in vivo* conversion theory of infringement, contending that “when one of Eon’s capsules is ingested an infringing hydrosol is formed when the capsule mixes with the aqueous environment of the user’s stomach.”<sup>168</sup> The district court granted Eon summary judgment of noninfringement, both literally or under the doctrine of equivalents, based on the court’s construction of the claim element “hydrosol” as including only synthetic mixtures, and excluding those produced by *in vivo* conversion in a patient’s stomach.<sup>169</sup> Novartis appealed to the Federal Circuit.

A split Federal Circuit panel affirmed the district court’s claim construction and agreed that “‘hydrosol’ as used in the ’382 patent [was] limited to an aqueous medicinal preparation prepared outside the body.”<sup>170</sup> The majority affirmed the grant of summary judgment of noninfringement.<sup>171</sup>

In arriving at its decision, the panel majority distinguished two previous decisions of the Federal Circuit involving *in vivo* conversion.<sup>172</sup> It distinguished *Zenith* because the claim there involved a “specific chemical compound,” cefadroxil monohydrate, and the plain language of the claim

---

168. *Id.*

169. *Id.* at 1308.

170. *Id.* at 1312.

171. *Id.*

172. Judge Clevenger dissented from the panel majority’s decision and would have defined “hydrosol” broadly enough to place a hydrosol of cyclosporin within the scope of the ’382 patent’s claims. *Novartis*, 363 F.3d at 1313–14. He also disagreed with the panel majority’s interpretation of “medicine” as “things made outside the body.” *Id.* at 1315. Rather, Judge Clevenger characterized “medicines” as much broader:

Our case law has long recognized that medicines claimed in patents can be made inside or outside the body, and that infringement will lie in either case if the proper proofs are made. . . . In all of them, we have a “medicine” whose ordinary meaning carries no manufacturing site limitations. *See Schering Corp. v. Geneva Pharms. Inc.*, 339 F.3d 1373 (Fed. Cir. 2003); *Hoechst-Roussel Pharms., Inc. v. Lehman*, 109 F.3d 756, 759 (Fed. Cir. 1997); *Zenith Labs.*, 19 F.3d at 1421–22. Each of these precedents involved medical preparations. But until this case, no one had suggested that a suspect dictionary definition of the term “medicine” should be used to deny a patentee the right to prove infringement when the claimed composition is formed as a medicine in the body following the ingestion of a different composition that was manufactured outside the body.

*Id.* at 1316. Interestingly, in none of the three cases cited here by Judge Clevenger did the Federal Circuit find infringement. This reasoning stands in clear contrast to the position of the panel majority, where “medicine” was limited to “a preexisting product that is administered to treat disease and therefore must necessarily be prepared outside the body.” *Id.* at 1309.

was clear and unambiguous.<sup>173</sup> Furthermore, the claim contained “no express or implied pre-ingestion limitation,” unlike the ’382 patent claims.<sup>174</sup> Next, the panel majority distinguished *Schering Corp. v. Geneva Pharmaceuticals, Inc.*<sup>175</sup> by noting that *Schering* involved inherent anticipation of a claim covering a metabolite that the parties specifically agreed was produced by *in vivo* conversion. Here, the majority noted, the parties disagreed about whether the product of *in vivo* conversion—the hydrosol—was covered by a ’382 patent claim.<sup>176</sup>

#### 6. Discussion

Prior to *in vivo* conversion cases involving therapeutic drugs, the court in *Feed Service v. Kent Feeds* noted a crucial distinction between deliberate addition of a chemical compound and generation of that same chemical compound through a natural process. As the court explained:

Defendants do not add alcohol to their feed supplements and plaintiff does not charge them with that. The charge of infringement is based on the use by defendants of fermented molasses which provides the alcohol in question as a natural occurring event. We have concluded that the patent in suit is limited to the teaching of the addition of alcohol in feed supplements. The fact that the defendants’ Bovino product may reach the same result as plaintiff’s Morea is not conclusive of the determination of infringement.<sup>177</sup>

The distinction between addition of ethanol and generation of ethanol via “a natural occurring event” is particularly striking because the court imputed an element recited only in process claims 1-7 of the ’332 patent (that is, “incorporating . . . ethanol”)<sup>178</sup> to product claims 8–21 lacking that element.<sup>179</sup> It is this distinction between ethanol deliberately added to feed and ethanol generated in the feed *in situ* by a “natural occurring event” that determines the results of this case. Similarly, on several occasions, courts in *in vivo* conversion cases have employed claim construction to limit the scope of claims to synthetic versions of metabolites, thus ex-

---

173. *Id.* at 1311.

174. *Id.*

175. *Schering Corp. v. Geneva Pharms.*, 339 F.3d 1373 (Fed. Cir. 2003).

176. *Novartis*, 363 F.3d at 1311-12.

177. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 528 F.2d 756, 764 (7th Cir. 1976) (emphasis added).

178. U.S. Patent No. 2,808,332 col.7 l.32-65 (filed Feb. 17, 1955).

179. ’332 Patent col.7 l.66, col.8 l.1-62.

cluding coverage of the same metabolites produced within the human body by *in vivo* conversion.

The courts in both *Marion Merrell v. Baker Norton* and *Mylan v. Thomas* found noninfringement based on a distinction between naturally occurring (by *in vivo* conversion) and synthetically produced metabolites. The court in *Marion Merrell v. Baker Norton* used various strands of evidence, including the specification's silence on *in vivo* conversion and the absurd implications of construing claims to cover products of *in vivo* conversion,<sup>180</sup> to support its conclusion that only "synthetically produced TAM" was covered by claims of the '129 patent, and, therefore, that terfenadine would not infringe.<sup>181</sup> Similarly, the court in *In re Mylan* interpreted the '365 patent claim, covering administration of 6-hydroxy-buspirone, as likely to exclude 6-hydroxy-buspirone produced as a metabolite by *in vivo* conversion of its precursor drug, buspirone, and instead likely to cover only direct administration of the product.<sup>182</sup>

In at least two separate cases, the courts have construed words or phrases narrowly to avoid a finding of infringement triggered by *in vitro* conversion. The court in *In re Busiprone* construed the word "dose" in the claim of the '365 patent to exclude metabolites produced by *in vivo* conversion:

The idea of a "dose" as a quantity that is "taken at one time" has a clear meaning in reference to an externally-measured amount of a substance that is to be ingested or administered into the body all at once, but would have no precise meaning if used to refer to *in vivo* levels in the bloodstream, which are constantly changing.<sup>183</sup>

---

180. See *Marion Merrell Dow Inc. v. Baker Norton Pharms., Inc.*, 948 F. Supp. 1050 at 1054. The court stated:

Baker Norton persuasively points out that if as MMD suggests the term "compound" refers to impure TAM created in the body by metabolism, claim 10 could be construed as the removal of impure TAM from human bodies to then be combined pharmaceutically with a synthetic, or pure, carrier, which as a practical matter the Court finds to be a tenuous assertion leading to an absurd result.

*Id.*

181. See *Marion Merrell Dow Inc. v. Baker Norton Pharms., Inc.*, 948 F. Supp. 1050, 1054-55, 1057 (S.D. Fla. 1996).

182. *Mylan Pharms., Inc. v. Thompson*, 139 F. Supp. 2d 1, 22-26 (D.D.C. 2001).

183. *In re Busiprone Patent Litig.*, 185 F. Supp. 2d 340, 353 (S.D.N.Y. 2002).

In comparison, the majority in *Novartis v. Eon* construed the “hydro-sol” in claims of the ’382 patent to be “medicinal” in nature.<sup>184</sup> Consequently, the claims were interpreted to be “limited to a medicinal preparation consisting of a dispersion of solid particles in a liquid colloidal solution prepared outside the body.”<sup>185</sup> Thus, claim construction has been employed repeatedly to exclude from patent claims products of *in vivo* conversion that arise within the human body.

#### IV. THE PHYSIOLOGICAL STEPS DOCTRINE

Courts in the United States have repeatedly considered whether transformation of a drug via *in vivo* conversion into a metabolite can trigger infringement of patent claims covering the metabolite or methods of using the metabolite. A growing number of such infringement disputes have yielded final judgments, but not one of them found infringement.<sup>186</sup>

Courts have employed diverse rationales to avoid finding infringement in *in vivo* conversion cases. At least one court has pointed to difficulties of obtaining sufficient evidence of infringing products from within the human body. Others have relied upon inherency (*i.e.*, inherent anticipation) where there has been previous use, public knowledge, or sale of a precursor compound that is necessarily transformed by *in vivo* conversion into a claimed product. Other courts have interpreted as meaningful difference between “synthetic” and “natural” biochemicals. Still others have attributed significance to whether claimed medicinal substances are located inside or outside the body. The most parsimonious explanation for this diversity of rationales, but unanimity of findings of noninfringement, is a discomfort with the very idea that a product arising naturally within the body, or the activity of the body itself, can infringe, let alone be the sub-

---

184. *Novartis Pharms. Corp. v. Eon Labs Mfg., Inc.*, 363 F.3d 1306, 1310-11 (Fed. Cir. 2004).

185. *Id.* (defining medicinal preparation as “a preexisting product that is administered to treat disease and therefore must necessarily be prepared outside the body”).

