BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 24 NUMBER 4 FALL 2009

TABLE OF CONTENTS

ARTICLES

NOTES & COMMENTS COMPETITION WINNER

MEMO TO CONGRESS

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, 2850 Telegraph Avenue, Suite 561 #7220 Berkeley, CA 94705-7220. Periodicals Postage Rate Paid at Berkeley, CA 94704-9998, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications, 311 U.C. Berkeley 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, 2850 Telegraph Avenue, Suite 561 #7220 Berkeley, CA 94705-7220; (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 text and citations in the *Journal* conform generally to the 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 24 BERKELEY TECH. L.J. ___ (2009).

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.

HIGH TECHNOLOGY ENTREPRENEURS AND THE PATENT SYSTEM: RESULTS OF THE 2008 BERKELEY PATENT SURVEY

Stuart J.H. Graham,[†] Robert P. Merges,[††] Pam Samuelson,[†††] & Ted Sichelman[††††]

ABSTRACT

We offer description and analysis of the 2008 Berkeley Patent Survey—the first comprehensive survey of patenting and entrepreneurship in the United States—summarizing the responses of 1,332 early-stage technology companies founded since 1998. Our results show that entrepreneurs have varied and subtle reasons for using the patent system, many of which diverge from the traditional theory that patents provide an "incentive to invent." Somewhat surprisingly, startup executives report that patents generally provide relatively weak incentives to conduct innovative activities. But while a substantial share of early-stage companies hold no patents, we also find that holding patents is more widespread than previously reported, with patenting patterns and motives being highly industry, technology, and context specific. When early-stage companies patent, they are often seeking competitive advantage, and the associated goals of preventing technology copying, securing financing, and enhancing reputation. We find substantial differences between the health-related sectors (biotechnology and medical devices), in which patents are more commonly used and considered important, and the software and Internet fields, in which patents are reported to be less useful. Startups with venture funding hold more patents regardless of industry, although unlike software companies, venture-backed IT hardware firms show a patenting pattern more similar to that of health-related firms. When choosing not to patent major innovations, early-stage companies often cite to cost considerations, and report substantially higher patenting costs than the

† Assistant Professor, Georgia Institute of Technology, College of Management & Affiliated Scholar, Berkeley Center for Law and Technology, University of California, Berkeley, School of Law.

†† Wilson Sonsini Goodrich & Rosati Professor of Law and Technology, University of California, Berkeley, School of Law.

††† Richard M. Sherman Distinguished Professor of Law, University of California, Berkeley, School of Law.

†††† Assistant Professor, University of San Diego, School of Law & Affiliated Scholar, Berkeley Center for Law and Technology, University of California, Berkeley, School of Law.

prior literature has estimated. Our unique findings help inform the ongoing debate about the role and usefulness of the patent system.

TABLE OF CONTENTS

I. INTRODUCTION: WHY INVESTIGATE HOW ENTREPRENEURS USE (AND ARE AFFECTED BY) THE PATENT SYSTEM?

Entrepreneurs contribute significantly to economic growth in the U.S. and global economy.[1] They create new organizations, products, services, jobs, and opportunities for complementary economic activities.[2] Intellectual property (IP) law is an important policy lever that affects not only the opportunities for engaging in entrepreneurship, but also the success or failure of many entrepreneurial efforts.[3]

Classical economic theory holds that investments in technology development—of which entrepreneurial activities are an integral part—will be suboptimal if too little intellectual property protection exists.[4] However,

1. *See* JOSEPH A. SCHUMPETER, THE THEORY OF ECONOMIC DEVELOPMENT: AN INQUIRY INTO PROFITS, CAPITAL, CREDIT, INTEREST, AND THE BUSINESS CYCLE 74–94 (1934) (examining the role that entrepreneurship plays in the dynamism of capitalist economies); ZOLTAN ACS & DAVID AUDRETSCH, INNOVATION AND SMALL FIRMS 19–24 (1990) (discussing innovation rates among large and small firms); DAVID B. AUDRETSCH, MAX C. KEILBACH & ERIK E. LEHMANN, ENTREPRENEURSHIP AND ECONOMIC GROWTH (2006) (examining the relationship between entrepreneurship and economic growth); ANDRÉ VAN STEL, EMPIRICAL ANALYSIS OF ENTREPRENEURSHIP AND ECONOMIC GROWTH (2006) (same as above); John C. Bound et al., *Who Does R&D and Who Patents?*, *in* R&D, PATENTS AND PRODUCTIVITY 21 (Zvi Griliches ed., 1984); Martin A. Carree & A. Roy Thurik, *The Impact of Entrepreneurship on Economic Growth*, *in* HANDBOOK OF ENTREPRENEURSHIP RESEARCH 437 (Zoltan J. Acs & David B. Audretsch eds., 2003) (examining the relationship between entrepreneurship and economic growth); Samuel Kortum & Josh Lerner, *Assessing the Contribution of Venture Capital to Innovation*, 31 RAND J. ECON. 674, 682 (2000) (examining the systematic measurement of the relationship between venture capital and innovation).

2. *See supra* note 1. *See generally* Ingrid Verheul et al., *An* Eclectic Theory *of Entrepreneurship: Policies, Institutions and Culture*, *in* ENTREPRENEURSHIP: DETERMINANTS AND POLICY IN A EUROPEAN-US COMPARISON 11, 18 (2002) (describing the supply and demand for entrepreneurial activity, including culture; industry structure; individual attributes; and technological, market, and financing opportunities).

3. *See* David H. Hsu & Rosemarie Ziedonis, Patents as Quality Signals for Entrepreneurial Ventures 2 (Apr. 2007) (unpublished manuscript, on file with the Mack Ctr. for Technological Innovation) (finding that a doubling in the patent stock of venture-backed semiconductor companies leads to a 24% premium in market valuation); *see also* Andreas Panagopoulos & In-Uck Park, *Patent Protection, Takeovers, and Startup Innovation: A Dynamic Approach* 19 (Ctr. for Mkt. & Pub. Org., Working Paper No. 08/201, 2008), *available at* http://www.epip.eu/conferences/epip03/papers/Panagopoulos_ip-aug08v2-sw.pdf (providing an economic model wherein "positive but not excessive IP protection" leads to increased acquisitions of startup firms by incumbents, thereby increasing overall innovation); Ted Sichelman, *Commercializing Patents*, 62 STAN. L. REV. 341 (2010) (exploring the relationship between patent protection and the successful commercialization of invention).

4. *See generally* WILLIAM D. NORDHAUS, INVENTION, GROWTH, AND WELFARE: A THEORETICAL TREATMENT OF TECHNOLOGICAL CHANGE (1969) (discussing the standard economic theory of IP); Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for*

because entrepreneurial activity may be redirected or halted by intellectual property rights claimed by others, an equally serious impediment to investment for entrepreneurs may arise if IP protection is too strong or uncertain.[5]

Although a considerable body of previous work has explored the relationship between IP rights and innovation, far less scholarship has focused on the more particular relationship between IP rights and entrepreneurship.[6] The basic economic principle underlying IP rights is that the process of developing innovative products and practices is an expensive, time-consuming, labor-intensive, and risky endeavor.[7] Once these innovations exist, however, they can be cheap and easy to copy.[8] IP rights protect innovators from copying by free riders and allow them to recoup the investment incurred during the creation, development, and commercialization processes, either directly by manufacturing and distributing products and services embodying the innovation, or indirectly through licensing to other firms that incorporate the innovation in their products and services.[9] Although this basic economic principle applies to all companies, because early-stage firms tend to lack the kinds of complementary assets (such as well-defined marketing channels, manufacturing capabilities, and access to cheap credit) that ease entry into the market, they are arguably even more sensitive to IP rights than their more mature counterparts.[10] Nevertheless, research and policy analysis has not

---

Innovation, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609 (William D. Nordhaus ed., 1962) (same as above).

5. *See* Stuart J.H. Graham & Ted Sichelman, *Why Do Start-Ups Patent?*, 23 BERKELEY TECH. L.J. 1063, 1080–81 (2008) (exploring how incumbents may use patent portfolios to restrain startups from entering markets); Colleen V. Chien, *Of Trolls, Davids, Goliaths, and Kings: Narratives and Evidence in the Litigation of High-Tech Patents*, 87 N.C. L. REV. 1571, 1573–74 (2009).

6. *See, e.g.*, NORDHAUS, *supra* note 4; Arrow, *supra* note 4; Lee G. Branstetter, Raymond Fisman & C. Fritz Foley, *Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U.S. Firm-Level Panel Data*, 121 Q.J. ECON. 321 (2006); Nancy Gallini & Suzanne Scotchmer, *Intellectual Property: When Is It the Best Incentive System?*, 2 INNOVATION POL'Y & ECON. 51 (2002); Elhanan Helpman, *Innovation, Imitation, and Intellectual Property Rights*, 61 ECONOMETRICA 1247 (1993).

7. Arrow, *supra* note 4, at 619.

8. *Id.* at 614

9. Concerning these latter transactions in the markets for technology, see ASHISH ARORA, ANDREA FOSFURI & ALFONSO GAMBARDELLA, MARKETS FOR TECHNOLOGY: THE ECONOMICS OF INNOVATION AND CORPORATE STRATEGY 171–96 (2004).

10. *See* Scott Shane, *Technological Opportunities and New Firm Creation*, 47 MGMT. SCI. 205, 209 (2001) ("Although established firms might also be more likely to commercialize broad patents, they are disproportionately important to independent entrepreneurs who lack complementary assets."). *See generally* David J. Teece, *Profiting from Technological Innovation*:

adequately addressed how particular IP laws differentially affect entrepreneurial firms relative to more established ones.[11]

In response to this noticeable gap in knowledge, the Berkeley Center for Law and Technology (BCLT) conducted a wide-scale survey in 2008 of high-technology startup firms in the United States to determine how these companies use and are affected by the patent system. Part of our aim in conducting the survey was to identify those aspects of the patent system that substantially encourage or hinder entrepreneurial activity, particularly in high-growth technology industries such as the Internet, computer software and hardware, medical device, and biotechnology fields.

We were also concerned with the paucity of data regarding the potential effects of pending patent reform measures on entrepreneurs and on the entrepreneurial process more generally. Patent reform is currently the subject of a vigorous national debate.[12] Recent studies by the Federal Trade Commission[13] and the National Academy of Sciences[14] made specific policy recommendations that inspired this debate. This led to the drafting and discussion of various legislative proposals in Congress, starting in 2005 and continuing through the current session.[15] There has also been a renewed interest in patent law in the Supreme Court, with the Court deciding several high-profile cases over the last several years.[16] Recently, the Court granted certiorari in the highly controversial case of *Bilski v. Doll*, which will address

Implications for Integration, Collaboration, Licensing and Public Policy, 15 RES. POL'Y 285 (1986) (discussing the benefits of "complementary assets").

11. *See* Ted Sichelman & Stuart Graham, *Patenting by Entrepreneurs: An Empirical Study*, 16 MICH. TELECOMM. & TECH. L. REV. (forthcoming 2010) (reviewing most of the existing literature on patenting and entrepreneurship).

12. *See, e.g.*, Dan L. Burk & Mark Lemley, *Don't Tailor Make Patent Act: The Key to Accommodating Competing Interests Lies in the Courts*, 31 NAT'L L.J. 18 (2009).

13. FEDERAL TRADE COMMISSION, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY 4–17 (2003), *available at* http://www.ftc.gov/opa/2003/10/cpreport.htm.

14. COMM. ON INTELLECTUAL PROP. RIGHTS IN THE KNOWLEDGE-BASED ECON., NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., A PATENT SYSTEM FOR THE 21ST CENTURY 81–130 (Stephen A. Merrill, Richard C. Levin & Mark B. Myers eds., 2004), *available at* http://www.nap.edu/catalog.php?record_id=10976.

15. *See generally* Kevin R. Davidson, *Retooling Patents: Current Problems, Proposed Solutions, and Economic Implications for Patent Reform*, 8 HOUS. BUS. & TAX L.J. 425 (2008) (describing various congressional bills); Carl E. Gulbrandsen et al., *Patent Reform Should Not Leave Innovation Behind*, 8 J. MARSHALL REV. INTELL. PROP. L. 328 (2009) (same as above).

16. *See, e.g.*, MedImmune, Inc., v. Genentech, Inc., 549 U.S. 118 (2007); KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398 (2007); eBay Inc. v. MercExchange, L.L.C., 547 U.S. 388 (2006).

the scope of patentable subject matter for business methods.[17] *Bilski* may also have implications for the patentability of software, a subject matter, along with business methods, of critical concern to many technology startups. Also of note, the U.S. Patent and Trademark Office (USPTO) has recently considered dramatic changes in the way patent applications are handled.[18] The opinions of large corporations and the organized patent bar have been expressed very clearly in the judicial, legislative, and Patent Office reform debates,[19] but input on how the proposed changes to patent law and policy will specifically affect entrepreneurship has been lacking.[20]

In the detailed results presented below, we do not offer simple answers to the difficult unanswered questions about the patent-entrepreneurship relationship. As one important example illustrates: although we find that holding patents is more widespread among technology startups than previously believed, company executives report that patents provide relatively weak incentives for core activities in the innovation process, such as invention and commercialization. These results raise the question: why would startups incur the substantial costs of patenting if they find patents are not

17. *In re* Bilski, 545 F.3d 943 (Fed. Cir. 2008), *cert. granted sub nom*. Bilski v. Doll, 129 S. Ct. 2735 (2009).

18. *See, e.g.*, Tafas v. Doll, 559 F.3d 1345 (Fed. Cir.), *reh'g en banc granted*, 328 Fed. App'x 658 (2009) (finding that the proposed continuation rules were inconsistent with the Patent Act); James W. Beard, *Weeds in the Docket: Patent Continuation Reforms and Their Impact on Patent Applications in the Biotechnology Industry*, 90 J. PAT. & TRADEMARK OFF. SOC'Y 423 (2008) (describing the PTO's proposed changes to the continuation rules). The USPTO recently withdrew its proposed changes, essentially mooting the case. *See* Susan Decker, *U.S. Patent Office Drops Rules That Led to Glaxo Suit*, BLOOMBERG, Oct. 8, 2009, http://www.bloomberg.com/apps/news?pid=newsarchive&sid=aIKiD9D4dOgc.

19. *See, e.g.*, Brief Amicus Curiae of the Federal Circuit Bar Ass'n in Support of Respondents, Teleflex, Inc., KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398 (2007) (No. 04-1350), 2006 WL 2983161; Amicus Brief for Pharmaceutical Research & Manufacturers of America, MedImmune, Inc., v. Genentech, Inc., 459 U.S. 118 (2007) (No. 05-608), 2006 WL 2091231; Brief for Amicus Curiae Computer & Communications Industry Ass'n in Support of Petitioners, eBay, Inc. v. MercExchange, L.L.C., 547 U.S. 388 (2006) (No. 05-130), 2005 WL 2381066; Kevin Bogardus, *Stakeholders Inch Toward Final Deal on Patent Reform, but Rifts Remain*, THE HILL, Feb. 13, 2008, *available at* http://thehill.com/business-a-lobbying/3467-stakeholders-inch-toward-final-deal-on-patent-reform-but-rifts-remain.

20. *See, e.g.*, Joe Mullin, *Small Companies Suffering in Surge of Infringement Suits*, THE RECORDER, June 4, 2009, at 3 ("When the Senate began debating patent reform in March, there was plenty of testimony from patent experts, but nothing from small entrepreneurs."). Although independent inventors have been active in the patent reform debates, not all independent inventors are entrepreneurs. *See, e.g.*, Rick Merritt, *Inventors' Group to Address Patent Reform*, EE TIMES, June 4, 2009, *available at* http://www.eetimes.com/news/latest/showArticle.jhtml?articleID=217701966. As such, a unique "entrepreneurship voice" has likely been missing from the reform debates.

offering particularly strong incentives to engage in invention and commercialization?

Our response to this question is nuanced and multi-faceted. Specifically, a large share of early-stage companies, especially in the software industry, avoid the patent system altogether. At the same time, we find that startups hold more patents and patent applications than prior commentators have reported.[21] These differences appear to be primarily attributable to our survey method, which captured many patents and applications—including those acquired from founders and from other sources—that previous studies had missed.

A dominant theme in our findings is that the usefulness of patents to technology entrepreneurs is driven by industry characteristics. So while we find that patents help many startups compete in the market with their technology, this role tends to be much more pronounced among biotechnology and hardware companies (including both medical hardware, such as surgical devices, and IT hardware, such as computers and semiconductors). Conversely, patents are much less important as a means by which most software firms—the majority of which hold no patents—capture competitive advantage from their innovations.[22]

Our results also shed light upon startups' motivations for filing for patents and for choosing not to patent major innovations. When electing to patent, startup executives tend to be most influenced by a desire to prevent copying and, to a lesser extent, by reputational and financial motives, including successful exit (such as being acquired or going public). In an important showing, we demonstrate that patenting may play a previously underappreciated role in helping startups to secure investment from various sources of entrepreneurial capital, including not only angel and venture investors, but also "friends and family" and commercial banks. Also notable is our finding that the costs of prosecuting and enforcing patents are a substantial barrier to technology entrepreneurs attempting to access the patent system. But the explanation for startups choosing not to patent is also context-specific: biotechnology company executives are much more likely to cite concerns about information disclosure than those in other industries.

Our results also show subtlety in the ways that startups contend with patents in their operating environment and their executives' attitudes toward the patent system. While respondents told us that checking the patent literature is reasonably common while innovating, startups at times do so

21. *See infra* Section III.B.2.

22. *See id.*

only after launching their products or when considering patenting themselves. Licensing of others' patents is also not infrequent, and while startups are generally receiving information or know-how in the transaction, they are also commonly trying to settle a controversy—and in some cases they take patent licenses solely to avoid a lawsuit. On the whole, technology entrepreneurs tell us that the patent system is neither working particularly poorly nor well for their companies and industries.

Because many of these reasons for—and attitudes toward—patenting have little to do with the classical incentives and free rider stories, they are exemplary of many of the results we present in this Article. We find varied and subtle explanations for why technology entrepreneurs use the patent system: from the reasons they decide (or not) to file for patents on their innovations, to whether and when they examine the patent literature, to their views about the patent system as a whole. With this context in mind, Part I describes the specific research questions that motivated the survey, how the samples were selected, and some key profile characteristics of our respondents. Part II discusses the patent holding characteristics of technology startups, and the reportedly weak incentives patents provide to startups to engage in innovative activity. Part III offers our first set of findings that help explain why startups file for patents when they appear to offer relatively weak incentives to innovate. Part IV assesses the varied roles played by patents in helping the startup to compete, especially compared with other means available to the firm for capturing value from innovation. Part V provides another set of key findings—those uncovering the motivations that entrepreneurs report for filing patents on their innovations. Part VI complements that discussion by exploring the reasons why technology entrepreneurs forgo patenting. Part VII completes the analysis by examining how technology entrepreneurs respond when they face patents held by others in their competitive environment. Part VIII explores our key attitudinal question in the survey, and reports on the general opinion among high-technology executives that the patent system is neither working particularly well nor particularly poorly for their companies and industries. Part IX offers some concluding observations about the scope of the study, the implications of our findings, and continuing our research in this area.

II. APPROACH TO SURVEYING HIGH TECHNOLOGY ENTREPRENEURS ABOUT THEIR USE OF IP

We began our project with a set of research questions addressing a variety of topics concerning patenting and technology entrepreneurship, including the following:[23]

- Do startups apply for patents to protect their innovations and is this a successful strategy?
  - Do they apply for patents for other reasons, e.g., to attract investors or for cross-licensing to obtain the right to use another company's patents?
- What role do actual or potential patent rights play in decisions to invest in startups?
  - In what circumstances is strong patent protection a prerequisite to investment?
  - In what circumstances are patent rights less important than other factors?
- How do investors and entrepreneurs assess the scope and value of their own and other firms' patent rights in the course of deciding which business opportunities to pursue or to fund?
- Once entrepreneurs have initial funding, are patent rights more important to the entrepreneur than to more established companies, which can rely on manufacturing, distribution and marketing capacities, brands and other reputation qualities, and their existing customer base to protect their market position?
- What steps do startups take to avoid infringing other parties' patent rights? Are they able to acquire licenses to such rights when necessary? Is inadvertent infringement a problem?
- How often do startups receive allegations of infringement and what do they do in response?
  - Are entrepreneurs affected by "patent trolls" (that is, owners of patents who pursue patent litigation as a business model)?

23. We were aided in formulating our research design by insights and helpful comments during discussions with Hank Barry, John Barton, James Bessen, Tom Ciotti, Wes Cohen, John Duffy, Rebecca Eisenberg, Brad Feld, Richard Gilbert, Michael Goldberg, Josh Green, Bob Gunderson, Bronwyn Hall, Mitchell Kapor, Peter Menell, James Pooley, Walter Powell, Arti Rai, AnnaLee Saxenian, Carl Shapiro, Robert Strom, Lee Van Pelt, and David Yoffie, whom we thank.

- How would patent reform proposals affect entrepreneurial companies?

Given the breadth of these research questions, we chose a research design that would allow us to provide meaningful empirical data on these issues. While some useful data is available in archival sources, much of the most useful information about how well or poorly the patent system is working for entrepreneurs resides with those who are starting and managing startup and early-stage companies. We therefore designed a survey questionnaire targeted to these issues and administered it to a large sample of technology entrepreneurs in companies across the United States.

In order to transform our research questions into a survey, we undertook extensive research and conducted numerous interviews with experts in the field and with active entrepreneurs. During this process we examined the theoretical and empirical literature in law, economics, management, and other social sciences to develop a series of hypotheses.[24] We also drew heavily from prior innovation surveys.[25] Furthermore, our research team held discussions with scholars, university technology-transfer officers, independent inventors, startup founders and executives of early-stage companies, Silicon Valley lawyers, managers of venture capital (VC) firms, and angel investors in order to better understand the entrepreneurial environment and to craft questions

24. Some of the ideas covered by our questions were drawn from a number of sources. *See* Janice J. Jackson, *The Usefulness of the General Social Surveys Database in Entrepreneurship and Small Business Research*, *in* 4 ADVANCES IN ENTREPRENEURSHIP, FIRM EMERGENCE, AND GROWTH: DATABASES FOR THE STUDY OF ENTREPRENEURSHIP 393 (Jerome A. Katz ed., 2000); Jerome A. Katz, *The Logic and Opportunities of Secondary Analysis in Entrepreneurship Research*, *in* 4 ADVANCES IN ENTREPRENEURSHIP, FIRM EMERGENCE, AND GROWTH: DATABASES FOR THE STUDY OF ENTREPRENEURSHIP 5 (Jerome A. Katz ed., 2000); Richard C. Levin et al., *Appropriating the Returns from Industrial Research and Development*, 18 BROOKINGS PAPERS ON ECON. ACTIVITY (SPECIAL ISSUE) 783 (1987); Edwin Mansfield, *Patents and Innovation: An Empirical Study*, 32 MGMT. SCI. 173 (1986); Edwin Mansfield, Mark Schwartz & Samuel Wagner, *Imitation Costs and Patents: An Empirical Study*, 91 ECON. J. 907 (1981); Jerry G. Thursby & Marie C. Thursby, *Who Is Selling the Ivory Tower? Sources of Growth in University Licensing*, 48 MGMT. SCI. 90 (2002); Wesley M. Cohen, Richard R. Nelson & John P. Walsh, *Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (Or Not)* (Nat'l Bureau of Econ. Research, Working Paper No. 7552, 2000), *available at* http://ssrn.com/abstract=214952.

25. Multiple surveys were examined. *See* AM. INTELLECTUAL PROP. LAW ASS'N, AIPLA 2005 ECONOMIC SURVEY (2005); CARNEGIE MELLON UNIV. & ALFRED P. SLOAN FOUND., RESEARCH AND DEVELOPMENT IN THE UNITED STATES (1994); GA. INST. OF TECH., INVENTORS AND THEIR INVENTIONS: UNDERSTANDING THE INNOVATION PROCESS (2007), *available at* http://www.prism.gatech.edu/~jwalsh6/inventors/Inventor Questionnaire.pdf; NAT'L SCI. FOUND., THE MANUFACTURERS' INNOVATION SURVEY (1993); PURDUE UNIV., PURDUE LICENSING SURVEY (1998).

that would shed light upon topics that were not well understood, or for which there were conflicting explanations.[26]

A. OUR FOCUS IS ON TECHNOLOGY STARTUPS

1. Targeting "Entrepreneurial Companies" and Their Top Executives

The economist Daniel Spulber has critiqued the relative underrepresentation of the entrepreneur in neoclassical economic theory.[27] In his view, entrepreneurship does not begin until the individual forms a startup firm, after which the individual and the enterprise are inexorably intertwined.[28] Spulber writes:

> Entrepreneurs play a central role in the economy because they are the prime movers—the makers of firms. . . .
>
> . . . [Spulber's definition of t]he general theory of the firm places the entrepreneur at the center of microeconomic analysis. The entrepreneur engages in transactions that are needed to establish firms. In turn, firms create and operate markets and organizations.[29]

According to this theory, the period of entrepreneurship does not end until a true separation of ownership and control occurs.[30]

We approached the problem of surveying entrepreneurship in a manner consistent with that suggested by Spulber. While there are many pathways to better understanding the roles played by patenting in entrepreneurship, we

---

26. We had fruitful discussions concerning our questionnaire design with Hank Barry (Hummer Winblad Venture Partners), James Bessen (Boston University School of Law), Tom Ciotti (Morrison & Foerster LLP), Wes Cohen (Fuqua School of Business, Duke University), John Duffy (The George Washington University Law School), Brad Feld (Foundry Group), Rich Gilbert (UC Berkeley Department of Economics), Bob Glushko (UC Berkeley School of Information), Christoph Grimpe (Centre for European Economic Research), Dominique Guellec (Organisation for Economic Co-operation and Development), Bob Gunderson (Gunderson Dettmer Stough Villeneuve Franklin & Hachigian, LLP), Bronwyn Hall (UC Berkeley Department of Economics), Dietmar Harhoff (Ludwig-Maximilians University School of Business), Bob Lee (UC Berkeley Survey Research Center), Mark Myers (The Wharton School, University of Pennsylvania), Sean O'Connor (University of Washington School of Law), Lee Van Pelt (Van Pelt, Yi & James, LLP), and John Walsh (Georgia Institute of Technology School of Public Policy), whom we thank for their insights.

27. *See* DANIEL F. SPULBER, THE THEORY OF THE FIRM: MICROECONOMICS WITH ENDOGENOUS ENTREPRENEURS, FIRMS, MARKETS, AND ORGANIZATIONS 156–57 (2009).

28. *Id.* at 151.

29. *Id.*

30. *Id.* at 152. The separation of ownership and control has been a cornerstone of the economic theory of the firm for decades. *See also* ADOLF A. BERLE, JR. & GARDINER C. MEANS, THE MODERN CORPORATION AND PRIVATE PROPERTY 69 (1933).

elected to survey "entrepreneurial companies"—defined by us as companies founded in the United States during the last ten years.[31] Given our research questions stated above, we were particularly keen to uncover the realities of patenting and entrepreneurship, rather than the intentions or attitudes of entrepreneurs. Noting the almost complete lack of empirical data on what entrepreneurs actually *do* when faced with decisions about intellectual property, we focused not on the intentions of would-be entrepreneurs, but rather on the choices that entrepreneurs in firms actually made while innovating.

Such a research design fosters the ability to provide meaningful insights about the role that IP rights play in a wide class of entrepreneurial activities. For instance, focusing on the company allowed us to inquire into the entity's age, employment, and share of scientists and engineers, and then relate these aspects to the company's patenting choices. Moreover, this research design permitted us to ask about the sources and success in garnering external funding, and relate these responses to the usefulness and utility of seeking patents. Critically, unlike the few European studies of small firm patenting,[32] we avoided focusing solely on existing patent holders and, instead surveyed the broad class of entrepreneurial companies in order to offer insights into how the patent system is working for patentees and non-patentees alike.

Entrepreneurs generally become high-level executives of the firms they found and tend to be generalists who handle numerous aspects of their business.[33] Aware that many firms have more than one founder,[34] and driven

31. Entrepreneurs can run the panoply from "idea havers" to inventors to company founders. *See generally* David G. Blanchflower & Andrew J. Oswald, *What Makes an Entrepreneur?*, 16 J. LAB. ECON. 26 (1998) (outlining characteristics of individual entrepreneurs); James W. Carland et al., *Who Is an Entrepreneur? Is a Question Worth Asking*, *in* 2 ENTREPRENEURSHIP: CRITICAL PERSPECTIVES ON BUSINESS AND MANAGEMENT 178 (Norris F. Krueger ed., 2002) (same as above); William B. Gartner, *"Who Is an Entrepreneur?" Is the Wrong Question*, 12 AM. J. SMALL BUS. 11 (1988) (same as above).

32. *See* Sichelman & Graham, *supra* note 11, at 25–31 (reviewing these studies).

33. *See* Edward P. Lazear, *Balanced Skills and Entrepreneurship*, 94 AM. ECON. REV. (PAPERS & PROC.) 208, 210 (2004) (noting that entrepreneurs are generalists who put together teams of people and assemble resources and capital, and to be effective, they must have a general set of skills).

34. *See* Chan Chaiyochlarb, *If You Are Going to Launch a Startup, How Many Friends Would You Need?*, TESTING TESTING 1,2,3, Feb. 23, 2007, http://blogs.msdn.com/testing123/archive/2007/02/23/if-you-are-going-to-build-a-startup-how-many-friends-should-you-start-up-with.aspx (electing a non-scientific sample and arriving at a figure of 2.09 founders for successful firms); Brad Feld, *How Many Founders Does a Startup Need?*, FELD THOUGHTS, Feb. 25, 2007, http://www.feld.com/wp/archives/2007/02/how-many-founders-does-a-startup-need.html (observing that startups tend to have two or three founders).

by our interest in the innovation functions of the firm,[35] we targeted our survey questionnaire to chief executive officers (CEOs), presidents, and chief technology officers (CTOs). Over three-quarters of our respondents identified themselves as being one of these three officers.[36]

2. Selecting High Technology Sectors

Because we were principally interested in knowing how patenting affects the technology entrepreneurship process, we limited our field of study to certain high technology sectors. Given the extensive writing, opinion, and theory about the differences in innovation and patenting characteristics between the life sciences and information technologies firms, we focused primarily on companies in the biotechnology and software industries. As such, much of our analysis below compares and contrasts the results of our survey in these two sectors. At the same time, we understood that other high-technology industries are important drivers of dynamism and growth. Accordingly, we also surveyed companies in the medical device and IT hardware sectors (the latter defined as semiconductor, communications, and computer hardware), and when meaningful and relevant, we report briefly here on their responses. While we are mindful that patenting may play an important role for startups in other industries,[37] much of the patenting and venture-funding activity among small firms takes place in our chosen high technology sectors.[38]

35. We chose not to specifically target lawyers in the firm, on the theory that lawyers would tend to be comparatively more risk averse about divulging company information to outsiders.

36. This figure is actually an undercount, since some respondents who were presidents instead called themselves "Owner," and other officers neglected to identify themselves as "Chief Technology Officer" but instead as "VP of Engineering," "Chief Scientist," or the like. As such, our yield in these three categories was closer to 85%. Among the "other" executives responding, they serve in many different functions, including those from operations (about 15% of all others), finance (about 10% of all others), and in-house lawyers (about 7% of the others), as well as marketing/sales and development (each about 5% of all others).