186. In *Ortho Pharm. Corp. v. Smith*, 959 F.2d 936 (Fed. Cir. 1992), two of norgestrel’s *in vivo* products, norgestrel and norgestrel acetate, independently infringed claims to the ’332 patent under the doctrine of equivalents. Thus, though it did involve a product of *in vivo* conversion, it did not base its finding of infringement triggered by *in vivo* conversion of a product claimed in a patent. For expanded explanation see *supra* Section III.A.I.I.A.1. Another case, *Ortho-McNeil Pharm., Inc. v. Mylan Labs., Inc.*, 348 F. Supp. 2d 713 (N.D.W. Va. 2004), found infringement of a claim covering the antimicrobial compound levofloxacin; however, levofloxacin is the levorotatory enantiomer of a racemic mixture, and, once it enters the human body, though it is possible that it undergoes a physical separation from the dextrorotatory enantiomer, it does not undergo any change in chemical form.

ject matter of, a valid and enforceable patent claim. This Article offers a name for this explanation: the Physiological Steps Doctrine.

### A. The European Equivalent

Europe endorses explicitly what American courts appear to endorse implicitly. European patent law expressly limits the patentability of inventions relating to the human body, including methods of medical surgery, therapy, and diagnosis. The European Patent Convention (EPC) Article 53(c) places limits on patentable subject matter related to biological entities, stating that “European patents shall not be granted in respect of: . . . methods for treatment of the human or animal body by surgery or therapy and diagnostic methods practised on the human or animal body,” though it does add that “this provision shall not apply to products, in particular substances or compositions, for use in any of these methods.”<sup>187</sup> EPC Article 52(4), which was replaced by Article 53(c), declares that such subject matter “shall not be regarded as inventions which are susceptible of industrial application.”<sup>188</sup> More specifically, section (1) of Rule 29 (“The human body and its elements”) of the Implementing Regulations to the Convention on the Grant of European Patents declares that “[t]he human body, at the various stages of its formation and development, and the simple discovery of one of its elements . . . cannot constitute patentable inventions.”<sup>189</sup>

On the other hand, Rule 29(2) allows that “[a]n element isolated from the human body or otherwise produced by means of a technical process . . . may constitute a patentable invention, even if the structure of that element is identical to that of a natural element.”<sup>190</sup> The Guidelines for Examination in the European Patent Office clarifies that “[s]uch an element is not a priori excluded from patentability since it is, for example, the result of technical processes used to identify, purify and classify it and to produce it outside the human body, techniques which human beings alone are capable of putting into practice and which nature is incapable of ac-

---

187. Convention on the Grant of European Patents, art. 53(c), Nov. 29, 2000, 2001 O.J.E.P.O. SPEC. ED. NO. 1 38 [hereinafter EPC 2000].

188. Convention on the Grant of European Patents, art. 52(4), Oct. 5, 1973, 1065 U.N.T.S. 255 [hereinafter EPC 1973].

189. Implementing Regulations to the Convention on the Grant of European Patents, pt. II, ch. V, R29 § 1 (Dec. 13, 2006), available at <http://www.epo.org/patents/law/legal-texts/html/epc/1973/e/ma2.html> [hereinafter Implementing Regulations].

190. *Id.* at § 2.

complishing itself.”<sup>191</sup> Thus, the EPC recognizes a distinction between the patentability of chemical inventions practiced outside and inside the human body. An “element isolated from the human body”<sup>192</sup> or “produce[d] . . . outside the human body”<sup>193</sup> may constitute patentable subject matter, but, by implication, an element not isolated from, or produced inside, the human body is unpatentable. Similarly, the World Trade Organization Agreement on Trade-Related Aspect of Intellectual Property offers very comparable provisions in Articles 27(2) and (3).<sup>194</sup>

This patentability criterion is consistent with the Physiological Steps Doctrine, and with the *in vivo* conversion cases discussed above. It would thus appear that European patent law definitively encompasses a Physiological Steps Doctrine, in contrast with the implicit Physiological Steps Doctrine of U.S. patent law.

## B. Within the United States

U.S. law offers no existing theory that can explain why no court has ultimately found infringement of a patent claim by a product or process of *in vivo* conversion. It is highly improbable that such a one-sided outcome has occurred merely by chance. If the odds of a patent owner obtaining a finding of infringement in an *in vivo* conversion case were even (that is, 50%), the unanimous result of *in vivo* conversion cases in failing to find infringement would be equivalent to flipping a coin ten times in a row, and getting heads every single time, a result whose odds are less than 0.01%. Based on such stark math, it would appear that courts are reluctant to allow the involuntary activity of a human body to trigger patent infringement. This suggests an unrecognized explanation that underlies *in vivo* conversion court decisions.

During the middle of the 20th Century, the courts and the USPTO developed a legal doctrine governing the patentability of claims involving “mental steps.”<sup>195</sup> The Mental Steps Doctrine rendered unpatentable any patent claim to a process made up of purely mental steps.<sup>196</sup> In a famous

---

191. European Patent Office, *Guidelines for Examination in the EPO*, pt. C, ch. IV, § 3.2, available at [http://www.epo.org/patents/law/legal-texts/html/guix/e/c\\_iv\\_3\\_2.htm](http://www.epo.org/patents/law/legal-texts/html/guix/e/c_iv_3_2.htm) (last updated Dec. 13, 2007) [hereinafter EPO Guidelines].

192. Implementing Regulations, *supra* note 189, at § 2.

193. EPO Guidelines, *supra* note 191.

194. Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C art. 27(2)-(3), 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994) [hereinafter TRIPs Agreement].

195. 1 DONALD S. CHISUM, CHISUM ON PATENTS § 1.03[6] (2007).

196. *See* Gottschalk v. Benson, 409 U.S. 63, 67 (1972).

statement of this rule, the court in *In re Abrams*, declared that “[i]t is self-evident that thought is not patentable.”<sup>197</sup>

Patent law itself strongly suggests that human thought itself should not be patentable subject matter for at least two reasons. Natural phenomena, such as “laws of nature, physical phenomena, and abstract ideas, have been held not patentable.”<sup>198</sup> Human thought falls within at least two of these specific categories of unpatentable subject matter: thoughts themselves surely qualify as “abstract ideas”; and, the physiological processes involving neurons, neural networks, and electrical and neurochemical signals by which thoughts are generated within the brain are “physical phenomena.”

Just as thoughts result from natural human physiology, so are metabolites produced by the natural *in vivo* conversion of precursor chemicals. Thus, in humans neither thoughts themselves nor products of *in vivo* conversion themselves should qualify as patentable subject matter. In fact, the Mental Steps Doctrine can be viewed as merely a subset of a broader Physiological Steps Doctrine that precludes patentability of claims covering products of human physiological processes.

There are particular intimations of this Physiological Steps Doctrine in the judicial decisions involving *in vivo* conversion.<sup>199</sup> The court’s opinion in *In re Omeprazole* included a statement that lends a more direct form of support for Physiological Steps Doctrine. In this case, the ’499 patent included claims purporting to cover sulphenamides, metabolites produced by *in vivo* conversion of the drug omeprazole.<sup>200</sup> In explaining why claims to sulphenamides themselves would be invalid, the court stated that “[b]y claiming patent protection for sulphenamides formed *in vivo* after the oral administration of omeprazole, Astra has merely attempted to patent the unpatentable—‘a scientific explanation for the prior art’s functioning.’ ”<sup>201</sup> Despite the formal use of inherency doctrine as the rationale for its decision, the court classified metabolites produced by *in vivo* conversion within the category of natural phenomena. Once a patient has ingested a drug, metabolites of that drug produced within the human body through the processes of human physiology may provide “a scientific ex-

---

197. *In re Abrams*, 188 F.2d 165, 168 (C.C.P.A. 1951).

198. *Diamond v. Chakrabarty*, 447 U.S. 303, 309 (1980).

199. *See supra* Part III.

200. *In re Omeprazole Patent Litig.*, No. MDL 1291, 2001 WL 585534, at \*1-2 (S.D.N.Y. May 31, 2001).

201. *Id.* at 12 (quoting *Atlas Powder Co. v. Ireco, Inc.*, 190 F.3d 1342, 1347 (Fed. Cir. 1999)).

planation of [the drug's] functioning," but they are unpatentable subject matter.

After finding claims 1 and 3 of the '233 patent invalid, the Federal Circuit in *Schering v. Geneva* stated that its conclusion on inherent anticipation "does not preclude patent protection for metabolites of known drugs."<sup>202</sup> However, the Federal Circuit then outlined a very strict standard governing how patent protection for products of *in vivo* conversion might be attained through "proper claiming."<sup>203</sup> "[Naturally occurring] metabolites may not receive [patent] protection via compound claims . . . [because] [s]uch bare compound claims include within their scope the recited compounds as chemical species in any surroundings, including within the human body as metabolites of a drug."<sup>204</sup> Instead, "the metabolite may be claimed in its pure and isolated form . . . or as a pharmaceutical composition (e.g., with a pharmaceutically acceptable carrier). The patent drafter could also claim a method of administering the metabolite or the corresponding pharmaceutical composition."<sup>205</sup> However, according to this unanimous opinion of the Federal Circuit, one cannot obtain patent protection for a metabolite produced by *in vivo* conversion of a precursor drug, adding further support for the existence of an unarticulated Physiological Steps Doctrine.

Thus, whether a court employs evidentiary rationales, inherency doctrine, or claim construction, the result is consistently and predictably the same: patent claims purporting to cover products of *in vivo* conversion are either invalid, unenforceable, or not infringing. Unlike explanations involving lack of evidence and inherency, the Physiological Steps Doctrine is consistent with the ultimate decisions in all conversion cases. The "natural occurring event" of *Feed Service v. Kent Feeds*,<sup>206</sup> the "synthetically produced TAM" of *Marion Merrell v. Baker Norton*,<sup>207</sup> the direct administration of 6-hydroxy-buspirone in *Mylan Pharms., Inc v. Thompson*,<sup>208</sup> the externally administered "dose" in *In re Buspirone*,<sup>209</sup> the unpatentability of metabolites produced within the human body by *in vivo* conversion

---

202. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1381 (Fed. Cir. 2003).