37. While the vast majority of our respondents fell into these four industry classifications, about 12% came from other sectors. Approximately 7% fell into medical technologies that did not include either "biotechnology" or "medical devices," and another 5% fell completely outside our technology definitions. For more information, see Stuart Graham & Ted Sichelman, *Patents and Innovating Startups: Analysis of the 2008 Berkeley Patent Survey* (Nov. 24, 2009) (unpublished manuscript, on file with Social Science Research Network), *available at* http://ssrn.com/abstract=1512726.

38. *See* Robert E. Hall & Susan E. Woodward, *The Incentives to Start New Companies: Evidence from Venture Capital* 3 (Nat'l Bureau of Econ. Research, Working Paper No. 13056, 2007), *available at* http://imio.haas.berkeley.edu/WilliamsonSeminar/hall021408.pdf.

3. Choosing Companies to Survey

Because our research questions deal with how the patent system broadly affects technology entrepreneurship, we adapted our research design to two different types of entrepreneurial companies. First, we were interested in studying how patenting relates to the "average" entrepreneurial company. Second, we were keen to understand how the highest "quality" companies—namely, those most likely to generate innovations, succeed, and grow—were using the patent system. We therefore chose two groups to survey: those representing the population of firms, and those that had been successful in securing VC funding. Unfortunately, there is no single association to which all high-technology entrepreneurial firms belong.

After examining the available data sources, we selected Dun & Bradstreet (D&B) as a proximate window into the overall population of companies in the United States. D&B is a leading business credit-reporting and information source in the United States and holds over 140 million business records worldwide.[39] D&B conducts entrance interviews to assign a credit rating to the companies in its database during which it determines technology class and founding date. We used D&B's data on technology class[40] and founding date to construct a sample of 10,500 D&B-listed companies founded in our target industries after December 31, 1997.[41]

However, we also understand that venture-backed firms, though an important segment, make up only a small portion of the overall population of

39. The company reports that "77 percent of D&B's U.S. active file contains businesses with 10 or fewer employees." Dun & Bradstreet, Facts & Figures, http://www.dnb.com/us/about/db_database/dnbstatistics.html (last visited May 24, 2009). Based on a rough comparison with aggregate U.S. Census data, we estimated that D&B contains about 60% of all companies founded in our target industries. *See* U.S. Census Bureau, Statistics of U.S. Businesses, http://www.census.gov/csd/susb/susbdyn.htm (last visited Sept. 23, 2009).

40. Using the entrance interviews, D&B assigns companies to Standard Industry Classifications (SICs), which we relied upon to select which companies fell into each industry sector. SIC numbers are not sufficiently differentiated to identify "biotechnology" companies—in the SIC codes, these companies are lumped together with other commercial researchers in SIC code 8731 "Commercial Physical and Biological Research" companies. We therefore relied upon D&B's more fine-grained North American Industry Classification System (NAICS) coding to identify "biotechnology research" companies. The corresponding NAICS code is 541711. *See* U.S. Census Bureau, North American Industry Classification System, http://www.census.gov/eos/www/naics/ (last visited Sept. 23, 2009).

41. Our list included all D&B's biotechnology companies (642) assigned to NAICS 541711, all medical device companies (1,048) assigned to SIC 3841, and a random sample of its listed software companies (8,810) assigned to SIC codes 7371, 7372, 7373, and 7379.

companies.[42] Prior research shows that this small share of firms is disproportionately responsible for innovative output in the economy,[43] and that venture capital funding is related to patenting activity.[44] Because of our interest in understanding the role of patenting in startup investment, we chose to focus on this important class of companies by turning to the Thomson's *VentureXpert* data, which covers a substantial share of venture-backed companies in the United States.[45] Using Thomson's rich data on company characteristics, we selected 5,600 companies primarily in our target sectors founded in 1998 or later that received venture or similar institutional

42. *See* Kortum & Lerner, *supra* note 1, at 674 (noting that from 1983–1992, venture-backed firms' R&D spending accounted for 3% of the U.S. total).

43. *Id.* at 675 (noting that while venture-backed firms account for just 3% of all R&D spending, they account for 8% of the innovative output and a disproportionate amount of patenting).

44. *See* David B. Audretsch, Werner Bönte & Prashanth Mahagaonkar, *Financial Signalling by Innovative Nascent Entrepreneurs* (Ctr. for Econ. Policy Research, Discussion Paper No. DP7165, 2009), *available at* http://ssrn.com/abstract=1345692; Iain M. Cockburn & Megan MacGarvie, *Patents, Thickets, and the Financing of Early-Stage Firms: Evidence from the Software Industry* (Nat'l Bureau of Econ. Research, Working Paper No. W13644, 2007), *available at* http://ssrn.com/abstract=1037168; Carolin Haeussler, Dietmar Harhoff & Elisabeth Müller, *To Be Financed or Not . . . – The Role of Patents for Venture Capital Financing* 2 (Munich Sch. of Mgmt., Discussion Paper No. 2009-02, 2009), *available at* http://www.ip.mpg.de/shared/data/pdf/haeussler_et_al_vcpat_jan2009lmu_wp_reihe.pdf; Paul H. Jensen, Elizabeth Webster & Hielke Buddelmeyer, *Innovation, Technological Conditions and New Firm Survival* 5–6 (Melbourne Inst., Working Paper No. 26/06, 2006), *available at* http://ssrn.com/abstract=946827.

45. *VentureXpert* draws its data from portfolio companies funded by over 1,000 private equity partnerships, of which 700 are venture funds, and holds information on the funding transactions of over 4,350 VC firms derived from industry surveys as well as quarterly and annual fund reports. VentureXpert, About Us, http://vx.thomsonib.com/NASApp/VxComponent/VXMain.jsp (last visited Oct. 1, 2009). *But cf.* Steven N. Kaplan, Berk A. Sensoy & Per Strömberg, How Well Do Venture Capital Databases Reflect Actual Investments? 1 (Sept. 2002) (unpublished manuscript, on file with the Chicago Graduate School of Business), *available at* http://ssrn.com/abstract=939073 (suggesting that *VentureXpert* tends to be biased toward California companies, and that most of the data come from the investors, not from the companies themselves).

funding.[46] Combining our D&B and *VentureXpert* samples, our final list of target firms contained over 15,000 unique entities.[47]

B. PROFILING OUR RESPONDENT COMPANIES

In trying to profile the respondents in our survey, we are met with a difficult task. Since we purposefully set about to target entrepreneurial companies across a variety of characteristics (such as industry and age), there is no meaningful "average" respondent company. However, it is important to disclose the numerical breakdown of the responding companies' characteristics. We therefore offer descriptive statistics on the characteristics of the companies that answered our questionnaire, followed by statistics on the response rates to our survey.

Our median respondent (at the 50th percentile per category) is a self-described "startup" company founded in April 2002. It has nine employees, half of whom are scientists or engineers, and has neither had an initial public offering (IPO) nor been acquired.[48] The company's 2007 revenues were $300,000, and its founders had prior experience running another company. Geographically, the company's offices are located somewhere west of the Mississippi River. In terms of funding its operations, the median respondent from our D&B sample has received funding from "friends and family," at

46. Selection was made primarily on technology attributes. Thomson uses a proprietary industry classification that includes "Information Technology," "Medical/Health/Life Science," and "Non-High Technology." As of May 2008, of the over 11,000 U.S. companies founded after 1997 in their database, about 65% were classed "Information Technology" and about 20% were assigned to "Medical/Health/Life Science." All companies with at least one email address for a company executive were included in our sample (3,359 companies), supplemented with a random sample of all remaining companies meeting our technology definitions.

47. We also administered our questionnaire to a group of companies in the "cleantech" (i.e., environmental technologies) market in conjunction with GreenTech Media, an online cleantech industry clearinghouse. Because we have a strong reason to believe that this sample was not representative of the population of cleantech startup firms, we do not report these results here. However, we plan to make these effectively anecdotal reports available.

48. In this regard, over 10% our respondents had been acquired or had undergone an IPO. So while our sample may have excluded some firms that effectively dissolved upon a successful exit, presumably D&B and *VentureXpert* data did not exclude many of these firms. Furthermore, our sample also included many firms that had *unsuccessfully* dissolved, although we did not pursue extra efforts to locate former executives and founders of those firms. While in many, if not most of the circumstances, these executives were not reachable, there are responses in our data from companies that were winding down when the surveys were completed. Nevertheless, our respondent set generally excludes non-surviving firms. However, our bias testing of respondent versus non-respondent firm-level characteristics, described further *infra*, included both survivors and non-survivors in the latter category.

least one "angel" investor, and a commercial bank, but not from venture capital firms, investment banks, or other companies.

Excluding our pre-test responses, 1,332 unique companies responded to our survey. Overall, this figure produces an 8.7% response rate. This uncorrected response rate does not account for companies that could not be reached, either because the contact information we had was inaccurate, or because the firm had gone out of business or fundamentally reorganized. We revised this base response rate by first correcting for returned mail (evidence of inaccurate physical address),[49] and second by reference to the results of random sample telephone calls we conducted during the fall of 2008.[50]

For the D&B sample, we achieved a 7.0% response rate, and after correcting for bad addresses that figure becomes 8.4%. Correcting further for telephone failures, the rate rises to 10.6%. In the *VentureXpert* sample, we achieved a 12.4% response rate among companies for which we had emails.[51] After accounting for mailing and telephone failures, the corrected response rate is 17.9%.

In addition to higher response rates from *VentureXpert* companies, responses from medical firms were substantially more likely than from IT companies.[52] In the D&B sample, the (telephone and mail-corrected) response rate for biotechnology and medical device firms is 23.7% and 13.4%, respectively. For software and Internet companies, that same figure is 8.9%. Among the *VentureXpert* sample, the (fully corrected) response rates for medical (biotechnology and medical device) and IT (hardware and software) companies are 24.2% and 15.6%, respectively.[53]

Overall, we find that our respondents are not statistically different from the non-respondents on key company characteristics. Within industries, we

49. The U.S. Postal Service (USPS) returned approximately 17% of our mailings with "return to sender," "unable to forward," and "no longer at address" stamps. It is likely that this figure is low since the USPS returns less than 100% of non-deliverable mail, particularly when that mail, like ours, is not a first-class mailing.

50. We found on average that 23% of the companies could not be reached at the telephone numbers provided by our data sources, though this figure was substantially higher in some sectors (e.g., 41% for venture-backed biotechnology firms).

51. We had an email address for at least one officer in 68% of the *VentureXpert* companies. Our testing shows that there are generally no significant differences in the way that the emailed and non-emailed (mail only) *VentureXpert* companies answered our survey. *See* Graham & Sichelman, *supra* note 37.

52. In this regard, because of the reported importance of patenting to biotechnology companies, we took extra efforts to contact this group, telephoning every non-respondent, typically multiple times.

53. Based on our telephone calls, we found that approximately 23% of IT companies and 35% of biotechnology companies were not reachable at the telephone number provided by our data sources.

find generally no statistically significant differences in the age, sales volume, or employee counts between respondents and non-respondents.[54] We also recognized the possibility that more active patentees may have chosen to disproportionately respond to our "patent" survey, so we matched our respondent and non-respondent companies to the USPTO patent record. Again, we find no statistically significant differences between the two groups in terms of the number of patents held by the companies, nor the number of applications filed since 2001.[55] We have built statistical models based on firm characteristics to predict responses from non-respondents, and have also compared responses from our original pool of respondents to those garnered from a random telephone sample of initial non-respondents. While we do see a western bias in the companies responding to our survey, our testing generally shows that our respondents are not dissimilar in other important respects from our non-respondents.[56] In sum, by and large, we can detect no statistical differences in the answers from the companies we were able to reach and the companies that declined to answer our survey.[57]

One of the main findings of our survey is that venture-backed companies are significantly different in the way they view the innovation, technology competition, and patenting processes. As such, we often report their responses separately in this Article. When these differences are not important, we combine the results.

---

54. D&B software respondents are approximately one-half year younger than non-respondents when we measure on the year (not date) of founding (significant at the 95% confidence level). Responding D&B medical device companies have more employees than non-respondents (38 versus 14, a result which is weakly significant at the 90% confidence level). When we report confidence levels herein, we describe differences as either at 90%, 95%, or 99% level, but an actual confidence level may be higher than the reported value.

55. We can only see published applications after 2001 in the United States. We also tested for the number of patents (and applications) per year of age of the company, and again found no significant differences. *See* Graham & Sichelman, *supra* note 37.

56. *See id.* The one exception is in the medical device sectors, in which respondents do appear to be different from the non-respondents in some respects. For instance, D&B non-respondents tend to have significantly more employees at the mean than the D&B respondents in this sector, although this difference is largely driven by a few relatively large outliers (i.e., larger firms, represented in this sector did not choose to respond to our survey). This finding may account for the large (yet insignificant) differences in non-respondents' patents and applications we found when matching to the USPTO data. Venture-backed companies in medical devices also show differences, with the non-respondents being significantly more likely to be founded earlier than respondents, which may again explain differences we see (albeit non-significant ones) in the larger number of patents granted to non-respondents in our USPTO name matching. *See id.* As such, we tend to downplay the reporting on these respondents in this Article.

57. For interested readers, we describe this bias testing in a separate article. *See id.*

III. PATENT (AND APPLICATION) HOLDING IN TECHNOLOGY STARTUPS

A. PATENT HOLDING AMONG STARTUPS IS WIDESPREAD BUT NOT UBIQUITOUS

1. Startups' Patents Come from Several Different Sources

Because we asked respondents to report the numbers and origins of the patents and applications they held, our survey results offer a uniquely accurate window into the patenting behavior of early-stage technology companies. In the past, scholars have tried to match granted patents listed in the USPTO database to company names by using data in the assignee field.[58] This method tends to produce an undercount because it misses patents that were assigned to the company from the founders post-grant or otherwise acquired from outside the firm.[59] While scholars have generally not used the USPTO reassignment data to supplement the original assignments, even in this regard the USPTO records on patents reassigned to different entities after grant are notoriously incomplete.[60] Further compounding the difficulty

58. *See, e.g.*, Bronwyn H. Hall, Adam Jaffe & Manuel Trajtenberg, *Market Value and Patent Citations*, 36 RAND J. ECON. 16, 20 (2005).

59. Although published patent applications have recently become available in the United States, they too cannot be relied upon to give accurate results. First, they have only been published since 2001. Second, they are only published after eighteen months, which means that the most recent applications cannot be counted. Third, there is a class of patent applications that do not require publication. Applicants can disclose to the Patent Office that they do not intend to seek patent protection in any other jurisdiction with an eighteen month publication rule and opt out of the requirement. *See* 35 U.S.C. § 122(b)(2)(B) (2006). Estimates vary, but it is believed that between 5–15% of applications take advantage of this statute. *See* Stuart J. Graham, Continuation, Complementarity, and Capturing Value: Three Studies Exploring Firms' Complementary Uses of Appropriability Mechanisms in Technological Innovation 36 (Spring 2004) (unpublished Ph.D. dissertation, UC Berkeley) (on file with the UC Berkeley Library). Additionally, small firms are among the most likely to elect non-publication, since they are comparatively less likely to market their products outside the United States. *See* Mark D. Janis, *Patent Abolitionism*, 17 BERKELEY TECH. L.J. 899, 919 (2002).

60. Although the Patent Office offers a recording system, it is merely optional, and the recording of assignments appears to be routinely ignored. *See* 35 U.S.C. § 261 (2006). In essence, the law establishes a registration system with a notice statute protecting subsequent purchasers for value. *See id.*; U.S. PATENT & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE § 301 (8th ed., rev. 7, 2008). The registration system is organized in a fashion analogous to a "land deed" registry, in which the person or entity that purchases later without notice of an earlier transfer prevails over the earlier transferee, if it did not record within a grace period. *See* 35 U.S.C. § 261 (2006); U.S. PATENT & TRADEMARK OFFICE, MANUAL OF PATENT EXAMINING PROCEDURE § 301 (8th ed., rev. 7, 2008). Because the market for buying and selling patents is presumably subject to extensive due diligence and other forms of contracting that prevent fraudulent assignments of patents,

of knowing "what patents companies have" is the problem of name-matching, which stems from the multiple forms of company names, corporate name changes, and typographical errors in the patent records themselves.[61] Moreover, arriving at accurate counts may be particularly difficult for early-stage companies since—as our survey results show—it is common for startups' patents to originate with founders prior to the company's founding date. Such patents may first issue to founders (as individuals, with no initial corporate assignee) and only later be assigned to the company. Because startups are resource constrained (in money and time), it may be less likely that these companies update their assignment records at the Patent Office.[62] Effectively tracking these patents would require knowing the names of the founders and matching them to the USPTO patent inventor records. And with common names, that second task can be a herculean labor.[63]

Our survey results are not prone to these difficulties, and thus offer superior information about the patent holdings of early-stage technology companies. Instead of relying on matching, we asked our respondents to report to us the number of U.S. patents or filed applications held by the company from three unique sources: (1) those coming from founders that were applied for prior to the company founding date; (2) those acquired by the company from sources other than its founders; and (3) those filed by the company itself after its founding date.

2. Startups Hold Numerous Patents, but Explanations Are Nuanced

Because studies relying on the USPTO database had reported average patent holdings for *venture-backed* startup companies in the range of 0.5–6.0 patents,[64] we were surprised to find that, on average, the companies in the population of high-tech firms (as proxied by D&B) hold well over four (4.7)

---

arguably there is not as great a need to pay attorneys' and filing fees to record reassignments as with real property.

61. For instance, the assignee "Minnesota Mining and Manufacturing Company" also appears as assignee "3-M" in U.S. Patent No. 4,000,444 (filed Mar. 12, 1973).

62. *See* Dennis Crouch, *Another University Patent Ownership Dispute: Stanford Loses Rights Based on Researcher's Side Agreement*, PATENTLY-O, Oct. 1, 2009, http://www.patentlyo.com/patent/2009/10/another-university-patent-ownership-dispute-stanford-loses-rights-based-on-researchers-side-agreement.html ("Patent ownership regularly transfers without the new owner recording the assignment with the USPTO.").

63. *See* Manuel Trajtenberg et al., *The "Names Game": Harnessing Inventors' Patent Data for Economic Research* (Nat'l Bureau of Econ. Research, Working Paper No. W12479, 2006), *available at* http://ssrn.com/abstract=926058 (discussing the difficulty of the inventor name-matching problem).

64. *See* Ronald J. Mann & Thomas W. Sager, *Patents, Venture Capital, and Software Start-ups*, 36 RES. POL'Y 193, 197 (2007).

patents and applications (Table 1). Among the venture-backed firms in our survey—a more comparable sample to previous studies—the average firm holds just under nineteen patents and applications (18.7). These figures are somewhat misleading, though: among the D&B respondents, over six in ten companies (61%) hold no patents at all. Moreover, these average statistics are influenced by a few patent holders with very large portfolios. For instance, one of the respondents in our D&B sample reports holding 260 patents and applications from all sources. Similarly, one venture-backed company holds 570 total patents and applications.

We observed earlier that there are important differences between the very select venture-backed startups and companies drawn from the larger population of companies (most of this latter set having no venture investment)—and this observation also applies to patent ownership. For instance, among those startups drawn from the D&B sample (approximately 85% of which have no venture backing), the median company (the firm at the 50th percentile) holds no patents or applications. Venture-backed companies are substantially different from the D&B respondents: the median venture-backed firm holds six patents or applications. Among venture-backed firms, a comparatively small 18% hold no patents or applications, about one-third the share exhibited by the D&B companies.

These findings suggest that the holding of patents by technology startups is more widespread than previously believed, especially among venture-backed companies, but that holding patents is by no means ubiquitous among entrepreneurial firms. Substantial numbers of early-stage technology companies appear to be opting out of the patent system altogether, and these firms are not merely clustered among the younger companies. In fact, the likelihood of not holding any patents is virtually the same among the youngest and oldest companies in our study. For example, among the older D&B companies in our sample (those founded prior to 2003), 64% of firms report holding no patents.[65] Therefore, the likelihood of holding (or not holding) patents by technology startups does not appear to be driven by age effects, but instead by the company's business model, strategy, technology, or other factors, such as the cost of patenting and subsequent enforcement.[66]

65. We note that there is not a hidden "technology" effect to these findings: the share of companies that are identified as biotechnology, medical devices, or software are virtually identical among the older and younger companies.

66. Among those firms that chose to hold patents, however, there is a positive influence of age on the number of patents held. *See* Graham & Sichelman, *supra* note 37.

Table 1: Patents and Applications Held by Startup Companies[67]

Source / Industry	All respondents	Biotechnology	Medical Devices	Software/Internet	IT Hardware‡
Population of companies (D&B)					
Companies holding patents/applications (share)	39%	75%	76%	24%	—
Average # patents/applications held (all companies)	4.7	9.7	15.0	1.7	—
Average # filed by company (patent holders only)	8.1	8.5	13.0	5.0	—
Average # from founders (patent holders only)	1.9	2.0	3.0	1.2	—
Average # acquired (patent holders only)	2.1	2.4	3.7	0.9	—
Venture-backed companies					
Companies holding patents/applications (share)	82%	97%	94%	67%	91%
Average # patents/applications held (all companies)	18.7	34.6	25.2	5.9	27.4
Average # filed by company (patent holders only)	15.8	22.9	16.1	7.1	23.6
Average # from founders (patent holders only)	2.5	3.8	3.8	0.7	3.1
Average # acquired (patent holders only)	4.2	9.0	6.5	0.7	3.5

‡ Available only for *VentureXpert* listed companies.

3. Technology Entrepreneurs Suggest the Patenting Story Is Complex

To better understand the underlying rationale of our respondents' answers, we conducted several hour-long, follow-up interviews by telephone in the early months of 2009.[68] The comments of one such executive, call him Neil, are illustrative of what is often a tension for the entrepreneur in deciding whether to seek patent protection. Neil is the inventor of an innovative biometrics information technology and CEO of a startup he founded in 2003 to commercialize the technology. While his company has

67. Results are based upon two questions from the survey. (1) "Does your company own U.S. patents or has it applied for U.S. patents (including any patents or applications acquired through any purchase or transfer)?" and (2) "How many, if any, U.S. patents or pending applications does your company have from the following sources: Patents or pending applications applied for by Founders prior to the company's founding date? Patents or pending applications purchased, transferred, or assigned to company? Patents or pending applications filed by your company since founding?"

68. Our interviews were conducted on a convenience sample of our respondents. We began with a random sample of respondents who had provided email addresses to us (approximately 80% of the total), and disproportionately contacted respondents in Northern California. We did, however, speak with all respondents who were willing to speak with us.

filed one patent application that has not yet been granted, he has generally tried to avoid patents even though his investors, which include VC firms, are interested in him having them. He stated:

> Our business is a consumer services business; it was never my intention to be a monopoly and to protect [this technology] . . . . Instead, our strategy is to grow as the number one supplier of [our new product] while simultaneously allowing other, smaller companies to spread [the underlying technology] around the nation. In doing so, [the technology] will have a huge consumer base and our company will be viewed as *the* prominent player, much like "Kleenex" is in the tissue industry.[69]

Neil's comments help expose several themes that run through our survey responses. First, patenting is common among our respondents, but by no means ubiquitous. Second, patenting for the entrepreneurs choosing to hold patents is often motivated by reasons that have not commonly been well-understood or studied in the literature. And third, the incidence and usefulness of patents to technology entrepreneurs is very much determined by the industry and technology in which the company is operating.

B. IN PATENTING, NOT ALL STARTUPS ARE CREATED EQUAL

1. Industry Influences the Incidence of Startups Holding Patents

Similar to the patenting differences associated with the presence of venture funding, we find profound disparities in the likelihood, number, and original source of patents by the technology focus and industry of the company. Among the D&B sample, biotechnology and medical device companies are much more likely to hold patents and applications than are software and Internet firms (Table 1). In fact, three out of four of these medical companies report holding patents and applications compared with only one in four among software startups.

There are also substantial differences across industries in the number of patents held on average, with the total patents of medical and life science companies once again substantially greater than those of the software and Internet firms in the D&B sample. Medical device companies report holding fifteen patents and applications on average, compared with just under ten for biotechnology research companies. These figures are significantly higher than for software companies, which hold on average less than two patents and applications. Much of this difference is driven by the relatively high likelihood of software firms holding no patents. If we compare the average

69. Interview with anonymous executive, March 2009 (emphasis in original).

count of patents and applications held by firms that have chosen to hold patents, D&B software and Internet firms compare more closely with D&B biotechnology research companies (7.1 patents to 12.9 patents, respectively). Additionally, large patent-portfolio companies influence the result: among those D&B companies that hold patents, the median software startup (at the 50th percentile) holds two patents and applications and the median biotechnology firm possesses six. In the IT sector, hardware companies hold significantly more patents than their software counterparts: venture-backed IT hardware firms hold on average more than twenty-seven patents and applications, about five times more than similarly-funded software and Internet companies.

While there are substantial inter-industry differences at the level of the average D&B company, focusing instead on those companies choosing to hold patents demonstrates smaller or insignificant differences as to the *origin* of the patents held by the firm. Across the D&B patent holders, about two patents and applications come into the firm from its founders, and while this number falls to nearly one for the (fewer) software and Internet companies holding patents, the difference is not statistically meaningful. For every ten employees at a patent-holding D&B company, the firm files seven patents on average, and it acquires two patents or applications from sources other than its founders. These figures change little based on the technology of the company, and the differences are not statistically meaningful. In sum, for startup companies that hold patents, the likely origin of those patents remains virtually the same regardless of the technology focus of the firm. What is significantly different, though, is the likelihood of holding *any* patents—software and Internet companies in the population (as proxied by D&B) are much less likely to hold patents than similar companies involved in the health-related technologies.

2. Venture-Backed Companies Are More Likely to Hold Patents

These inter-industry differences also persist among the venture-backed firms, where the incidence of holding patents is much higher and the origins of those patents are more varied. Table 1 shows that virtually all (97%) companies in the biotechnology and medical device sectors hold patents—and while holding patents is less likely for venture-backed IT firms (hardware and software alike), the rates are still relatively high (about 90% and 70%,

respectively). Venture-backed firms are much more likely to hold patents, regardless of technology focus.[70]

In order to better understand the motivations for venture-backed companies to patent, we interviewed several partners at VC firms. One such partner holds an engineering degree, invests primarily in biotechnology companies, and has extensive experience in the technology, business, and investment environments in which his portfolio companies operate. He stated:

> When you go into life sciences—and in reality, with any [bio-related technology] that you're creating or acquiring—if it doesn't have a reasonably strong patent, and if you don't have the capability to expand the patent estate covering your technology and products, you are going to have complicating issues. [As a young company], you need to secure patents, and with the broadest claims and specifications that you can get.[71]

These comments support two clear messages that spring from our responses. First, early-stage biotechnology companies are much more likely to use, and to see utility in using, the patent system. Second, venture investors are interested in patents, and venture-capital backed companies are much more likely to hold and file for patents. Whether this second observation is primarily driven by investors demanding a more active patent "footprint" of the companies they fund or by companies that VCs fund simply being more likely to supply patents to investors, we are unable to say with certainty. Our evidence points, however, to a relationship that runs both ways. Firms that seek venture-funding appear to be patenting more actively prior to the funding event (and for the purpose of securing funding), and venture-capital investors appear much less willing to fund companies that hold no patents.

In this sense, the results of our survey differ substantially from several prior studies of startup patenting. In particular, we show that among venture-backed companies, patent-holding is more widespread and—in conjunction with additional data we present below[72]—more important to securing venture investment than previously reported. The best prior evidence on the topic comes from Ronald Mann and Tom Sager, who matched venture-backed companies to the Delphion patent database.[73] Their article reports that

70. The finding that venture-backed firms are substantially more likely to hold patents is consistent with an earlier study finding that venture-backing has a positive influence on patenting. *See* Kortum & Lerner, *supra* note 1.

71. Interview with anonymous venture capital firm partner (May 2009).

72. *See infra* Parts IV, V.

73. Mann & Sager, *supra* note 64, at 196.

venture-backed software firms hold on average just under three patents, while venture-backed biotechnology companies possess just under 5.5 patents.[74] Moreover, they state: "[O]nly 2% of software firms had more than 4 patents, and less than 1% of software firms had more than 10 patents; 19% of biotechnology firms had more than 4 patents and 6% had more than 10 patents."[75]

Our responses from startup company executives show that patents are more widely held, and held in greater numbers, than Mann and Sager estimate. Our respondent venture-backed software and Internet firms hold on average just under six patents and applications,[76] double the estimate of Mann and Sager, while venture-backed biotechnology companies hold just under thirty-five patents, more than six times what Mann and Sager estimate. If we restrict ourselves only to patents and applications filed by these firms (thus more closely approximating Mann and Sager's patent-matching-and-counting method), our venture-backed software respondents report filing 4.9 patents on average, while venture-backed biotechnology companies file 22.2 at the mean—still substantially higher than the figures estimated by Mann and Sager.

In contrast to Mann and Sager's suggestion that patent holding is reasonably uncommon,[77] our respondents tell us that 63% of venture-backed software and Internet startups hold more than four patents and applications, while more than 47% of these firms reported holding more than ten. Extending this analysis to venture-backed biotechnology startups, we find that 86% of these companies report having more than four patents and applications while 60% hold more than ten. These discrepancies with the figures cited in Mann and Sager may be partly accounted for by the earlier time period of their study[78] and the fact that not all applications result in an

74. *Id.* at 197.

75. *Id.* at 198.

76. We also find that venture-backed IT hardware firms hold on average more than twenty-seven patents and applications, about five times more than similarly-funded software and Internet companies.

77. Mann & Sager, *supra* note 64, at 197–98.

78. Mann and Sager examined companies that received their first rounds of funding in 1996–1998, measuring patent grant rates as of the end of 2004. As such, the higher patenting numbers we see among firms in our data could reflect an increasing patenting trend over time in startups—though such a trend is arguably unlikely to account for a substantial share of the large differences reported here. Additionally, although it may appear that Mann and Sager's time window is shorter than ours, it is not. In particular, because most startups do not receive venture capital financing until more than one year after founding, Mann and Sager's effective time horizon appears to be about seven to ten years. An analysis by the authors shows that firms listed in *VentureXpert* that have issued patents (matched to the USPTO data) have, on average, initial VC funding arriving about 2.5 years after company

issued patent.[79] This latter possible explanation may be undercut, however, insofar as the grant rate of startups' patent applications is higher than usual because (a) the quality of the invention may be higher than average[80] and (b) the patent applicants themselves may care more about winning a granted patent, and thus expend more time, effort, and money in the application process.[81]

Another noteworthy finding of our study is that founders bring substantial numbers of patents and applications with them into their startups. Another possible source for our differences with Mann and Sager's estimates is their failure to account for founder-added patents that go unrecorded at the USPTO.[82] When a D&B-listed biotechnology startup holds patents, its founders tend to bring on average two patents and applications into the firm, a figure that nearly doubles to 3.8 when the firm has venture backing.[83] Venture-backed medical device and IT/hardware firms that hold patents tend to acquire more than three patents and applications on average from their founders. However, for patent-holding software companies with

founding. Although our study focused on firms founded after 1997, because many of those firms were not founded until much later (the median founding date for all firms was 2002), the effective windows appear roughly equal, if not shorter in our study.