203. *Id.*

204. *Id.*

205. *Id.*

206. *Feed Serv. Corp. v. Kent Feeds, Inc.*, 528 F.2d 756, 764 (7th Cir. 1976).

207. *Marion Merrell Dow, Inc. v. Baker Norton Pharms., Inc.*, 948 F. Supp. 1050, 1057 (S.D. Fla. 1996).

208. *Mylan Pharms., Inc. v. Thompson*, 139 F. Supp. 2d 1, 23 (D.D.C. 2001).

209. *In re Buspirone Patent Litig.*, 185 F. Supp. 2d 340, 353 (S.D.N.Y. 2002).

of *In re Omeprazole*,<sup>210</sup> *Schering v. Geneva*'s rule that "[naturally occurring] metabolites may not receive [patent] protection via compound claims[.]"<sup>211</sup> and *Novartis v. Eon*'s "limit[] to a medicinal preparation . . . outside the body"<sup>212</sup> all suggest the existence of an implicit Physiological Steps Doctrine in U.S. patent law. Unlike explanations involving lack of evidence and inherency, the Physiological Steps Doctrine is consistent with the ultimate decisions in all *in vivo* conversion cases.<sup>213</sup>

## V. CONCLUSION

The unanimity of results in cases involving patent infringement triggered by *in vivo* conversion is striking. In fact, its very improbability suggests a common underlying explanation for why *in vivo* conversion does not ever seem to trigger patent infringement. Explanations based on inherency or a lack of evidence provide a satisfactory explanation for only a minority of *in vivo* cases. The Physiological Steps Doctrine, which suggests that products and processes of *in vivo* conversion are unpatentable subject matter under U.S. patent law, offers an explanation that spans all *in vivo* conversion cases. Though the rationales offered to explain the results in a number of *in vivo* conversion cases are suggestive, there are several advantages for a more explicit recognition of the Physiological Steps Doctrine. Consistent with much international, European, and U.S. patent law, the Physiological Steps Doctrine provides a theoretical underpinning to explain the results in cases involving products and processes of *in vivo* conversion. This theoretical underpinning not only has explanatory power for interpreting previous case law but is also useful in predicting the outcome of future *in vivo* conversion cases. In addition, the Physiological Steps Doctrine increases the understanding of where inventions involving human beings, and the biological products and processes thereof, fit within the spectrum of patentable subject matter.

---

210. *In re Omeprazole Patent Litig.*, No. MDL 1291, 2001 WL 585534, at \*12 (S.D.N.Y. May 31, 2001).

211. *Schering Corp. v. Geneva Pharms., Inc.*, 339 F.3d 1373, 1381 (Fed. Cir. 2003).

212. *Novartis Pharms. Corp. v. Eon Labs Mfg., Inc.*, 363 F.3d 1306, 1309-10 (Fed. Cir. 2004) (defining medicinal preparation as "a preexisting product that is administered to treat disease and therefore must necessarily be prepared outside the body").

213. Furthermore, an article by Prof. Dan Burk, entitled *Patenting Speech*, may even suggest a Constitutional justification for a Physiological Steps Doctrine. Burk writes that "there would seem to be profound First Amendment implications to the concept of infringement by 'thinking patented thoughts.'" Dan L. Burk, *Patenting Speech*, 79 TEX. L. REV. 99, 140 (2000). If thinking patented thoughts implicates the First Amendment, then surely involuntarily engaging in physiological processes, such as *in vivo* conversion, that trigger infringement would have equally profound Thirteenth Amendment implications.

# SCHOOL ADMINISTRATORS AS CYBER CENSORS: CYBER SPEECH AND FIRST AMENDMENT RIGHTS

*By Tova Wolking<sup>†</sup>*

## TABLE OF CONTENTS

I. INTRODUCTION .....	1507
II. STUDENT FREE SPEECH FRAMEWORK .....	1509
A. SUPREME COURT PRECEDENT .....	1510
B. TWO-PART FRAMEWORK FOR STUDENT FREE SPEECH .....	1512
III. HOW COURTS HAVE MODIFIED THIS FRAMEWORK TO APPLY TO “CYBER SPEECH” .....	1515
A. UNDERGROUND NEWSPAPER CASES IN CIRCUIT COURTS.....	1516
B. NOTABLE LOWER COURT DECISIONS ON CYBER SPEECH.....	1519
C. CYBER SPEECH MODIFICATIONS TO THE SUPREME COURT’S STUDENT FREE SPEECH FRAMEWORK .....	1523
IV. APPLICATION OF THE STUDENT “CYBER SPEECH” FRAMEWORK IN LAYSHOCK V. HERMITAGE SCHOOL DISTRICT .....	1525
V. CONCLUSION .....	1527

## I. INTRODUCTION

The First Amendment to the United States Constitution protects “freedom of speech,”<sup>1</sup> including freedom of “expression” (symbolic, written, or other nonverbal speech). This protection extends to students and teachers

---

© 2008 Tova Wolking.

<sup>†</sup> Assistant Regional Counsel, United States Social Security Administration; J.D., Berkeley School of Law (2008); M.P.H., University of California, Berkeley, School of Public Health (2003); B.S., University of Utah (1998). The author thanks Professor Jennifer Elrod, and the editors of the Berkeley Technology Law Journal for valuable comments. An earlier version of this Article received the Aldo J. Test Writing Award in Intellectual Property (2008) and won first prize in the 2008 Berkeley Technology Law Journal Notes and Comments Competition. The views expressed here are those of Ms. Wolking and not of the United States Social Security Administration or its staff.

1. U.S. CONST. amend. I.

in public schools, who do not “shed their constitutional rights to freedom of speech or expression at the schoolhouse gate.”<sup>2</sup> Even so, “in light of the special characteristics of the school environment,”<sup>3</sup> the expressive rights of public school students “are not automatically coextensive” with those of adults.<sup>4</sup> Therefore, school boards are given wide latitude to implement disciplinary policies that meet the needs of their district, and school administrators may be more restrictive of on-campus student speech than the government may be of citizens’ speech in general. This authority has limits, however, and schools cannot prohibit student speech simply to avoid controversy or unpleasantness.<sup>5</sup> “Under our Constitution, free speech is not a right that is given only to be so circumscribed that it exists in principle but not in fact.”<sup>6</sup>

The Supreme Court has been retreating from this expansive protection of student expression, and creating exceptions that circumscribe free speech. This Comment utilizes a two-part framework that encapsulates its jurisprudence on student speech in public schools.<sup>7</sup> This framework illustrates how the Court has attempted to balance protection of student speech against schools’ ability to assure that the educational process is not disrupted in a way that interferes with other students’ rights.<sup>8</sup> First, courts must look to the speech act itself, and whether school officials reasonably limited the expression based on legitimate pedagogical concerns.<sup>9</sup> If not, then the court must move on to the second inquiry, whether the speech act created an actual or foreseeable disruption of school functioning.<sup>10</sup>

The Supreme Court’s student free speech jurisprudence is grounded in expressive activity that either occurs at, or makes a physical appearance on school grounds or at school sponsored events. Cyber speech, on the other

---

2. *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503, 506 (1969).

3. *Morse v. Frederick*, 127 S. Ct. 2618, 2622 (2007) (quoting *Tinker*, 393 U.S. at 506).

4. *Bethel Sch. Dist. No. 403 v. Fraser*, 478 U.S. 675, 682 (1986).

5. *Tinker*, 393 U.S. at 509.

6. *Id.* at 513.

7. I created this framework based on the Court’s discussion in *Morse*, the Court’s most recent decision on the issue of student speech in public schools. While the Court did not explicitly call it a two-part test, it provided a detailed explanation of the two levels of analysis that it has used. *Morse*, 127 S. Ct. at 2625-27.

8. *See id.*

9. *Id.* at 2626-27 (allowing reasonable restriction of speech that pertains to “the special characteristics of the school environment” even where there is not a substantial disruption).

10. *Id.* at 2627 (allowing restriction of “political” speech that causes “substantial disruption” within the school).

hand, is boundary-less, presenting a quandary for schools and students, and in turn, the district courts that adjudicate the issues. As a result, courts have applied the Supreme Court's two-part framework and modified it to incorporate an analysis of the nexus between off-campus student speech and associated on-campus disruption.

This Comment will provide an overview of judicial decisions on the issue of student electronic speech that implicate public schools, public school teachers, and administrators. The Comment will examine the test that courts are using to determine the extent of schools' power to censor students' electronic expression produced off school grounds. Next, the Comment will deconstruct *Layshock v. Hermitage School District*, a 2007 district court case concerning a MySpace parody profile of a high school principal.<sup>11</sup> The Comment will use the case to illustrate how the framework is being applied to protect instances of student off-campus cyber speech. Finally, the Comment will conclude with a discussion about the importance of this lower court jurisprudence, which is providing vital push-back to Supreme Court restrictions on First Amendment protection of student speech.

## II. STUDENT FREE SPEECH FRAMEWORK

In a line of cases dating from 1969 to 2007, the Supreme Court has shaped a framework for evaluating constraints on student speech in public schools, differentiating between categorical restriction of speech that does not jibe with the educational mission of the public schools, and circumscription of otherwise unrestricted speech, because of its foreseeable or actual disruption of school activities.<sup>12</sup> The two-part analysis applies to speech that occurs on school grounds or at school sponsored events and can be broken into two stages. The first part entails an examination of the speech itself and whether it is protected by the First Amendment, given

---

11. *Layshock v. Hermitage Sch. Dist.*, 496 F. Supp. 2d 587 (W.D. Pa. 2007). As of this writing, *Layshock* is the most recent published federal case on the topic.