79. There is conflicting evidence about the USPTO's application grant rate. *See* Robert A. Clarke, *U.S. Continuity Law and Its Impact on the Comparative Patenting Rates of the US, Japan and the European Patent Office*, 85 J. PAT. & TRADEMARK OFF. SOC'Y 335 (2003) (responding to an earlier study by Quillen and Webster); Ron D. Katznelson, *Bad Science in Search of "Bad" Patents*, 17 FED. CIR. B.J. 1 (2008) (discussing problems involved in grant rate calculation); Cecil D. Quillen, Ogden H. Webster & Richard Eichmann, *Continuing Patent Applications and Performance of the U.S. Patent and Trademark Office - Extended*, 12 FED. CIR. B.J. 35 (2002) (evaluating prior date to revisit determination of the USPTO grant rate); Mark A. Lemley & Bhaven Sampat, *Is the Patent Office a Rubber Stamp?*, 58 EMORY L.J. 181, 184–89 (2008) (pointing to data controversies in calculating patent grant rates).

80. *See* Kortum & Lerner, *supra* note 1, at 690 (finding that venture-backed firms' patents have significantly more forward citations than the population of patents). *But cf.* Jean O. Lanjouw & Mark Schankerman, *Protecting Intellectual Property Rights: Are Small Firms Handicapped?*, 47 J. L. & ECON. 45 (2004) (finding that patent portfolio size is positively related to forward citations).

81. *See infra* notes 188–89 and accompanying text (finding that startups expend twice or more on out-of-pocket costs on patenting than previously-reported averages). *See generally* Stuart J.H. Graham & Dietmar Harhoff, *Separating Patent Wheat from Chaff: Would the U.S. Benefit from Adopting a Patent Post-Grant Review?* (Oct. 14, 2009) (unpublished manuscript, on file with Social Science Research Network), *available at* http://papers.ssrn.com/abstract_id=1489579 (suggesting higher grant rates among triadic patents demonstrate that extra effort pays off in winning patents).

82. *See* Mann & Sager, *supra* note 64, at 196–97.

83. These differences are significant at the 95% confidence interval.

venture funding, founders are likely to bring fewer than one patent or application with them into the startup.[84]

After founding, venture-backed biotechnology and medical device companies are also more likely than software and Internet firms to file patent applications and to acquire patents (or applications) from sources other than their founders. For these biotechnology and medical device firms that hold patents or applications, they report that on average eight originate from acquisitions (other than from founders), and that they directly file twenty applications. These figures are substantially higher than the just over 0.5 patents or applications acquired on average by venture-backed software and Internet firms, and the seven directly filed by these startups. Interestingly, IT hardware firms are the outlier: these companies are more like the health-related firms in terms of their number of founder-originating patents and patent filings, but are also similar to their software counterparts in that they tend to acquire comparatively few patents from sources other than founders.

C. PATENTS MAY OFFER ONLY MIXED TO WEAK INCENTIVES TO ENGAGE IN INNOVATION

Given that the patent monopoly is most commonly justified on the ground of providing incentives to innovate, we were surprised to find that, in general, the technology startup executives responding to our survey report that patents offer relatively mixed to weak incentives to engage in innovation. In this context, we refer to the term "innovation" in its Schumpeterian sense—the series of steps taken from idea to invention to development to commercialization.[85] To uncover patents' incentive value, our questionnaire asked all respondents how strong or weak an incentive (1=not at all, 2=weak, 3=moderate, and 4=strong incentive) patents served for undertaking four innovation-related activities: (a) inventing new products, processes, or services; (b) conducting initial research and development; (c) creating internal tools or processes to build or implement final products, processes, or services; and (d) undertaking the risks and costs of making, selling, and marketing a commercial product. In general, the executives we surveyed reported that patents serve as only slight to moderate incentives for each of these stages in the innovation process.

---

84. It is noteworthy that among software firms, the average number of founder-originating patents is lower for venture-backed companies (0.7 patents) than for those in the D&B population sample (1.2 patents), although this difference is not statistically significant at conventional levels.

85. *See* SCHUMPETER, *supra* note 1, at 66 (contending that innovation consists of novel goods, production methods, markets, production inputs, and forms of organization).

We recognized during the design stage of our survey instrument that the meaning of the term "incentive" could be interpreted in several ways by our respondents. While we cannot dismiss the possibility that our respondents may not have understood the term incentive in the way that neoclassical theory generally presents it,[86] our preliminary testing and interviews suggested to us that the business-savvy technology entrepreneurs we surveyed do interpret that term in the way that theory presents it. Moreover, there is ample evidence that the patents-as-incentives rationale is a well-known concept in the mainstream business literature and media.[87] Of course, while the respondents may have understood our questions, they may not fully comprehend the role patenting actually plays in their innovation process, which is often subtle.[88] So while we are reasonably confident that our responses capture at least a proximate window into the incentive role that

86. *See* JEAN TIROLE, THE THEORY OF INDUSTRIAL ORGANIZATION 390 (1988) (inducing innovation through an award of a temporary monopoly ultimately benefits the public good).

87. *See, e.g.*, Robert J. Barro, *Drugs Via Canada? The Side Effects Could Hurt*, BUS. WK., Aug. 30, 2004, at 34 ("Pursuing these policies would be a mistake because the current setup strikes a reasonable balance between incentives for invention and the prices of existing drugs."); Hiawatha Bray, *Ethernet Pioneer: Patent System Is Flawed, but Still the Best We've Got*, BOSTON GLOBE, May 13, 2007, at D3 ("When you grant private property, you create an incentive to invest and protect and make the best use of that property. A patent, which is a form of private property, is itself a form of technology for encouraging innovation." (quoting Bob Metcalfe, partner, Polaris Venture Partners)); Michael Fitzgerald, *A Patent Is Worth Having, Right? Well, Maybe Not*, N.Y. TIMES, July 15, 2007, at BU3 ("Patents are supposed to give inventors an incentive to create things that spur economic growth."); Gloria Lau, *U.S. Policy on Prescription Drugs Keeps Raising Eyebrows*, INVESTOR'S BUS. DAILY, July 25, 2005, at A7 ("There's got to be a worldwide push to honor intellectual property. Otherwise there's no incentive to do drug research."); Gary L. Reback, *Patently Absurd*, FORBES, July 21, 2002, at 44 ("But as an incentive to innovate, a patent holder gets a free pass from the rigors and challenges of competition. The right amount of such incentive may well spur invention.").

88. For instance, it may be that people think of "an incentive" as somehow the proximate cause or reason for conducting R&D and other innovation-related activities, in which case a respondent may not have connected the term "incentive" with more "secondary" incentive effects that patents may play in spurring the activity. In other words, people may well think: "I would have conducted this innovation anyway, even without a patent." Indeed, because one rarely knows the outcome of an R&D project, and never knows the complete shape of the prior art, very few inventions would be created if the certainty of patenting was required to begin R&D. What people may not realize, however, is that the "secondary" effects of patents—such as attracting capital and enabling arm's-length transactions—may in the aggregate contribute enough of a "plus factor" to make certain projects viable, even if people do not think of patents in those terms. Thus, respondents may not have appreciated the idea of a "marginal incentive effect," thinking of incentives as more of an "all or nothing" concept.

patents play for technology entrepreneurs, we nevertheless recognize that further investigation of our findings is warranted.

1. The Reported Relatively Weak Incentive Value of Patents

Among the D&B companies, respondents told us that on average, patents offer just above a "slight incentive" to engage in invention, R&D, and commercialization, and between "slight" and "no incentive at all" to create internal tools and processes. While venture-backed startup executives rate the incentive value more highly than do those at D&B companies, in no category are patents reported to provide even a "moderate" incentive for any of the four entrepreneurial activities about which we queried.

An interview we conducted with the sole proprietor of a medical device company assisted us in understanding some of the low-powered incentives that patents offer. This physician-turned-entrepreneur—whom we will call Jeremy—has chosen largely to opt out of the patent system. As is the case with many innovations in medical devices, Jeremy was a "user-innovator," seeing a practical problem in his medical practice that needed a practical response.[89] He founded his company to offer a product to meet that need, and quickly secured a patent. This lone patent, however, has not allowed him to compete effectively, and he perceives very little incentive to seeking patent protection today:

> Fifteen years ago, patents were probably very useful and offered a lot of protection. But not today. In fact, today they are not very valuable at all, and, even if I were to get a patent on my [updated technology], odds are that I would still find a copy for sale on the side of the road in China.[90]

Jeremy's comments highlight two aspects associated with the incentive value of patents expressed in our survey responses. First, respondents tell us in general that the incentive value provided by patents is typically low for a range of innovative activities. Second, the pattern of reporting relatively weak incentives generally holds not just for those companies that do not patent, but also among those companies that do.

To verify that those companies holding patents were also reporting relatively low incentives, we divided our respondents into two categories: companies that told us they held at least one patent or application and companies holding none. We show that, as we expected, companies

89. *See generally* ERIC VON HIPPEL, DEMOCRATIZING INNOVATION (2005) (describing "user innovation" as technology supplied by technology users who experience a need and innovate to fill that need).

90. Interview with anonymous company founder (Feb. 2009).

expending resources to acquire patents rate their incentive value higher in all categories than do those companies holding none. However, the incentive value that these active patent holders ascribe to patents still does not, for any of the four innovative activities, reach on average even the "moderate incentive" level. And for those companies that report holding zero patents or applications, the incentive value of patents is ranked by respondents on average between "slight incentive" and "no incentive at all" for each of the four innovation activities.

An important caveat to these findings, however, is a marked inter-industry difference in the incentive values that patents offer for innovation. For example, biotechnology companies report that patents generally provide "moderate" incentives in the innovation process, whereas software firms report that they generally provide at best "slight" incentives. This finding is consistent with inter-industry differences in the incidence and numbers of companies holding patents discussed above.[91] While these incentive results are also consistent with anecdotes about software firms,[92] we were somewhat surprised to find that, in the biotechnology sector, companies did not report a stronger connection between the incentives offered by patents and innovation. Interestingly, these incentive results do not substantially change when limiting the sample only to those firms that are actually holding patents.

Our findings on the incentive value of patents are noteworthy given the legal and theoretical bases for the "patents as incentives" view. The U.S. Constitution, the foundation for the intellectual property laws in the United States, provides for patents primarily upon the incentive view.[93] And at least since Nobel Laureate Kenneth Arrow's 1962 work on the value of intellectual property in spurring research and development, economists have examined the value that society gains from offering patents (in the form of limited monopolies) to "idea havers" as an incentive to convert those ideas into inventions.[94] Noting that patent rights generally live well beyond the creative spark of invention,[95] other scholars have suggested that patents serve as incentives throughout the innovation process, across a range of

91. *See supra* Section III.A.

92. *See* Ronald J. Mann, *Do Patents Facilitate Financing in the Software Industry?*, 83 TEX. L. REV. 961, 1001 (2005).

93. *See* U.S. CONST. art. I, § 8, cl. 8.

94. *See* Mansfield et al., *supra* note 24, at 915; Roberto Mazzoleni & Richard R. Nelson, *The Benefits and Costs of Strong Patent Protection: A Contribution to the Current Debate*, 27 RES. POL'Y 273, 274 (1998); Brian D. Wright, *The Economics of Invention Incentives: Patents, Prizes, and Research Contracts*, 73 AM. ECON. REV. 691, 696 (1983).

95. *See, e.g.*, SUZZANNE SCOTCHMER, INNOVATION AND INCENTIVES 138–43 (2004).

entrepreneurial activities.[96] So, if as our findings suggest, patents provide relatively weak incentives to undertake the risks of innovation, what purposes do they serve?

2. The Multiple Meanings of Startup Patenting

The overall picture of startup patenting suggested by these two chief findings is surprising. On the one hand, startup companies in all high technology sectors are patenting much more widely, and in greater numbers than we had reason to believe from the prior literature. However, when we ask top executives at these early-stage firms whether the patents that they are seeking (and for which they are devoting scarce resources) offer incentives to create, develop, and commercialize the technology that is at the core of the venture, they answer that, in general, patents are not serving that purpose particularly well.

Our questionnaire design enabled us to shed light on this important conundrum. We find, and detail below, that patents are playing significant roles in technology entrepreneurship, even if the executives of startups assess their incentive value to be relatively weak. Some of these functions do concern patents' traditional role in diminishing competition from third parties in the technological marketplace and the related concern of securing profits from innovation—especially for biotechnology, medical device, and hardware firms (but notably less so for software and Internet companies). At the same time, patents appear to be supporting other activities crucial to *technology* startups: securing the necessary investment to develop and grow; increasing the odds and quality of a liquidity event, such as an acquisition or IPO; and serving strategic roles in negotiation and defending against patent infringement suits.

What clearly springs from our data is a recognition that patents serve as important institutional support for activities that are important to the success of technology entrepreneurs. But as some theorists have recently suggested,[97] for many of these companies, patents do not provide the sorts of incentives to innovate that we long expected, though our findings indicate that they often serve other important roles in the entrepreneurship process. We now turn to a fuller examination of these roles.

96. *See generally* Sichelman, *supra* note 3 (examining incentives to commercialize significant patented inventions).

97. *See* Clarisa Long, *Patent Signals*, 69 U. CHI. L. REV. 625 (2002); Mann, *supra* note 92.

IV. FIRST FINDING: PATENTS HELP STARTUPS IN TECHNOLOGICAL COMPETITION

A. PATENTS CAN SERVE TO PROMOTE STARTUPS' COMPETITIVE ADVANTAGE

A major finding of our survey, and a partial explanation for the widespread use of patents by technology entrepreneurs, concerns the function that patents serve in helping the startup compete in the marketplace with its innovative technology. Theory has long suggested that formal IP rights are critical in protecting the lead-time or first-mover advantages that fast innovators possess.[98] Moreover, earlier surveys of managers at large U.S. companies indicated that IP rights are an important means of gaining a competitive edge but found mixed results when specifically considering patents.[99]

Our results demonstrate that patenting plays a substantial role for many high-technology startups in securing competitive advantage from their technology innovations, but that this finding is also context-specific. In asking our sample companies about how meaningful patents were in the quest for profits and success, we attempted to disaggregate some of the answers from previous surveys, and also focused on startups. Specifically, previous studies tended to examine how mainly large and publicly traded U.S. companies profit from their innovation in fairly broad terms,[100] generally failing to distinguish among important legal mechanisms, such as copyrights and trademarks.[101] While making these distinctions, our questionnaire also reproduced several elements from these previous surveys, such as "patenting" and "secrecy," for the sake of consistency and comparison.[102] In the end, our questionnaire asked the respondent to indicate how important (or unimportant) the following seven items were to the company in securing

98. *See* Teece, *supra* note 10, at 286–87 (suggesting that both technological complexity and the strength of IP rights are critical to protect the innovator's advantage).

99. *See* Cohen et al., *supra* note 24; Levin et al., *supra* note 24.

100. *See supra* note 99.

101. For example, the study by Levin and others included both patenting and secrecy, but not copyrights and trademarks as appropriability means. *See* Levin et al., *supra* note 24, at 794. The study by Cohen included patenting, secrecy, and "other legal" means. Cohen et al., *supra* note 24, at 5–6.

102. The use in prior surveys of the term "secrecy" instead of "trade secret" presented us with a conundrum: we believed that the term "secrecy" could be interpreted by respondents to mean both more, and less, than "trade secret." At the same time, we saw value in being as consistent as possible with prior surveys. Preliminary testing with respondents convinced us that the subjective understanding of the terms "secrecy" and "trade secret" did not differ substantially among respondents. As such, we decided to use the term "secrecy."

competitive advantage from its technology innovations: first-mover advantage over competitors; secrecy; patents; copyrights; trademarks; difficulty of reverse engineering; and other production, implementation, or marketing capabilities. Hereafter, we call these seven methods "appropriability strategies."[103]

Our results show that these early-stage firms use multiple appropriability strategies. Among our aggregated respondents,[104] first-mover advantage is clearly ranked the most important: in fact, it is the only appropriability strategy ranked between "moderately important" and "very important" on average by all companies.[105] Three methods are grouped together in the next ranked position—secrecy, complementary assets, and patenting—which respondents rate on average between "slightly important" and "moderately important," although closer to the latter.[106] Following these are the remaining three appropriability strategies—difficulty of reverse engineering, trademarks, and copyright—which fall between "slightly important" and "moderately important."[107]

Such an aggregation of our responses[108] offers a basic descriptive picture into our results: first-mover advantage tends to dominate the other appropriability strategies. This view of our data also suggests that other methods—including patents—are generally rated as having some importance. In fact, the difference between the average scores assigned to the other six methods is relatively small (even though some of these differences are statistically significant). However, this aggregation of our responses hides a highly nuanced story. In fact, when we examine companies' responses by their differing characteristics, such as their technological focus or the type of funding they have secured, the importance of these appropriability strategies, and pointedly patenting, tends to shift radically.

---

103. In so doing, we follow the nomenclature of the Cohen study. *See* Cohen et al., *supra* note 24, at 5.

104. 1,236 companies responded to this set of questions.

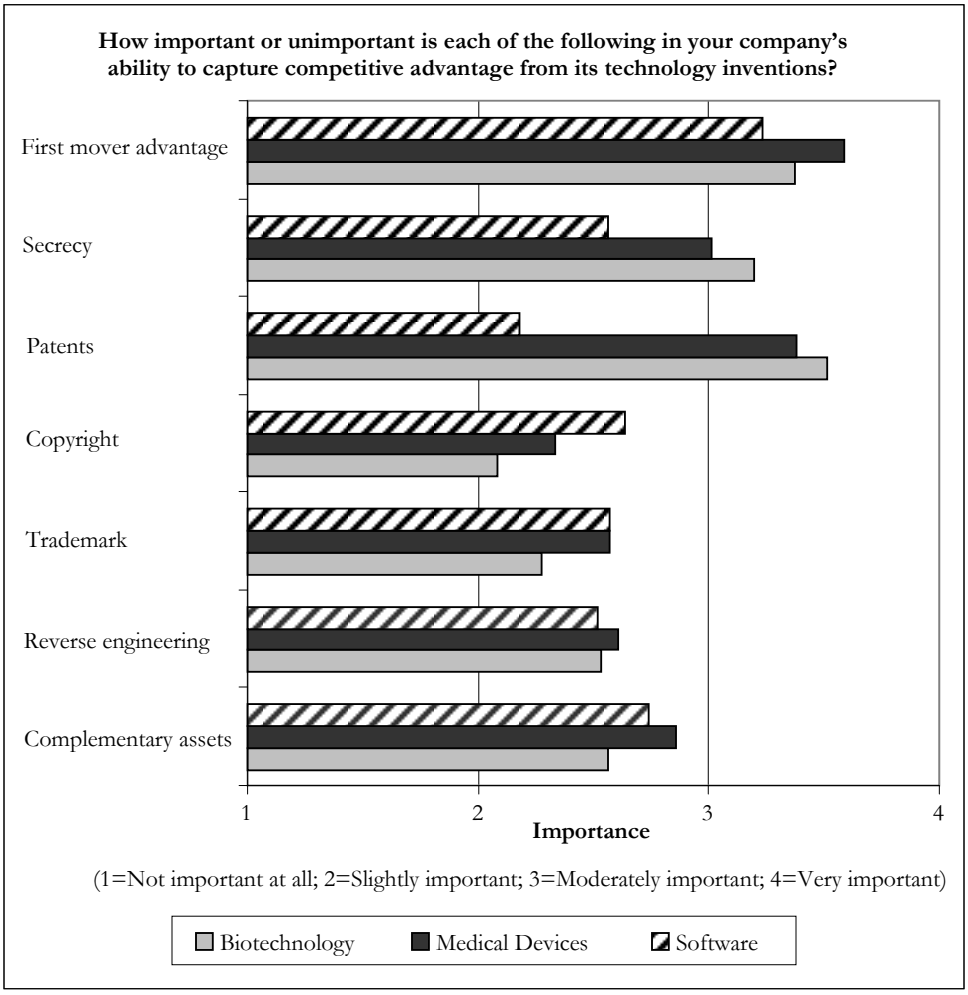
105. Scoring at the mean 3.3 on a 1–4 scale; significantly different from the next-ranked method at the 99% confidence interval.

106. These methods score at the mean 2.8, 2.8, and 2.7, respectively, on a 1–4 scale; these three are not significantly different from each other, but are significantly different from the next ranked method at the 99% confidence interval.

107. These methods score at the mean 2.6, 2.5, and 2.5, respectively, on a 1–4 scale; these three are not significantly different from each other in rank order, but the mean score for reverse engineering is significantly different from the mean score for copyright at the 99% confidence interval.

108. *See supra* note 104.

Figure 1: Capturing Competitive Advantage from Technology, by Industry[109]



1. Industry Influences Patents' Role in Competitive Advantage

One cut of our data that allows us to see profound differences in the reported importance of patents, as well as other appropriability strategies, is segmenting by industry: for biotechnology, medical devices, and even IT hardware firms in our sample,[110] patenting is ranked among the most important appropriability means (Figure 1). Among biotechnology

110. We note that IT hardware companies are found only in our *VentureXpert* sample, and thus their responses presented here must be viewed through the overall lens of the increased likelihood of venture-funded firms to use the patent system.

companies, patenting is ranked the most important appropriability strategy.[111] For medical device startups and venture-backed IT hardware companies, respondents rank patenting second, behind first-mover advantage.[112]

Our results, for all biotechnology companies combined, underscore that a firm's technological focus strongly influences startup executives' view of the importance of different appropriability strategies. For this group of firms, patenting is ranked as the most important means of capturing competitive advantage.[113] Even when we exclude the *VentureXpert* firms and focus only on the D&B firms, patenting is still rated the most important appropriability method.[114] This finding is noteworthy given that, among the sample of both large and small pharmaceutical firms reported by Cohen, Nelson, and Walsh ("Cohen et al."), patenting was considered less effective than secrecy in protecting competitive advantage from innovation.[115]

We were also surprised to find that the importance of patenting by (venture-backed) IT hardware companies is much more similar to that expressed by healthcare startups than by software and Internet companies.[116] This is a noteworthy finding because in the 1994 Carnegie-Mellon Survey, IT

111. The difference over the second most important, "first-mover advantage," is statistically significant at the 90% confidence interval.

112. In both cases, the differences between patenting and first-mover advantage were significant at the 95% confidence interval. The results for all medical device firms show that patenting is more important than the next most important method cited, secrecy, with a difference significant at the 99% level. Among the *VentureXpert* IT hardware firms, patenting was indistinguishable in importance from secrecy and reverse engineering, with all three essentially "tied" for the second most important means—although each was significantly different from the next most important method, complementary assets, at the 95% confidence interval.

113. Among these 171 respondents, "patents" is ranked first, and is significantly different from the second ranked "first-mover advantage" at the 90% confidence level.

114. However, among the 101 D&B companies responding to this question, we cannot statistically differentiate between first-mover advantage, secrecy, and patenting.

115. *See* Cohen et al., *supra* note 24. The study by Cohen et al. asked respondents to report on the "effectiveness" of appropriability means, while we chose to follow the study of Levin et al., *supra* note 24, and ask about their "importance." While comparison is made more difficult by these semantic differences, Cohen et al. nevertheless report that patenting is a less effective appropriability strategy for product innovation, and particularly for process innovation, than secrecy. *See* Cohen et al., *supra* note 24, at 9–11. Although not discussed in the Cohen et al. working paper, the original Carnegie-Mellon survey reportedly contained responses from some biotechnology firms indicating that in 1994, secrecy was substantially more effective than patenting—more so in fact than among larger pharmaceutical firms. Communication from Wesley M. Cohen, Professor, Duke Univ., to Stuart Graham, Assistant Professor, Ga. Inst. Tech. Coll. Mgmt. (Aug. 12, 2009) (on file with authors).

116. We note that IT hardware firm responses exist only in the *VentureXpert* data, and that no D&B "hardware" respondents exist in our sample, so the reported averages represent only venture-backed companies.

hardware firms (such as semiconductors and communications equipment) reported that patenting was only effective at protecting about one-quarter of their product innovations, compared with secrecy, which was effective at protecting about one-half.[117] In our results, venture-backed IT hardware firms rank patenting at least as important as secrecy.[118] Clearly, for this select sample of IT hardware companies, patenting plays a much more significant role in capturing competitive advantage from innovation than for a sample of generally larger firms surveyed by Cohen et al. over a decade ago.[119] Like the biotechnology companies, part of this difference may reflect our respondents' small size and lack of reliance upon complementary assets. Another explanation could lie in industry shifts, such as an increasing trend in intra-industry cross-licensing, as well as the rise of "fabless" semiconductor firms, which appear to be more dependent on patents than secrecy since they are more likely to license, rather than commercialize, their inventions.[120]

The value of patenting among startups in biotechnology and medical devices (and venture-backed IT hardware) stands in stark contrast to the (un)importance ascribed to patents by software and Internet firms. The limited function served by patenting in technology competition for early-stage software firms is underscored in Figure 1. In software, patenting is rated the least important among all the appropriability strategies.[121] When we focus only on software companies in the D&B sample, patenting is still the least important method,[122] ranked on average barely above "slightly important." Likewise, patenting is the lowest-rated method by venture-

117. But, interestingly, undifferentiated "computer" companies in the Cohen et al. survey reported that both patenting and secrecy were about equally effective. Cohen et al., *supra* note 24, at tbls.1 & 2.

118. While the average importance given to patents is greater than that of secrecy, the difference is not statistically significant at conventional levels.

119. *See* Cohen et al., *supra* note 24, at tbls.1 & 2. Interestingly, in our findings, preventing reverse engineering—presumably through technical design—is more important to competitive advantage for IT hardware startups than to companies in other sectors.

120. *See* Ashish Arora & Marco Ceccagnoli, *Patent Protection, Complementary Assets, and Firms' Incentives for Technology Licensing*, 52 MGMT. SCI. 293, 293 (2006); Bronwyn H. Hall & Rosemarie Ham Ziedonis, *The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979–1995*, 32 RAND J. ECON. 101, 107–08 (2001).

121. Patents are ranked last, with the difference between patenting and the next lowest ranked method, reverse engineering, significant at the 99% confidence interval.

122. For both the overall and D&B sample of software firms, patenting is rated the lowest by statistically significant margins (at the 99% confidence interval).

backed software companies, although for these companies it is not statistically different from other low-ranked methods.[123]

2. Patents' Role in Technology Competition Differs by Innovation Focus

We find that "product innovators" are substantially more likely than "process innovators" to report that patenting is important in capturing competitive advantage.[124] To relate the aforementioned appropriability strategies with the innovation focus of early-stage companies, we asked our respondents to disclose the importance (or unimportance) of various sources of innovation to their overall business strategy.[125] For two of these categories, *product* and *process* innovation, we were able to identify companies considering one or the other as their primary innovation focus.

Our analysis demonstrates that, in general, patenting is almost twice as important for product innovators than for process innovators. In fact, patenting is rated as second only to first-mover advantage by product innovators, but is rated last by process innovators. First-mover advantage, secrecy, copyright, trademark, and the difficulty of reverse-engineering are also rated as more important by product than by process innovators,[126] but not to the same extent as the differential reported in the importance of patenting between the two types of innovators. Among all the methods, only complementary assets are ranked (in absolute terms) as a more important appropriability strategy by process innovators when compared with product innovators.[127]

Noting that biotechnology companies rated patenting as more important overall than did software and Internet firms, we also examined product innovators within each of these two technologies. A strong technology effect on the ranking of the importance of patents is once again evident: biotechnology product innovators are much more likely to rate patents as

123. Venture-backed software firms rank "patents" on average the lowest in importance, but their average ranking is not statistically different from "copyright" and "trademark." Patenting is statistically different from the next most lowly ranked method, "secrecy," at the 95% confidence interval.

124. We define "product innovators" here as those that rated product innovation as "very important" but all other types of innovation as less important. Similarly, we define "process innovators" as those that rated process innovation as "very important" but all other types of innovation as less important.

125. These included: (1) product innovation; (2) process or internal tools innovation; (3) business-model innovation; and (4) design innovation (including product shape and packaging).

126. The differences associated with copyright and trademark are not statistically significant at conventional levels.

127. This difference is not statistically significant at conventional levels.

important when compared with software product innovators. Among these product-focused biotechnology companies, patenting was rated the most important appropriability strategy. Among product-focused software companies, however, patenting remains rated the least important means of successfully competing. Therefore, we believe that our main findings are driven more by technology differences than the type (i.e., process vs. product) of innovation.

3. Startups Use Multiple Methods to Compete with Technology Innovations

As was clear in the aggregated statistics presented earlier, startups across the high-technology landscape use different methods to compete in the marketplace. While patenting is playing a substantial role in all but the software and Internet sector, startups in all sectors use other appropriability strategies—possibly in complementary ways. It is noteworthy, however, that software and Internet startups, compared with companies in other sectors, tend to rate all methods about which we queried as less important. Software and Internet companies rate on average only one method, first-mover advantage, as at least "moderately important." Conversely, companies in other industry sectors rate at least three methods on average as more than "moderately important" (Figure 1).

Our in-depth interviews with respondents support these observations. One executive at a biotechnology startup—we will call him Glen—reported that the company held over 150 patents, many of which his company filed, but a substantial number of which were acquired from other entities. In describing his company's competitive strategy, he reported:

> We have three tiers when thinking about how to protect our technology with intellectual property. Tier one comprises [our basic technologies]; tier two involves the "clumping" of our [basic technologies]; tier three is a combination [of our technologies] to very specific uses. Beyond these, we adopt different strategies for different products and for different reasons. For one, we file [fewer] manufacturing patents in hopes of keeping these as trade secrets.[128]

Glen also related that, while first-mover advantage is a key component of his company's strategy, "[first-mover advantage] plays more of a role in the drug industry than it does in biotech . . . . It is also important [for us] to establish relationships within the network of similar companies and investors. Finding key partners is critical for a platform company like us."[129]

128. Interview with anonymous company founder (Mar. 2009).
129. *Id.*

These comments reflect several findings that arise from our survey results. While biotechnology firms are active patent-holders and are more likely to say that patents are important to their competitive position, they also tend to rank several, if not all, of the various means of capturing competitive advantage from technological innovation more highly relative to software companies. Moreover, to effectively compete, startups tend to report that multiple methods of appropriability are useful, even though the clustering of and relative importance ascribed to these methods do not follow a common pattern.

For instance, in terms of startups' use of intellectual property, it is noteworthy that both copyrights and trademarks play varied appropriability roles. These forms of protection are particularly salient for software firms, although even in this sector there is divergence. Among the population of software companies (as proxied by D&B), executives rate copyright as second, not statistically different from complementary assets (but both methods ranked behind first-mover advantage, which is clearly rated as the most important method).[130] For these same software-firm respondents, trademarks rank just behind copyrights and complementary assets, and are considered just as important as secrecy in capturing competitive advantage. Among venture-backed software firms, however, trademarks, copyrights, and patents are ranked behind all others, statistically undifferentiated among each other as the least important methods of capturing competitive advantage. The responses of D&B software companies differ markedly not only from venture-backed software firms, but also from all companies in other sectors. Among non-software startups, copyright protection is rated the least important of the several methods, while trademarks tend to be among the lower-ranked items.

In sum, while we find that various appropriability strategies are important to technology startups, our chief finding is that, outside of the software and Internet sector, patenting plays a substantial role in helping early-stage technology companies compete. But having learned that patenting is important in securing competitive advantage does not answer a key question: What are the specific mechanisms by which this competitive advantage is achieved? Is competitive advantage attained through added financing that patents help facilitate for the company, thus enabling it to develop a better technology and get it to market faster and more effectively than its competitors? Or is this competitive advantage won when a patent signals to suppliers and would-be customers that a company has a valuable and

130. These differences are significant at the 99% confidence interval.

important technology? Or do patents permit the company to secure its innovations, and keep competitors at bay while it develops a monopoly position that patents then serve to protect and solidify? We could answer these questions only by inquiring into the specific factors motivating companies seeking patent protection on their innovations. As such, we asked our respondents whose companies had actually filed at least one U.S. patent application about the reasons they pursued a patenting strategy.