12. See *infra* section II.B. In *Morse*, the Court differentiated a *Fraser*-type analysis, which permits school officials to categorically limit certain types of speech that pertain to the "special characteristics of the school environment," from a *Tinker*-type analysis, which allows school officials to constrain otherwise protected speech if it causes or may cause "substantial disruption." *Morse*, 127 S. Ct. at 2626-27; see also *Layshock*, 496 F. Supp. 2d at 596 (discussing the rule from *Morse* that "student speech cases must be resolved 'in light of the special characteristics of the school environment' " and "substantial disruption" is a separate inquiry) (citations omitted).

the institutional environment of the public school system.<sup>13</sup> If the speech is protected, then the second part of the analysis inquires into whether the speech can be permissibly restricted by the school if it substantially disrupts school activities<sup>14</sup> or if school administrators have a specific fear of disruption.<sup>15</sup>

### A. Supreme Court Precedent

The foundational 1969 case, *Tinker v. Des Moines Independent Community School District*, created the initial rule that protects students' non-disruptive speech on public school campuses.<sup>16</sup> Amidst the heat and divisiveness of the Vietnam War, Mary Beth Tinker and two other Iowa teenagers wore black armbands to school in silent protest of United States' involvement in the war.<sup>17</sup> In response, school administrators who wished to avoid controversy suspended the students until they agreed to remove their armbands.<sup>18</sup> In its majority opinion, the Court upheld students' constitutional right to express themselves, particularly their communication of political views, so long as it does not "materially or substantially" interfere with school activities or "collid[e] with the rights of others."<sup>19</sup> *Tinker* thus shielded on-campus student speech from discipline, so long as it did not disrupt school activities.

In 1983 the Supreme Court retreated from the broad protections provided by the Warren Court in *Tinker*, when the conservative majority in *Bethel School District Number 403 v. Fraser* sided with the school district to limit student speech.<sup>20</sup> In *Fraser*, Matthew Fraser, a student at Bethel High School in Washington, had been suspended for presenting a speech consisting of an extended sexual metaphor. The Court determined that his speech was not protected by the First Amendment.<sup>21</sup> The Court marked a

---

13. See *Morse*, 127 S. Ct. at 2626-27 (explaining that, while the analysis in *Fraser* was "not entirely clear," its holding stood for the proposition that students' on-campus speech receives less protection than speech by adults in other settings, and because of the "special characteristics" of the school environment, some speech can be further restricted even if it does not disrupt school activities).

14. See *Morse*, 127 S. Ct. at 2626-27 (explaining that a court making a *Fraser*-type finding need not move on to the *Tinker* "substantial disruption" analysis).

15. *Layshock*, 496 F. Supp. 2d at 597 ("It is clear that school administrators need not wait until a 'substantial disruption' has already occurred prior to taking action.").

16. *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503 (1969).

17. *Id.* at 504.

18. *Id.*

19. *Id.* at 512-13.

20. *Bethel Sch. Dist. No. 403 v. Fraser*, 478 U.S. 675 (1986).

21. *Id.* at 685.

new exception to *Tinker*, such that schools could discipline students for “lewd” speech, free from First Amendment restraint.<sup>22</sup> Notably, the Court did not base its holding on the effect of the speech, meaning it did not point to a “substantial disruption,” but rather the authority of the school to limit a special type of speech that would not be restricted outside of the school environment.<sup>23</sup> Further, the Court held that the definition of this special category of “lewd” speech rests with the school, as part of its “basic educational mission”<sup>24</sup> to teach the “boundaries of socially appropriate behavior.”<sup>25</sup>

In *Hazelwood School District v. Kuhlmeier*, the Court refined the concept of “basic educational mission” described in *Fraser*.<sup>26</sup> The Court explained that “[a] school must be able to set high standards for the student speech that is disseminated under its auspices . . . and may refuse to disseminate student speech that does not meet those standards.”<sup>27</sup> A school may refuse to disseminate the speech even if it is not disruptive of the school environment.<sup>28</sup> In *Kuhlmeier*, school administrators censored student articles involving issues of teen pregnancy and divorced parents.<sup>29</sup> The Court deemed this forum (a school newspaper) a “supervised learning experience” that comprised school sponsored speech.<sup>30</sup> The Court held for the school, allowing it to limit student speech that could be reasonably perceived as “bear[ing] the imprimatur of the school”<sup>31</sup> so long as the restriction is “reasonably related to legitimate pedagogical concerns.”<sup>32</sup>

The Court’s most recent restriction of the *Tinker* rule occurred in mid-2007 when it decided *Morse v. Frederick*.<sup>33</sup> Writing for the majority,

---

22. *Id.*

23. *Fraser* created a conscribed category of “lewd” speech, reasoning that “schools must teach by example the shared values of a civilized social order.” *Fraser*, 478 U.S. at 683. The Court in *Morse* summarizes this reasoning, noting that in *Fraser*, it did not employ a “substantial disruption” analysis, but rather linked school discipline of “lewd” speech with its interest in upholding its educational mission. *See Morse v. Frederick*, 127 S. Ct. 2618, 2626-27 (2007).

24. *Fraser*, 478 U.S. at 685.

25. *Id.* at 681.

26. *Id.* at 685. *Fraser* drew on dicta in *Tinker*, regarding the “special characteristics of the school environment.” *See Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503, 506 (1969).

27. *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260, 271-72 (1988).

28. *Id.* at 271.

29. *Id.* at 263.

30. *Id.* at 270.

31. *Id.* at 271.

32. *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260, 273 (1988).

33. *Morse v. Frederick*, 127 S. Ct. 2618 (2007).

Chief Justice Roberts etched a new exception into the realm of protected student speech; any speech on school grounds or during a school activity that advocates illegal drug use now falls outside the realm of protected speech.<sup>34</sup> Specifically, the Court withheld First Amendment protection from an 18-year-old student in Alaska who held up a “BONG HiTS 4 JE-SUS” banner during a school sponsored viewing of the 2002 Olympic Torch Relay.<sup>35</sup> The Court explained that *Tinker* was “not absolute,” such that the “substantial disruption” analysis was not required when issues affecting the special characteristics of the school (such as “lewd” speech and “drug promoting” speech) were implicated.<sup>36</sup>

Thus, without calling it a “test,” the Court in *Morse* explained its line of jurisprudence from *Tinker* to *Fraser* to *Kuhlmeier*<sup>37</sup> and implicitly separated student free speech analysis into two parts, where some “inappropriate” types of speech can be restricted even without evidence or reasonable fear of “substantial disruption” of the school environment.<sup>38</sup>

## **B. Two-Part Framework for Student Free Speech**

This line of Supreme Court jurisprudence on student free speech falls neatly into a two-part framework (see Figure 1). The first part of the framework looks at the content of the speech or expression, and the second part looks at its effect (or foreseeable effect).

---

34. *Id.* at 2625.

35. *Id.* at 2622.

36. *Id.* at 2626-27.

37. *Id.* at 2625-28.

38. *See id.* (distinguishing between its analyses in *Fraser* and *Tinker*, and citing *Kuhlmeier* to explain that the analysis in *Fraser* rested on the school’s educational mission to prohibit inappropriate sexual content, whereas *Tinker* was based on a substantial disruption analysis).

**Fig. 1: FRAMEWORK FOR STUDENT EXPRESSION THAT OCCURS ON SCHOOL GROUNDS OR AT SCHOOL SPONSORED EVENTS**

**Part I:**

*Is the expression contrary to the school's educational mission within the "special characteristics of the school environment"?*

*Speech that is lewd, vulgar, obscene or plainly offensive.*

→ *Not protected (Fraser)*

*Speech advocating illegal drug use.*

→ *Not protected (Morse)*

*Speech that bears the imprimatur of the school.*

→ *Not protected (Hazelwood)*

**Part II:**

*Even though the expression is constitutionally protected, was it appropriate for the school to restrict speech or discipline the student?*

*Caused, or was reasonably likely to cause "substantial disruption" of school activities.*

→ *Permissible to restrict speech (Tinker)*

With some notable exceptions, almost any sort of verbal speech,<sup>39</sup> written speech,<sup>40</sup> or nonverbal conduct<sup>41</sup> in which a student expresses an opinion amounts to "speech" under the First Amendment. But the Court has observed that "political" speech "is at the core of what the First Amendment is designed to protect."<sup>42</sup> However, what constitutes "political" speech, and how this differs from expression of opinions that may or may not have political implications, is far from clear. For example, the Court considered the anti-war armband worn by Mary Beth Tinker to be "political" whereas the student government nomination speech given by Matthew Fraser was not.<sup>43</sup>

While the line between "political" and presumably "apolitical" speech is indistinct, the Court has made it clear that some sorts of speech are not

39. See, e.g., *Bethel Sch. Dist. No. 403 v. Fraser*, 478 U.S. 675 (1986).

40. See, e.g., *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260 (1988).

41. See, e.g., *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503 (1969).

42. See *Morse*, 127 S. Ct. at 2626 (quoting *Virginia v. Black*, 538 U.S. 343, 365 (2003)).

43. Compare *Tinker*, 393 U.S. at 514 (finding Tinker wore the armband to show "disapproval of the Vietnam hostilities"), with *Fraser*, 478 U.S. at 685 ("Unlike the sanctions imposed . . . in *Tinker*, the penalties imposed in this case were unrelated to any political viewpoint.").