V. SECOND FINDING: STARTUPS HAVE DIFFERING MOTIVES FOR PATENTING

We find that when technology entrepreneurs seek patent protection, they often do so for varying and, sometimes, complementary reasons. Comments made by Glen, the CEO of a biotechnology firm holding over 150 patents, are illustrative:

> We have a patent committee that decides, within a complex framework of factors, whether it is important to patent, whether competitors will copy the technology regardless of IP protection, and whether the patent will have foreseeable future value . . . . Patents also tend to legitimate [our product]. A patent can also provide a source of supplementary income and can be a badge, a branding, of a successful innovative high-tech company.[131]

Glen also told us that patents are important to his company when making deals and seeking investment:

> When doing deals, sometimes we only show our stack [of patents], and sometimes the other party wants to do a lot of due diligence on our individual patents. But we never fail to give a presentation of about one hour on them—patents play a huge role in securing investment.[132]

He also indicated how the issues of investment, copying, and competition were interrelated in the ways his company approached patenting:

> I have fidelity in my investors, and have raised over $100 million in capital. In securing my investors' support, I see our patent portfolio as an integral piece of the puzzle. Our competitors are concerned about our patents, or otherwise they'd do it themselves. Our customers don't really care—they just want the best product at the cheapest price. But if startups like ours are not diligent about securing patents, they will be crushed by larger

131. Interview with anonymous company founder, *supra* note 128.
132. *Id.*

> corporations, who will steal your technology and make it cheaper—and obliterate you.[133]

This chief executive's comments touch upon several of the factors that our study of the literature and our pre-survey conversations with entrepreneurs and investors suggested were motivators for startups to seek patent protection. During this process, we identified those factors that were most prominent, including the following: preventing others from copying products or services; improving the chances of securing investment; obtaining licensing revenues; improving the chances/quality of liquidity (e.g., acquisition/IPO); preventing patent infringement actions against the company; improving the company's negotiating position with other companies (e.g., cross-licensing); and enhancing the company's reputation/product image. In the survey, we asked whether the company had filed at least one patent, and among those answering in the affirmative, we inquired into the importance of these several motivations for patenting.[134]

A. STARTUPS FILE PATENTS TO PREVENT COPYING, SECURE FINANCIAL GOALS, AND ENHANCE REPUTATION

Across all respondents who report filing for patents,[135] the most important reason for patenting is to prevent others from copying the startup's products and services (Figure 2).[136] This result is notable because it contradicts some prior anecdotal reports indicating that the high costs of patenting and enforcing patents generally precluded startups from using patents to prevent copying and competition.[137] We approach this finding with a certain amount of caution, however, since we are mindful that the premier position given by our respondents to "preventing copying" may be at least in part the result of socially desirable response bias (in that respondents generally know patents are supposed to prevent copying and so tell us that they do just that).[138] Our qualitative interviews with respondents, however,

---

133. *Id.*

134. *See also* Sichelman & Graham, *supra* note 11 (providing a more detailed account of the survey findings on the motivations to patent).

135. 610 companies responded to this set of questions.

136. The mean for this response was statistically different from the next most important (securing investment) at a 99% confidence level.

137. *See* Mann, *supra* note 92, at 981 (finding that the benefits of patenting to earning supernormal profits are weak for startup software firms, especially for ones at very early stages).

138. Communication from Wesley M. Cohen, Professor, Duke Univ., to Stuart Graham, Assistant Professor, Ga. Tech. Coll. Mgmt. & Ted Sichelman, Assistant Professor, Univ. San Diego, Sch. Law (July 22, 2009) (on file with authors). *See generally* Catherine E. Ross & John Mirowsky, *Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health*,

offer us confidence that preventing copying is an important consideration when startups choose to patent.

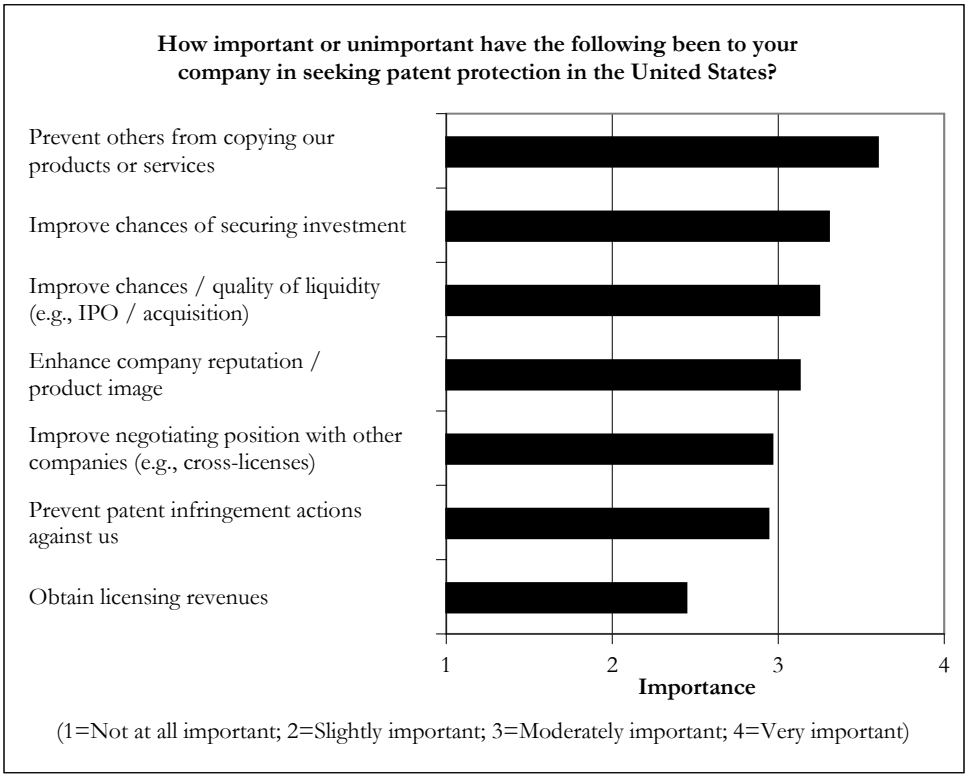
Next in importance is a cluster of several motives: improving chances of securing investment; improving chances and quality of a liquidity event (another species of securing capital);[139] and, of somewhat lesser importance, enhancing company reputation and product image.[140] This latter result, in which financial motives dominate reputation, is driven entirely by the *VentureXpert* companies in our survey. When we consider only D&B respondents, enhancing reputation is indistinguishable from securing funding and improving liquidity, although among the D&B startups, preventing copying is still the most important reason reported for filing patents.

25 J. HEALTH & SOC. BEHAV. 189 (1984) (noting the incidence of socially-desirable response bias). Further support for such a view is lent by our positioning the "preventing copying" choice as the first among our set of options presenting in the questionnaire.

139. We consider the "liquidity event" here as a means of securing capital for growth. We recognize, of course, that successful exit can also provide a return on investment for financiers and personal wealth for founders and employees.

140. There were no statistically significant differences between the rankings of securing investment and improving the liquidity event, but enhancing company reputation ranks as significantly different from both of these former reasons at a 95% confidence level.

Figure 2: Motivations for Patenting – All Startups Filing Patents



The above question was asked of those reporting that their company had filed for at least one U.S. patent (averages reported).

These results contrast with earlier large-firm surveys in which respondents ranked patenting for securing capital as relatively unimportant.[141] Our results are consistent, however, with a finding in the Carnegie Mellon survey showing that relatively smaller firms in that study tended to rank the importance of patenting to enhance firm reputation as significantly higher than did larger firms.[142] Our findings are also consistent with studies showing that patenting plays a positive role in valuation during fundraising and upon exit for venture-backed firms.[143]

The qualitative interviews we conducted with respondents following the administration of our survey add flavor to these findings. Patenting to prevent copying was an important motivator for the CEO of a software firm,

141. There is a slightly stronger tendency by the venture-backed firms to rate improving the chances of securing investment and liquidity events as more important than the D&B firms. However, the order of the reasons listed is the same for each sample set as the aggregate presented above.

142. *See* Cohen et al., *supra* note 24, at 18 n.41.

143. *See* Cockburn & MacGarvie, *supra* note 44, at 42; Hsu & Ziedonis, *supra* note 3, at 2.

whom we call Anna. She created a piece of software and filed a patent prior to founding a startup to market the software in 2003. That patent became important to her company's survival and success. Anna relates:

> A large public company copied the code of our product and tried to sell it on the market . . . . Without my patent, I wouldn't have been able to stop it . . . . [Ultimately], our company settled on the courthouse steps—literally—and we got our expenses covered, picked up a bit of money, and also established a license agreement [with the large company] to license it and pay us royalties.[144]

Regarding the relationship of patenting to entrepreneurial capital, she remarked: "Venture capital investors place a high value on companies with patents. From 2003 through 2007, I sat in on many startup and venture capital boards and, generally speaking, I found that patents were key to funding—in fact, they were the differentiator between companies."[145] These comments mirror our general findings that motivations to patent are varied, but that "preventing copying" and "financing and reputation" motives loom large for startups.

The next most important reasons for technology startups to file patents deal with defensive and strategic motives[146]—namely, to prevent infringement lawsuits and to improve negotiating position, for example in cross-licensing deals (Figure 2).[147] That startup and early-stage firms rate these motives for patenting as "moderately important" is a novel finding, insofar as previous work had implicitly assumed—at least outside of the biotechnology industry—that these firms were not targeted in enforcement (litigation and licensing) activity at sufficiently high rates to justify using patents defensively.[148] Although it appears that our early-stage technology firms state that these motives for patenting are less important than reported in earlier surveys of larger firms,[149] our result that startup companies may be

144. Interview with anonymous company executive (Apr. 2009).

145. *Id.*

146. While the means of both of these motivations are significantly different at the 99% confidence level from that of "enhancing reputation," there are no statistically significant differences between the two reasons. Taking the lesser important reason of this group—preventing patent actions against us—there was a statistically significant difference at the 99% confidence level when compared with the next most important reason—licensing.

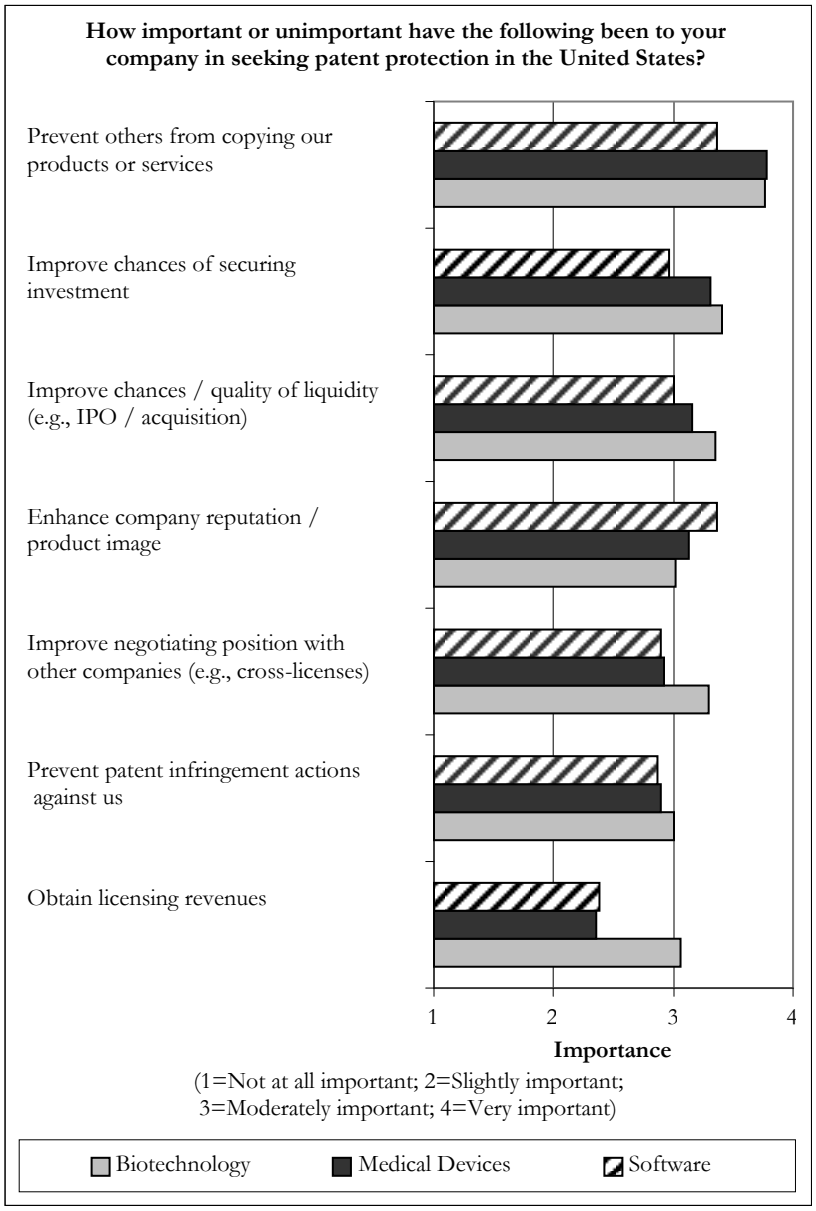
147. We are unable to differentiate, however, whether the "negotiation" value of patents is related to strategic motives or reflects the "deal-making" aspects of financing and the exit event mentioned earlier.

148. *See* Graham & Sichelman, *supra* note 5, at 1096 ("We would be somewhat surprised if many start-ups are filing for patents to improve their position in cross-licensing negotiations . . . .").

149. *See* Cohen et al., *supra* note 24, at tbls.8, 9, 10 & 11.

engaging in sophisticated uses of patents for strategic and defensive purposes is nonetheless noteworthy and deserves further study.

Figure 3: Motivations for Seeking Patent Protection, by D&B Industry



Question asked of those reporting that their company had filed for at least one U.S. patent (averages reported).

Finally, all startups rated the importance of securing licensing revenue significantly lower than other reasons, with the mean of all respondents falling between "slightly important" and "moderately important." This

finding might seem to conflict with the markets-for-technology view that small firms are more likely to license their patents because vertical specialization allows these firms to operate in upstream technology markets and provide technology inputs to (generally large) firms operating in downstream product markets.[150] Indeed, a recent survey of European patentees showed that small firms are much more likely to patent to secure licensing revenue than larger ones.[151] While our main findings show that licensing revenue is generally a comparatively unimportant consideration in startup patenting, some evidence from our study supports the view that the smallest of startup firms rely more on patenting for licensing than larger firms.[152] And as we detail in the next Section, firms in the biotechnology industry—which is often used as an exemplar of vertical specialization—are more likely to rate licensing income as an important reason to patent than are firms in other sectors we surveyed.

B. STARTUPS' MOTIVATIONS TO SEEK PATENTS DIFFER BY INDUSTRY

Consistent with the anecdotes we collected during our qualitative interviews, our results also show significant inter-industry differences in the motives for filing patents (Figure 3). We find that the health and life science companies (biotechnology and medical devices) tend to cluster in the importance they ascribe to the different motives. From a statistical standpoint, the averages presented in Figure 3 for biotechnology and medical device respondents are indistinguishable, with the exception of "obtaining licensing revenues" and "improving negotiating position," which biotechnology firms rate as significantly more important motivations to file patents.[153] Highlighting the industry distinctions, software and Internet firms' answers are all significantly different from the biotechnology and medical device firms, with the exception of "preventing patent infringement actions."[154]

In particular, biotechnology and medical device firms list preventing copying as nearly "very important" overall, while software firms place less emphasis on this motive (though still rating it between "moderately" and

150. *See, e.g.*, Arora & Ceccagnoli, *supra* note 120, at 304–05.