protected by the First Amendment. Schools may place heightened limitations on “lewd” speech,<sup>44</sup> speech that could be perceived as school sanctioned or approved,<sup>45</sup> and speech that promotes illegal drug use.<sup>46</sup> All of these excepted areas of speech have been grouped under the rubric of “special characteristics of the school environment.”<sup>47</sup> This nebulous catch-all could feasibly be used to restrict a wide swath of student speech, but for the purpose of practical application, schools may categorically prohibit student expression that falls within these “unprotected” realms as long as the restriction is reasonably related to the school’s educational mission.<sup>48</sup>

Speech that is not “lewd,” drug-promoting, or school-sponsored moves on to the second part of the test, which looks at the effect of the speech. Even if an instance of student expression is constitutionally protected, it may be restricted if it causes substantial disruption of school activities or impinges on the rights of other students.<sup>49</sup> Conversely, school officials’ mere disagreement or discomfort with the message is not sufficient to show a substantial disruption. For example, a student athlete who complains in writing about the abusive tactics of his basketball coach is constitutionally protected.<sup>50</sup> But if the student “substantially disrupts school activity” by engaging in a last minute team-wide boycott of a varsity game, then school administrators could permissibly discipline him for his expressive conduct.<sup>51</sup>

Further, a school need not wait for actual disruption. If it can “point to a well-founded expectation of disruption—especially one based on past incidents arising out of similar speech—the restriction may pass constitutional muster.”<sup>52</sup> Extrapolating from the example above, if the student athlete plans to distribute flyers at the varsity basketball game, and school administrators reasonably believe that the flyers could incite fights or vi-

---

44. See *Fraser*, 478 U.S. at 685.

45. See *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260, 273 (1988).

46. See *Morse v. Frederick*, 127 S. Ct. 2618, 2622 (2007).

47. See, e.g., *id.* at 2629 (“The ‘special characteristics of the school environment,’ and the governmental interest in stopping student drug abuse . . . allow schools to restrict student expression that they reasonably regard as promoting illegal drug use.”) (citations omitted).

48. Note that *Morse* stretches the school’s educational mission to encompass a federal legislative mission to combat illegal drug use in schools. *Id.* at 2629.

49. See *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503 (1969).

50. *Pinard v. Clatskanie Sch. Dist.* 6J, 467 F.3d 755, 768 (9th Cir. 2006).

51. *Id.* at 769.

52. *Killion v. Franklin Reg’l Sch. Dist.*, 136 F. Supp. 2d 446, 455 (W.D. Pa. 2001) (quoting *Saxe v. State Coll. Area Sch. Dist.*, 240 F.3d 200, 212 (3d Cir. 2001)).

olence, the principal's confiscation of the flyers and suspension of the student would likely pass constitutional muster.

### III. HOW COURTS HAVE MODIFIED THIS FRAMEWORK TO APPLY TO "CYBER SPEECH"

Written speech that occurs on school grounds, including the hypothetical flyers described above, as well as comments posted on a website are subject to the same test as other forms of student speech. Even so, when out-of-school conduct has in-school implications, the web has created a new dimensional twist in the two-part analysis. For example, a comment published to a website while at home or a parodic profile created on a non-school computer may penetrate the literal and electronic "walls" of the public school. Therefore, such speech may be received as if it was created at school or distributed on school grounds. Increasingly, courts are being called upon to resolve students' claims that schools' censorship of cyber speech created off-campus has violated their First Amendment free speech rights.

In *Morse*, the Supreme Court acknowledged "some uncertainty at the outer boundaries as to when courts should apply school-speech precedents."<sup>53</sup> Unfortunately, even the most applicable Supreme Court case on which lower courts can rely, *Kuhlmeier*, pertained to in-school speech. *Kuhlmeier* involved the publication of student writing in a school newspaper, and the Court determined that school censorship of student articles was permissible.<sup>54</sup> Under this rubric of "school sponsored speech," it follows that if a student publishes something to the Internet (such as an offensive MySpace profile) as part of a computer class or school-sponsored project, it seems clear that the school has disciplinary discretion. But lower courts are being called upon to adjudicate instances of student expression that have merely penetrated school grounds after they were published. Perhaps the closest historical analogues to cyber speech disputes are cases concerning privately distributed, student published "underground newspapers."<sup>55</sup>

---

53. *Morse*, 127 S. Ct. at 2623-24. The Pennsylvania Supreme Court also noted that "there is little case law" addressing the issue of student speech "that occurred *off of school premises* and was communicated to others via the Internet." *J.S. v. Bethlehem Area Sch. Dist.*, 757 A.2d 412, 419 (Pa. 2000).

54. *Hazelwood Sch. Dist. v. Kuhlmeier*, 484 U.S. 260, 274-76 (1988).

55. See Thomas E. Wheeler II, *Lessons from the Lord of the Flies: The Responsibility of Schools to Protect Students from Internet Threats and Cyber-Hate Speech*, 215 ED. LAW REP. 227, 231 (2007).

### A. Underground Newspaper Cases in Circuit Courts

Before the Internet, high school students who wanted to let off steam about their public school teachers or administrators produced “underground newspapers.” These newspapers and magazines were written and published by students, off-campus, and they were privately distributed.<sup>56</sup> Although they were not produced at school, if students distributed them on campus and they disrupted school, this “speech” fell within the disciplinary purview of the school when the educational process was threatened.<sup>57</sup> In such cases, courts would side with schools, upholding their discipline of the responsible student or students.<sup>58</sup> Conversely, if the publication was about the school, but it was distributed “beyond the schoolhouse gate” and only minimally affected school functions, then it did not fall under the authority of the school.<sup>59</sup> The publication was treated as any other form of speech in a public forum—where only extreme forms of expression, such as obscenity, are stripped of constitutional protection because of their content.<sup>60</sup>

In *Thomas v. Board of Education*,<sup>61</sup> the Second Circuit upheld student expression via an off-campus newspaper. The court held that the high school students’ newspaper did not form a sufficient connection with the school to pull it into the school’s disciplinary reach.<sup>62</sup> The publication, modeled after *National Lampoon*, contained satire about “school lunches, cheerleaders, classmates, and teachers,” as well as articles about prostitution and masturbation.<sup>63</sup> The students took pains to keep the publication separate from their school by writing articles in their homes and after school, paying to have it printed at a local business, and selling it at an off-campus general store.<sup>64</sup> Despite their efforts, a copy of the paper was

---

56. See Jamin B. Raskin, *WE THE STUDENTS: SUPREME COURT CASES FOR AND ABOUT STUDENTS* 67 (2003).

57. *Id.* Court decisions in the 1970s disallowed school authorities from censoring privately produced and distributed publications, unless they could be foreseen to substantially disrupt the educational process.

58. *See id.*

59. *See Thomas v. Bd. of Educ.*, 607 F.2d 1043, 1050 (2d Cir. 1979).

60. *See, e.g., id.* at 1048 (citing several cases in which “we have excluded libel, obscenity, and incitement from the First Amendment’s protective cloak”).

61. *Id.*

62. *Id.*

63. *Id.* at 1045.

64. *Id.* at 1045.

found on-campus.<sup>65</sup> Subsequently, after determining that the newspaper's contents were inappropriate, their principal suspended them.<sup>66</sup>

In a strongly worded opinion, the circuit court expressed immense concern about the implications of allowing school administrators to control the content of students' off-campus speech where it bore little nexus with the school.<sup>67</sup> The court was "intentionally frugal in exposing expression to government regulation."<sup>68</sup> It upheld student free speech because the student publishers "diligently labored to ensure that [the newspaper] was printed outside the school, and that no copies were sold on school grounds."<sup>69</sup> Though some of the articles were transcribed on school typewriters and the finished product was stored in a teacher's closet, the court found that student's activity within the school was nonetheless *de minimis*.<sup>70</sup> Further, the school's punishment could not "withstand the proscription of the First Amendment" allowed for on-campus speech, because "school officials [had] ventured out of the school yard and into the general community where the freedom accorded expression is at its zenith."<sup>71</sup>

An important component of this Second Circuit opinion is its clarification of the distinction between a student's speech as a "student," under the authority of the school, and her speech as a member of the general community. Because of the "unique requirements of the educational process," student speech that falls under a school's disciplinary authority may be held to the *Tinker* "substantial disruption" standard—a higher standard than speech that would ordinarily be protected.<sup>72</sup> But the authority of school officials encompasses the "power to punish one who sp[ea]ks out of turn in class or who disrupt[s] the quiet of the library or study hall,"<sup>73</sup> not the power to police student's off-campus speech.<sup>74</sup> The court chastised the school for overreaching its authority and punishing what it deemed "obscene" speech without any grounding in the constitutional standard for such speech.<sup>75</sup> It also found that if the wholly off-campus student newspa-

---

65. *Id.* at 1045.

66. *Id.* at 1046.

67. *Id.* at 1050-53.

68. *Id.* at 1048.

69. *Id.* at 1050.

70. *Id.*

71. *Id.*

72. *Id.* at 1049.

73. *Id.*

74. The term "off-campus" speech does not include school-sponsored events or field trips that occur off-campus.

75. *Thomas*, 607 F.2d at 1051-52.

per truly fell within the “narrow categories of words that the state may punish” outside of the school environment, the judgment would have to be made by an impartial arbiter, not a school official with a “vested interest in suppressing controversy” for the sake of institutional decorum.<sup>76</sup>

The most recent decision regarding an underground newspaper came out of the Seventh Circuit Court of Appeals.<sup>77</sup> Here, the court sided against the student, Justin Boucher, who had authored an article entitled, “So You Want to Be a Hacker,” in an off-campus publication.<sup>78</sup> The article explained how to “hack” into school computers, and the paper was distributed to students on campus.<sup>79</sup> Upon discovering the article the school board expelled Justin for one year, believing that the article posed a serious threat to school property.<sup>80</sup> The district court determined that the harm to Justin outweighed the harm to the school, and enjoined the school from imposing discipline.<sup>81</sup> The court of appeals agreed that a one-year expulsion was an unreasonable response, but it found that the school’s substantiated fear of tangible harm to its computer systems warranted its choice to impose discipline.<sup>82</sup> Because the article was distributed on campus and advocated on-campus activity, the appellate court overturned the injunction against his school.<sup>83</sup>

This case provides an interesting bridge to cyber speech because it involved a student advocating a mild form of cyber terrorism. It is precisely because school administrators had a reasonable fear that students might be able to overwhelm their school’s technological capabilities that the circuit court determined that the harm to the school outweighed the student’s free speech rights. Further, even though the publication was produced off-campus, it was distributed on the school campus and advocated on-campus activity, which pulled it into the school’s disciplinary authority.<sup>84</sup> Though the Seventh Circuit did not call this link a “nexus,” it follows the reasoning used by other courts in similar underground newspaper disputes.<sup>85</sup>

---

76. *Id.* at 1047, 1051.

77. *Boucher v. Sch. Bd. of Sch. Dist. of Greenfield*, 134 F.3d 821 (7th Cir. 1998).

78. *Id.* at 823.

79. *Id.* at 822.

80. *Id.* at 823.

81. *Id.* at 826.

82. *Id.* at 826-27.

83. *Id.* at 829.

84. *Id.* at 828.

85. *See, e.g., Bystrom v. Fridley High Sch.*, 822 F.2d 747 (8th Cir. 1987) (allowing restriction of student speech, where underground newspaper was distributed on campus, and advocated vandalism and violence against teachers).

Today, instead of underground newspapers, students blow off steam by publishing their opinions in electronic formats, such as blogs and electronic profiles. Unlike newspapers, these electronic publications can be distributed instantaneously among many people at once. So even if the student has no intention of sharing his creation with classmates, it can still easily find its way into the classroom.<sup>86</sup> “*Tinker*’s simple armband, worn silently and brought into a Des Moines, Iowa classroom, has been replaced by [a student’s] complex multi-media website, accessible to fellow students, teachers, and the world.”<sup>87</sup> Further, once the blog or web profile is discovered in a school it can be very difficult for school administrators to erase its electronic imprint or block student access.<sup>88</sup> In response, lower courts are amassing a body of case-law on the extent to which schools can abridge First Amendment protection of students’ off-campus cyber speech.

### **B. Notable Lower Court Decisions on Cyber Speech**

In a 1998 case, *Beussink v. Woodland R-IV School District*, a Missouri district court found that a high school student, Brandon Beussink could not be punished for his off-campus cyber speech.<sup>89</sup> Using his home computer during non-school hours, Brandon created a website that criticized his school, including his teachers, principal, and the school webpage.<sup>90</sup> Brandon did not intend his website to be viewed at school, but a friend saw it on Brandon’s home computer, and later brought it to the attention of school administrators.<sup>91</sup> Upset by the contents of the site, Brandon’s principal suspended him for ten days.<sup>92</sup> In response, Brandon sought a court-ordered injunction.<sup>93</sup> Despite his use of vulgar language, the court upheld Brandon’s free speech rights.<sup>94</sup> The court blocked the school district from imposing the suspension and restricting Brandon from reposting the page,

---

86. See, e.g., *Beussink v. Woodland R-IV Sch. Dist.*, 30 F. Supp. 2d 1175 (E.D. Mo. 1998) (noting that a student showed his website to a friend who, after a subsequent falling out, showed the website to a teacher).

87. *J.S. v. Bethlehem Area Sch. Dist.*, 569 Pa. 638, 665 (2002).

88. See, e.g., *Layshock v. Hermitage Sch. Dist.*, 496 F. Supp. 2d 587, 591-93 (W.D. Pa. 2007) (noting that it took more than a week for the school to block student access to parody MySpace profiles of school principal, and the district had been attempting to block access to MySpace for approximately two months).

89. *Beussink*, 30 F.Supp.2d at 1177.

90. *Id.*

91. *Id.* at 1177-78.

92. *Id.*

93. *Id.* at 1177.

94. *Id.*

noting that schools may not limit student speech merely because they dislike the content of the speech or are upset by it.<sup>95</sup> The court determined that this electronic posting did not form a nexus with the school and did not materially or substantially interfere with school discipline.<sup>96</sup>

Another 2001 Pennsylvania district court case, *Killion v. Franklin Regional School District*, reiterated the principle that disliking or being upset by privately produced student speech that is nonthreatening and nondisruptive is insufficient justification for discipline.<sup>97</sup> Here, the school suspended a student for an email he produced on his own time and on his home computer, but which he sent to a list of friends.<sup>98</sup> The email contained a "Top Ten List" of disparaging comments about the school's athletic director, including statements about the size of his genitals.<sup>99</sup> Like Brandon Beussink, this student did not distribute the list at school, but rather, another "undisclosed student" printed out the list and brought it to campus where it was discovered by school administrators.<sup>100</sup> The district court held that the school district violated the First Amendment by punishing the student for "mere creation" of a web-based list, in the absence of school disruption, and where the student had created the list while in the privacy of his home.<sup>101</sup>

Though not a district court case, *J.S. v. Bethlehem Area School District*, decided by the Pennsylvania Supreme Court in 2002, is illustrative of what it takes to form a nexus between a student's off-campus speech and a resulting school disruption.<sup>102</sup> Like *Beussink*, *J.S.* involved a student who created a website from his home computer.<sup>103</sup> However, the website in this case, entitled "Teacher Sux," contained threatening graphic depictions.<sup>104</sup> The student's algebra teacher was shown with her head severed and dripping with blood, and her face morphing into that of Adolph Hitler.<sup>105</sup> The site also contained a page purportedly devoted to soliciting donations in order to hire a hit man to kill her.<sup>106</sup> In this instance, the court held that the

---

95. Brandon had voluntarily removed the page after receiving disciplinary notice from his school. *Id.* at 1179.

96. *Id.* at 1181.

97. *Killion v. Franklin Reg'l Sch. Dist.*, 136 F. Supp. 2d 446, 455 (W.D. Pa. 2001)

98. *Id.* at 448.

99. *Id.*

100. *Id.* at 449.

101. *Id.* at 458.

102. *J.S. v. Bethlehem Area Sch. Dist.*, 569 Pa. 638 (2002).

103. *Id.* at 643.

104. *Id.* at 644.

105. *Id.* at 645.

106. *Id.*

school district was justified in disciplining the student.<sup>107</sup> The algebra teacher suffered physical and emotional distress, had to take an extensive leave of absence after viewing the page, and her classes had to be taught by substitute teachers.<sup>108</sup>

Because of the effect on both the teacher and the school community,<sup>109</sup> the *J.S.* court agreed with the school that the website, while created off-campus, had substantially disrupted school activities, and had been accessed on-campus.<sup>110</sup> In this iteration of the substantial disruption test, actual disruption or targeting the specific audience of the school and its teachers is a precursor to the nexus analysis.<sup>111</sup> Pursuant to this reasoning, a student website, which was created off-campus and designed to harass a teacher, would not qualify for discipline by school authorities *per se*. However, if the website is accessed at school and brings about its desired effect of frightening the teacher or riling up the school community, then a court would likely find that the student's website was sufficiently connected to on-campus disruption.<sup>112</sup> A sufficient nexus can be found, even if the website was only accessed once on-campus by a student.<sup>113</sup> In coming to this conclusion, the court focused on the fact that the website had threatened a teacher.<sup>114</sup>

Off-campus cyber speech, or any speech for that matter, is not protected by the First Amendment if it moves beyond parody or criticism and poses a "true threat" to personal safety.<sup>115</sup> A "true threat" is a written or oral statement that could be reasonably perceived as a "serious expression of an intention to inflict bodily harm upon or take the life of" the target of

---

107. *Id.* at 643.

108. *Id.* at 646.

109. *Id.* ("[C]omparable to the . . . death of a student or staff member.")

110. *Id.* at 673.

111. *See id.* at 668.

112. *See id.* at 673-74

113. *See* Wheeler, *supra* note 55, at 235 (describing the "J.S. approach" as "[i]f the website is accessed by students at school then the speech will be deemed to have taken place on-campus and the school will be able to regulate it").

114. *J.S.*, 569 Pa. 638, 674 ("The most significant disruption caused by the posting of the website to the school environment was direct and indirect impact of the emotional and physical injuries [to the teacher].").

115. *See id.* at 652 ("A 'true threat' is a certain class of speech that the United States Supreme Court has determined is beyond the protective ambit of the First Amendment."). In *J.S.*, the court found, the "lack of immediate steps taken directly against J.S., and the lack of immediate notification of his parents about the web site, for the extended time period that passed in this case, strongly counters against a conclusion that the statements made in the web site constituted true threats." *Id.* at 658-59.

the statement.<sup>116</sup> Courts use a “reasonable person” standard to determine if an expressive act constitutes a “true threat.”<sup>117</sup> If student cyber speech does pose a threat, it is not constitutionally protected. In such a case, analysis under the on-campus “substantial disruption” portion of the student free speech test would be irrelevant. Only speech that would be ordinarily protected moves on to the second half of the test.

A 2002 Michigan case, *Mahaffey v. Aldrich*, provides a good example of the application of the interaction between “true threat” and “substantial disruption.”<sup>118</sup> In *Mahaffey*, a district court determined that a student’s off-campus contribution to a website, “Satan’s web page,” was an obvious parody that neither posed an actual threat nor disrupted school.<sup>119</sup> Though the site included a list of “people I wish would die,” it contained no direct threats, and it was only one of several lists, including “people that are cool,” “movies that rock,” and “music I hate.”<sup>120</sup> Further, the student had created it “for laughs” without showing it to anyone, and had even included a disclaimer: “PS: NOW THAT YOU’VE READ MY WEB PAGE PLEASE DON’T GO KILLING PEOPLE AND STUFF THEN BLAMING IT ON ME. OK?”<sup>121</sup> Accordingly, the court determined that a reasonable person could not view this site as a serious expression of intent to harm the people on the list.<sup>122</sup> The court then applied the second part of the test for student speech, and because the school officials had no evidence of school disruption, found that their discipline of the student was unconstitutional.<sup>123</sup>

Another cyber speech case addressing both “true threat” and “substantial disruption” was *Emmett v. Kent School District No. 415*.<sup>124</sup> Here, the court sided with a student that had created a mock version of his high school’s webpage, finding that his publication of faux obituaries did not pose a true threat or disrupt school.<sup>125</sup> The page included disclaimers that

---

116. *Mahaffey v. Aldrich*, 236 F. Supp. 2d, 779–785 (E.D. Mich. 2002) (quoting *United States v. Lincoln*, 462 F.2d 1368, 1369 (6th Cir. 1972)).

117. *See Wheeler*, *supra* note 55, at 240 (discussing courts diverging applications of the reasonable person standard—some use the view of the speech-maker and others use the view of the subject of the speech).

118. *Mahaffey*, 236 F. Supp. 2d at 782.

119. *Id.* at 784.

120. *Id.* at 782.

121. *Id.* at 786.

122. *Id.*

123. *Id.*

124. *Emmett v. Kent Sch. Dist. No. 415*, 92 F. Supp. 2d 1088 (W.D. Wash. 2000).

125. *Id.* at 1090.

the page was for entertainment and not sponsored by the school.<sup>126</sup> It also included “tongue-in-cheek” obituaries of two of the student’s friends, and an opportunity to vote for the subject of the next “obituary.”<sup>127</sup> The court acknowledged that, in the face of school shootings, “[w]eb sites can be an early indication of a student’s violent inclinations, and can spread those beliefs quickly to like-minded or susceptible people.”<sup>128</sup> But where the school had “presented no evidence that the mock obituaries and voting on this web site were intended to threaten anyone, did actually threaten anyone, or manifested any violent tendencies whatsoever,” and where “the speech was entirely outside of the school’s supervision or control,” it could not suspend the student.<sup>129</sup>

### C. Cyber Speech Modifications to the Supreme Court’s Student Free Speech Framework

As the cases above illustrate, lower courts have only allowed school authorities to discipline off-campus cyber speech in limited circumstances. In essence, the framework described in Part II has been amended to include two modifications (see Figure 2). The first modification applies to the first part of the student free speech test—the writing itself. When a website or blog is created off-campus the court does not apply the *Fraser*, *Kuhlmeier*, or *Morse* exceptions, but it still must decide whether the speech would “ordinarily” be protected by the First Amendment. Therefore, schools do not have special authority to categorically censor “lewd” or drug-promoting speech that is created off-campus.<sup>130</sup> For example, in *Killion*, the court agreed that the student’s “Top Ten List” contained “lewd and vulgar” statements but since the email was created at the student’s home, “far removed from any school premises or facilities,” and not “associated in any way with his role as a student,” the fact that it was “lewd” was irrelevant to the court’s analysis.<sup>131</sup>

---

126. *Id.* at 1089.

127. *Id.* (“The obituaries were written tongue-in-cheek, inspired, apparently, by a creative writing class last year in which students were assigned to write their own obituary.”).

128. *Id.* at 1090.

129. *Id.*

130. *See, e.g.,* Layshock v. Hermitage Sch. Dist., 496 F. Supp. 2d 587, 595 (W.D. Pa. 2007) (“[B]ecause . . . *Morse* involved school-related speech, *Morse* is not controlling of the instant matter.”).

131. *Killion v. Franklin Reg’l Sch. Dist.*, 136 F. Supp. 2d 446, 457 (W.D. Pa. 2001).

**Fig. 2: FRAMEWORK FOR ELECTRONIC (“CYBER”) SPEECH CREATED OFF SCHOOL GROUNDS & WITHOUT SCHOOL RESOURCES**

**Part I:**

*Is the expression ordinarily protected by the First Amendment?*

*Poses a “true threat” (or is obscene or libelous).*

*→ Not protected*

**Part II:**

*Even though the expression occurred off-campus, did it form a nexus with the school?*

*Student brought it to school or it was accessed at school.*

*→ Nexus formed and Part III applied*

**Part III:**

*Because of this nexus, was it reasonable for the school to discipline the student?*

*Speech caused, or was reasonably likely to cause “substantial disruption” of school activities.*

*→ Permissible to restrict speech*

Instead, courts apply traditional free speech analysis, which tends to be highly protective of individual rights, but with more limited exceptions for speech that is obscene, libelous, or meant to incite harm.<sup>132</sup> More specifically, student cyber speech cases often focus on the “true threat” exception to ordinary free speech.<sup>133</sup> In *Thomas*, the court explained, “we have granted First Amendment protection to much speech of questionable worth, rather than force potential speakers to determine at their peril if words are embraced within the protected zone.”<sup>134</sup> Accordingly, the bar for a “true threat” is high. But even in the absence of this finding, a court might still determine that the emotional toll of the threat (on the algebra teacher in *J.S.*, for example) amounts to a substantial disruption of school activities, and it might thereby uphold school discipline.

Therefore, if cyber speech originates outside of school, it is not subject to a school’s disciplinary authority unless it causes or is reasonably likely to cause a substantial disruption. Thus, the second modification that lower courts have applied is an additional analytic step inserted between the con-

132. See *Thomas v. Bd. of Educ.*, 607 F.2d 1043, 1048 (2d Cir. 1979).

133. See *Mahaffey v. Aldrich*, 236 F. Supp. 2d 779, 785 (E.D. Mich. 2002).

134. *Thomas*, 607 F.2d at 1048.

tent and effect of the speech. This step pertains to the “nexus” between the creation of the speech and the extent to which it correlates with a substantial disruption of school functioning. School officials must have actual evidence of substantial on-campus disruption caused by the off-campus speech, or a reasonable fear of potential disruption to justify disciplining students for such cyber speech. Nevertheless, the nexus between off-campus cyber speech and any real or feared disruption may be established through mere on-campus access of that webpage or blog posting. The discussion of *Layshock v. Hermitage School District*, below, will illustrate the application of the modified student free speech test.

#### IV. APPLICATION OF THE STUDENT “CYBER SPEECH” FRAMEWORK IN *LAYSHOCK V. HERMITAGE SCHOOL DISTRICT*

In July 2007, the Pennsylvania district court heard *Layshock v. Hermitage School District*, a case similar to *Killion*. Again, the court upheld the student’s free speech rights. Specifically, it held that a school district could not discipline a student, Justin Layshock, for posting a parody profile of his principal, Eric Trosch, on MySpace.<sup>135</sup>

MySpace is a hosted community of users, which advertises itself as “a social networking service.”<sup>136</sup> MySpace allows users to create blogs and fill-in-the-blank profiles, which can only be removed by the user or a MySpace administrator. Justin created a profile of Trosch in which he prefaced most answers with “big.”<sup>137</sup> For example, in response to the profile question, “in the past month have you smoked?” Justin wrote “big blunt”; and to the question, “in the past month have you gone on a date?” Justin answered “big hard-on.”<sup>138</sup>

Justin’s was not the only MySpace profile of Trosch, though. During the period of time in which students were able to access Justin’s profile (approximately 5-9 days), students and teachers could also view three other parodies of Eric Trosch, all of which were more vulgar and offensive than Justin’s.<sup>139</sup> When administrators discovered the profiles, they spent several days attempting to block access to the profiles, as well as seeking out the students who created them.<sup>140</sup> After Justin admitted that he had

---

135. *Layshock v. Hermitage Sch. Dist.*, 496 F. Supp. 2d 587, 595 (W.D. Pa. 2007).

136. MySpace, <http://www.myspace.com> (last visited Dec. 10, 2007).

137. *Layshock*, 496 F. Supp. 2d at 591.

138. *Id.*

139. *Id.*

140. *Id.* at 592-93.

created a Trosch profile, the administrators suspended him for ten days, placed him in an alternative learning program at school, barred him from participating in school events or activities, and prohibited him from attending his graduation.<sup>141</sup>

The court looked to the case-law for an applicable framework for deciding the issue. It found that the most recent Supreme Court decision, *Morse*, was not controlling because it applied only to school-sanctioned and school-supervised events, as well as school-related speech.<sup>142</sup> The court also listed several recent cases decided in the Western District of Pennsylvania, including *Killion*, which found that schools could not punish students for speech originating outside of school.<sup>143</sup> It explained that student free speech cases involving off-campus expression must be treated differently than instances of restricted on-campus speech.<sup>144</sup> While courts “must defer to school administrators’ determinations regarding whether student behavior within their supervision merits punishment,”<sup>145</sup> they may not defer to schools regarding the initial “jurisdictional” question of whether the school even has such supervisory authority over a student’s off-campus expression.<sup>146</sup> “The mere fact that the internet may be accessed at school does not authorize school officials to become censors of the world-wide web.”<sup>147</sup>

Consequently, school authority to censor web content or other speech generated off-campus must derive from the extent to which that expression is connected or linked to an actual or potential on-campus disruption.<sup>148</sup> The court applied this “appropriate nexus” test in order to determine the reach of the school’s disciplinary authority.<sup>149</sup> It found that Justin only accessed the website once and showed it to others before their class started, which “in theory, might support the punishment issued by the administration,” but there was “no evidence from which a reasonable jury could conclude that this incident caused a material and substantial disruption of school operations.”<sup>150</sup>

---

141. *Id.* at 593-94.

142. *Id.* at 595.

143. *Id.*

144. *Id.* at 596.

145. *Id.* at 597.

146. *Id.* at 599.

147. *Id.* at 597.

148. *Id.* at 599.

149. *Id.*

150. *Id.* at 600-01.

While there was technically a “nexus” between Justin’s off-campus speech and his on-campus activity, there was no evidence of substantial disruption, so it was not an “appropriate nexus” to justify school punishment.<sup>151</sup> Similar to the student in *Killion*, Justin had neither created the “writing” while on school grounds nor had he used school resources. He had used his grandmother’s computer during nonschool hours.<sup>152</sup> He had accessed the profile once while at school but it had caused no classroom disruption.<sup>153</sup> Further, Justin’s was only one of four similar profiles that students could access on school computers, and the school had no evidence that his specific profile had caused any disruption.<sup>154</sup> The Court weighed the minimal disruption in this case against the more substantial disruption in *Thomas* (which the Second Circuit deemed *de minimis*) and concluded that the “school administration lacked authority to punish Justin for his off-campus creation of a Trosch profile.”<sup>155</sup>

*Layshock* demonstrates that even if school administrators can reasonably foresee that off-campus student publications might find their way onto campus, such expression is protected by the First Amendment unless it can be demonstrably tied to substantial on-campus disruption.<sup>156</sup> The *Layshock* court also touched on an important policy issue, specifically the danger that schools overreach their authority and encroach on the supervisory rights of parents and others.<sup>157</sup> “Public schools are vital institutions, but their reach is not unlimited.”<sup>158</sup> A school’s reach is limited to conduct within the scope of school activities.<sup>159</sup>

## V. CONCLUSION

While lack of guidance from the Supreme Court has created some uncertainty among the lower courts, its silence is not necessarily a bad thing. Whereas the Supreme Court has been busy creating limiting “exceptions” to protected student speech, lower courts have clung to *Tinker*; and more

---

151. *Id.*

152. *Id.* at 591.

153. *Id.* at 591-92.

154. *Id.* at 600.

155. *Id.* at 601.

156. *Id.* at 598.

157. *Id.* at 597.

158. *Id.*

159. *Id.*

often than not, they rely on *Tinker* to uphold student rights.<sup>160</sup> Upholding students' free speech rights is a real demonstration that the Constitution is alive and applies to all of us,<sup>161</sup> and it is important that "cyber speech" remains a realm in which students can criticize the governmental institutions (primarily, public schools) that control their lives. "Students in school as well as out of school are 'persons' under our Constitution"<sup>162</sup> with the undoubted right to question governmental authority. So, if we truly value First Amendment protection of speech, including inculcation of such values into the learning experience of public school students, the lower courts' more flexible interest balancing approach is preferable to the Supreme Court's bright-line "exceptions" to protected student speech.

In cyber speech cases, district courts have, for the most part, applied the core premise of *Tinker*—that students and teachers do not shed their rights at the schoolhouse gate. These courts are charged with, and best equipped for assessing the facts of each case by balancing the actual instance of student speech against valid factually supported instances of school disruption. They have used this balancing test to uphold students' off-campus cyber expression unless it is outweighed by the countervailing rights of teachers or other students. For example, school discipline will be upheld if a student's off-campus cyber speech poses a "true threat" to a teacher's safety or incites on-campus disruption, but the student's free speech rights will be upheld if he merely posts unflattering comments about his teacher or uses vulgar language on the Internet.

Even in cases that do not involve cyber speech, lower courts have been loath to apply the Supreme Court's exceptions to protected student speech. They have already interpreted both *Fraser* and *Morse* as creating only limited, narrow exceptions to *Tinker*'s interest balancing approach.<sup>163</sup> For example, the Ninth Circuit decision in *Fraser* would have protected the student's speech, because there was no evidence that it actually infringed

---

160. See, e.g., *Killion v. Franklin Reg'l Sch. Dist.*, 136 F. Supp. 2d 446, 454-55 (W.D. Pa. 2001) ("The overwhelming weight of authority has analyzed student speech (whether on or off campus) in accordance with *Tinker*.").

161. *Beussink v. Woodland R-IV Sch. Dist.*, 30 F. Supp. 2d 1182 (E.D. Mo. 1998) ("The public interest is not only served by allowing [the student's] message to be free from censure, but also by giving the students . . . this opportunity to see the protections of the United States Constitution and the Bill of Rights at work.").

162. *Tinker v. Des Moines Indep. Cmty. Sch. Dist.*, 393 U.S. 503, 511 (1969).

163. *DePinto v. Bayonne Bd. of Educ.*, 514 F. Supp. 2d, 633, 638-39 (D.N.J. 2007); *Lowry v. Watson Chapel Sch. Dist.*, 508 F. Supp. 2d 713, 722 n.3 (E.D. Ark. 2007); *Layshock v. Hermitage Sch. Dist.*, 496 F. Supp. 2d 587, 595-96 (W.D. Pa. 2007).

on other students' rights or disrupted school functioning.<sup>164</sup> Also, even though the Supreme Court did not consider his speech "political," Fraser himself viewed the speech as such. He was subjected to a harsh lesson in constitutional interpretation when, despite the factual findings of the lower court, his nondisruptive political speech was deemed unworthy of constitutional protection by the highest court in the land.

Fraser's personal experience points to the broader societal interest in preserving our constitutional rights and limiting government imposed thought control. By restricting speech, the government, especially a student's public school, exerts a dangerous chilling effect on speech. "[A] school official acts as both prosecutor and judge when he moves against student expression," because the "short duration of most sanctions" makes "the promise of judicial review . . . an empty one."<sup>165</sup> *Tinker* warned that "students may not be regarded as closed-circuit recipients of only that which the State chooses to communicate. They may not be confined to the expression of those sentiments that are officially approved."<sup>166</sup> But public school administrators have a special role as both state actors and quasi-caregivers to students, with the ability to "mete out punishment without incurring the costs of procedural safeguards."<sup>167</sup> Therefore, courts must act as gatekeepers, preserving student free speech rights. Barring substantial disruption or dangerous threats, students must be allowed to express opinions about the things that matter to them, even if the messages themselves or the ways they are communicated seem distasteful.

The chilling effect of punishing student speech merely because it is unpleasant or disagreeable threatens the foundations of democracy. "Embodied in our democracy is the firm conviction that wisdom and justice are most likely to prevail in public decisionmaking if all ideas, discoveries, and points of view are before the citizenry for its consideration."<sup>168</sup> The fact that our Constitution is designed to limit governmental censorship of citizens' opinions is what differentiates us from nondemocratic countries. A student criticizing the U.S. government while standing in Central Park is protected by our Constitution in a way that a student protesting the Chinese government while standing in Tiananmen Square is not. It follows that discouraging students from engaging in discourse and critical thinking, even if it is juvenile or silly, is antithetical to a healthy democracy.

---

164. *Fraser v. Bethel Sch. Dist.* No. 403, 755 F.2d 1356 (9th Cir.1985), *rev'd*, 478 U.S. 675 (1986).

165. *Thomas v. Bd. of Educ.*, 607 F.2d 1043, 1051-52 (2d Cir. 1979).

166. *Tinker*, 393 U.S. at 511.

167. *Thomas*, 607 F.2d at 1052.

168. *Id.* at 1047.

Therefore, courts “must remain profoundly skeptical of government claims that state action affecting expression can survive constitutional objections.”<sup>169</sup>

Clearly, there are some limits to free speech, where lives are endangered or a student truly hijacks the educational environment to disrupt other students’ ability to learn. But the danger of allowing public schools to become “enclaves of totalitarianism,”<sup>170</sup> allowed to suppress students’ off-campus speech, is far greater. The court in *Thomas* stated it well:

When school officials are authorized only to punish speech on school property, the student is free to speak his mind when the school day ends. In this manner, the community is not deprived of the salutary effects of expression, and educational authorities are free to establish an academic environment in which the teaching and learning process can proceed free of disruption. Indeed, *our willingness to grant school officials substantial autonomy within their academic domain rests in part on the confinement of that power within the metes and bounds of the school itself.*<sup>171</sup>

When the school day ends, electronic communication provides an outlet for students who may not have any other comfortable venue in which to express their thoughts and criticize their schools. It provides a “breathing space in which expression may flourish”<sup>172</sup> and provides a balance to restrictions on expression that may be imposed within the confines of her school.

By disallowing schools from unduly extending their disciplinary authority to smother students’ off-campus cyber speech, district courts are preserving not just student free speech rights, but freedom of expression in general.<sup>173</sup> These students are the voters, policy-makers, and school administrators of the future. We can only preserve their faith in democracy and ask for their buy-in as future voters and politically astute citizens by demonstrating that the Constitution works to protect their rights.

---

169. *Id.*

170. *Tinker*, 393 U.S. at 511.

171. *Thomas*, 607 F.2d at 1052 (emphasis added).

172. *Id.* at 1048.

173. *Id.* at 1047 (“At the heart of the First Amendment is the ineluctable relationship between the free flow of information and a self-governing people.”).