151. *See* ALFONSO GAMBARDELLA ET AL., THE VALUE OF EUROPEAN PATENTS: EVIDENCE FROM A SURVEY OF EUROPEAN INVENTORS 41 (2005), *available at* http://www.alfonsogambardella.it/PATVALFinalReport.pdf.

152. When we segment our respondent firms by total revenue, high-revenue entrepreneurial firms report that licensing is significantly less important to patenting than for low-revenue firms. *See* Sichelman & Graham, *supra* note 11.

153. These differences are significant at the 99% and 95% level of confidence, respectively.

154. These differences are significant at the 95% confidence interval or above.

"very" important).[155] The biotechnology and medical device companies also cite patenting to secure investment and to improve the chances and quality of a liquidity event as much more important motivations than do software firms.[156] Finally, biotechnology firms place much greater emphasis on patenting to obtain licensing revenue than all other firms, including medical device firms.[157]

C. PATENTS SERVE AN IMPORTANT FUNCTION IN THE FINANCING OF STARTUPS

Another noteworthy finding of our study is that patents play an important role in the financing of many startup companies, both during the initial stages and subsequent development of the firm, and also at the liquidity or exit event. Entrepreneurs and startup firms can face substantial barriers when seeking to secure the financial resources necessary to grow and to survive, largely due to their small size and limited experience.[158] Compounding these problems, startups often lack observable measures of success since they generally have few assets and little to no operating history.[159] The uncertainty created by this limited information makes it difficult for potential investors to appraise the quality and profit potential of the enterprise, and as a result, these investors must assess the value of startups through other readily available measures.[160]

Several commentators have suggested that patents can serve as quality signals for startup investors.[161] For instance, David Hsu and Rosemarie Ziedonis have recognized that while many characteristics of the firm can be used as quality signals associated with future profits, patents have been

155. Biotechnology and medical device firms showed statistically significant differences in their mean responses for the importance of preventing copying from hardware and software firms at a 99% confidence level.

156. Biotechnology and medical device firms showed statistically significant differences from hardware and software firms at a 99% confidence level.

157. Biotechnology firms showed statistically significant differences from medical device, hardware, and software firms at a 99% confidence level.

158. *See* A.L. Stinchcombe, *Social Structure and Organizations*, *in* HANDBOOK OF ORGANIZATIONS 142–93 (James G. March ed., 1965).

159. *See* Tyzoon T. Tyebjee & Albert V. Bruno, *A Model of Venture Capitalist Investment Activity*, 30 MGMT. SCI. 1051, 1053 (1984).

160. *See* Toby E. Stuart, Ha Hoang & Ralph C. Hybels, *Interorganizational Endorsements and the Performance of Entrepreneurial Ventures*, 44 ADMIN. SCI. Q. 315 (1999).

161. *See* Hsu & Ziedonis, *supra* note 3, at 1–2 (suggesting that entrepreneurial lineage, founder backgrounds, and affiliations with reputable third parties such as corporate partners, venture capitalists, and investment bankers, can serve as important quality signals); Long, *supra* note 97, at 655–59 (setting forth a "signaling" theory of patents).

underappreciated in that role.[162] Moreover, patents are costly assets, and therefore fit the cost criteria for a quality signal laid out by Michael Spence, who suggested that a credible signal ought to be costly, both in terms of direct pecuniary costs and effort.[163]

While these theories can be read to apply to many types of investments, existing empirical tests of the value of patents tend to examine only venture capital, a species of investment that comprises a relatively small slice of the overall entrepreneurial finance pie.[164] In contrast to relatively difficult-to-secure VC financing, startups are more often funded by angel investors and commercial banks, and most often by friends and family.[165] The Kauffman Foundation Firm Survey, which tracks a cohort of companies founded in 2004, shows that while 1% and 5% of companies started in their founding year by exchanging ownership in the company for VC and angel funding, respectively, 7% took personal loans from friends and family, 13% supported the founding with personal loans, and 39% used personal credit cards to finance at least some part of the startup.[166]

1. Investors and Entrepreneurs Highlight the Role of Patents in Startup Financing

Unlike previous studies, our survey examined various sources of startup investment, and the results show that patenting plays a more substantial role in supporting many different species of entrepreneurial capital investment than has been commonly believed. Several of the qualitative interviews we conducted illuminate this finding. For instance, the partner we call Stan, who works with a VC firm and invests primarily in life-science companies, told us,

> When thinking about the life-cycle of a company, in many respects the value of the IP is really generally assessed at the early stage by the first stages of "professional" money. These early-stage

162. *See* Hsu & Ziedonis, *supra* note 3, at 10–12.

163. Michael Spence, *Job Market Signaling*, 87 Q.J. ECON. 355, 358 (1973).

164. *See, e.g.*, Dirk Engel & Max Keilbach, *Firm-Level Implications of Early Stage Venture Capital Investment - An Empirical Investigation*, 14 J. EMPIRICAL FIN. 150 (2007); Paul Gompers & Josh Lerner, *The Venture Capital Revolution*, 15 J. ECON. PERSP. 145 (2001); Steven N. Kaplan & Per Strömberg, *Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts*, 70 REV. ECON. STUD. 281 (2003). *But see* Bernard S. Black & Ronald J. Gilson, *Venture Capital and the Structure of Capital Markets: Banks Versus Stock Markets*, 47 J. FIN. ECON. 243 (1998).

165. Alicia Robb & David T. Robinson, The Capital Structure Decisions of New Firms (Feb. 11, 2009) (unpublished manuscript, on file with Social Science Research Network), *available at* http://ssrn.com/abstract=1345895.

166. Alicia Robb & David DesRoches, *Kauffman Firm Survey – Baseline/First/Second/Third Follow-Up* 446–48, 459–63 (Apr. 27, 2009) (unpublished manuscript, on file with Social Science Research Network), *available at* http://ssrn.com/abstract=1024312.

> [professional investors] will do a great deal of scrutiny of IP of all types, but especially of trade secrets and patents. If you think about a patent estate as having a life, most of its validity is presumably established when the first guys do their investment. Later investors experience less need to invest as much effort into due diligence as the early-stage investors did—there's generally no need to go back and repeat what's already been done. What will happen is that later investors will look at marginal change in the patents of a company since the last investigation. . . . A reason why patents are so important in the biotechnology industry in particular is that, when one makes a biotech investment, fundamentally one is making an IP investment. Consequentially, the early-stage venture investors dig very deeply into the validity of that IP.[167]

These comments resonate with our empirical findings: investors of many types value patents as an input into their investment decision, particularly venture capital investors in the life sciences.

Startup executives whom we interviewed also remarked about the importance that patents play in convincing investors to fund the startup. Neil, the CEO of a biometrics information company, suggested to us that "investors were interested in patents, and it was a key question that came up during negotiations. But our company does not hold patents as their [sic] core investment—instead we focus on our services to earn revenues."[168] This latter statement by Neil shows that there remains some ambiguity about the role played by patents in securing funding. Even among those companies that hold patents, we found similar ambiguity. For instance, our interview with Jeremy, the sole proprietor of a medical device company who filed one patent application, yielded this comment: "I applied for a patent to make it clear to investors what exactly it was that I owned. But seeing [the device] live is better than reading a patent, and I think that a live demonstration of [my device] is better in securing their investment."[169]

These comments from technology entrepreneurs are instructive, and help to highlight two important findings of our survey. First, patents appear to play a significant role in the funding decisions of many different types of startup investors. But, second, patenting may not be a necessary condition for access to entrepreneurial capital.

167. Interview with anonymous venture capital firm partner (May 2009).
168. Interview with anonymous executive (Mar. 2009).
169. Interview with anonymous founder (Feb. 2009).

2. *The Role of Patents in Attracting Entrepreneurial Capital*

Our survey specifically asked respondents about the role that patents play in securing investment from six different sources: friends and family; angel investors; venture capital investors; other companies as investors (corporate venture capital); investment banks; and commercial banks (such as credit or loans). We inquired whether the company had engaged in negotiations with any of these six sources of startup capital and whether the company had been funded by that source. Moreover, we asked respondents to report whether each of the funding sources had indicated that the startup having patents was an "important factor" in that source's funding decision.[170]

a) Startups Report That Patents Are Important to Their Investors

Our respondents report that patenting plays a more significant role in attracting funding than had been previously believed. It is widely held that VC investors rely on patents in their investment decisions, although the reasons for this reliance are unclear. Some suggest that a patent is an important signal of quality in an uncertain investment environment, and that by relying on the independent expertise of the Patent Office, the investor can dispel some of the information asymmetries between the investor and the startup.[171] A variant signaling theory maintains that while the Patent Office's pronouncements are not very meaningful, the fact that startup managers had the cognizance and wherewithal to file for patents is an important sign of their managerial sophistication, particularly in codifying inchoate knowledge.[172] An alternative explanation is that patents tend to provide sufficient freedom to operate, allowing a company to develop and commercialize its embryonic products.[173] Another theory suggests that

170. We understood that, like the problem of "hearsay," relying on one person's interpretation of the beliefs of another can be a problematic approach. In an attempt to mitigate against this problem, we asked the respondent whether the source had "indicated" that patents were an important factor.

171. *See* Hsu & Ziedonis, *supra* note 3, at 5; Long, *supra* note 97, at 649.

172. *See* Graham & Sichelman, *supra* note 5, at 1078–79.

173. *See* Gideon Parchomovsky, *Publish or Perish*, 98 MICH. L. REV. 926, 930 (2000). Technically, patents are only rights to exclude others from making, using, or selling the patented invention. 35 U.S.C. § 271 (2006). However, patents can also serve as "prior art" that can be used to prevent *others* from patenting. Thus, holding patents in a relevant field will tend to exclude others from patenting in that field, thereby providing the patent holder greater freedom to operate. Additionally, patents can enhance freedom to operate by preventing lawsuits from other patent holders, who might be wary of being countersued in litigation. *See* Graham & Sichelman, *supra* note 5, at 1065.

investors require the companies to have patents so as to enjoy these IP rights as residual claimants should the venture fail.[174]

While our data do not shed light on these different explanations *per se*, we are able to report some results on the perceived importance of patents in startup funding.[175] Our figures reflect the share of those respondents who reported that a potential funding source (e.g., friends and family) with whom the respondent negotiated had indicated that having patents was important to their funding decision. For D&B companies that negotiated with VC firms, just over two-thirds (67%) had indicated to the respondent that patents were an important factor. This share is higher among our venture-backed sample, with just above three-quarters (76%) indicating as much. There are also notable technology differences. Among the D&B companies, 60% of software firms reported that VC investors considered patents important, while that figure rose to 73% for D&B biotechnology firms and 85% for D&B medical device companies.

Surprisingly, respondents reported that patents are also important to many commercial banks and "friends and family" investors. In our D&B sample, companies negotiating with friends and family declare that 31% consider patenting important to making a funding decision. For those D&B companies negotiating with commercial banks—presumably, mainly for loans or credit—the figure is just over one-fifth (21%). The shares reported by the venture-backed sample are similar for friends and family (35%), but increase markedly (49%) for commercial banks.[176]

Our respondents also indicated that patents are important to other sources of entrepreneurial capital. D&B companies report that 57% of angel investors, 54% of "other companies" as investors, and 50% of investment banks indicated that patents were important to their investment decision. Among our venture-backed sample, the figure for angel investors is similar to that for D&B firms (59%), but patents were reported to be significantly more important to investment banks (61%) and other companies as investors (70%).[177]

---

174. *See id.* at 1078.

175. The importance associated with "enhancing reputation" among the startups that filed at least one patent may reflect, at least in part, a desire by the company to become more attractive to investors.

176. The importance of patents to commercial banks among venture-backed firms may be indicative of recent trends in "venture debt." Darian M. Ibrahim, *Debt as Venture Capital*, 2010 U. ILL. L. REV. (forthcoming 2010), *available at* http://ssrn.com/abstract=1418148.

177. These differences are significant at the 90% and 99% confidence interval, respectively.

b) Patents' Reported Importance to Investors Differs by Industry

As with other aspects of patenting by startups, the industry differences in the reported importance of patents to investors is noteworthy. Among D&B biotechnology firms for instance, respondents were much more likely to reveal that commercial banks considered patenting by the target firm important (43%) than were D&B software firm respondents (13%). This difference is also notable among friends and family, who considered patents to be important for more than half (55%) of D&B biotechnology firms, but less than one-quarter (23%) of D&B software companies. There are also significant differences for angel investors (71% for biotechnology and 53% for software), venture capital (73% and 60%, respectively), other companies as investors (64% and 42%, respectively), and investment banks (62% and 36%, respectively). Clearly, across all funding sources, respondents declare that potential investors view patenting as much less important for software and Internet as for biotechnology companies.

These industry differences are also marked in our venture-backed sample. Among the venture-backed biotechnology firms, for instance, friends and family were much more apt to indicate that patenting was important (61%) than among venture-backed software firms (18%). This difference is also notable in the responses concerning commercial banks, with companies reporting that this funding source indicated patenting as being important for almost three-quarters (74%) of venture-backed biotechnology firms, but for less than one-third (31%) of venture-backed software companies. There are also significant differences among angel investors (78% for biotechnology and 36% for software), venture capital (97% and 59%, respectively), other companies as investors (90% and 51%, respectively), and investment banks (81% and 40%, respectively).

In sum, our survey respondents report that patents are being widely demanded by different sources of entrepreneurial capital, though that demand does not extend to all funding negotiations, and the incidence of interest in patents is highly variable. While a caveat is in order—these findings are based upon the perceptions of the recipient about what was in the mind of the investor—the executives of the startups were nevertheless privy to negotiations, and can be expected to have unique insights into what occurred during their funding negotiations and what documents and information were requested of their companies. We are therefore reasonably

confident that our data present at least a proximate window into the importance of patenting to various startup investors.[178]

VI. THIRD FINDING: TECHNOLOGY ENTREPRENEURS VARY IN THEIR REASONS FOR NOT SEEKING PATENTS

There are several reasons for startup firms opting against patent protection,[179] including the belief that the technology is not patentable; the high costs associated with prosecuting and enforcing the patent; the perception that, with reverse engineering, that patents may afford relatively weak protection; the fear of disclosure;[180] and the availability of other forms of protection.[181] Rather than simply asking our respondents that hold no patents to report on their motivations for choosing against patenting, we wanted to uncover the nuances underlying decisions to forgo patenting—even among those that were patenting other inventions. As such, all respondents were asked whether the last major technology innovation they did not patent was a product or a process (or not), and what reasons motivated their company's decision not to patent.

Our study of the literature and previous surveys, as well as discussions with entrepreneurs and investors, allowed us to generate a list of the most common reasons why, reportedly, startups choose not to patent their

178. In fact, our numbers may be biased downward. Although some investors may uncaringly request a firm's patents through pro forma "due diligence lists," because patent information is often publicly available, investors may often *not* disclose their interest in a firm's patents, rather than *misrepresent* concern when they have none.

179. For more analysis, see Sichelman & Graham, *supra* note 11, which contains a more detailed discussion of our results on why startups choose not to patent.

180. Patenting is sometimes viewed as a "substitute" for secrecy, the former tending to vitiate any long-term attempts at secrecy through its disclosure requirements. *See, e.g.*, David D. Friedman, William M. Landes & Richard A. Posner, *Some Economics of Trade Secret Law*, 5 J. ECON. PERSP. 61 (1991). *But cf.* Stuart J.H. Graham & Deepak Somaya, Vermeers and Rembrandts in the Same Attic: Complementarity Between Copyright and Trademark Leveraging Strategies in Software (Feb. 23, 2006) (unpublished manuscript, on file with the Social Science Research Network), *available at* http://ssrn.com/abstract=887484 (suggesting that a complements view is more appropriate when viewing IP protection at the level of the innovation or the firm).

181. Other forms of protection include copyright, which is available for computer programs. *See* Apple Computer, Inc. v. Franklin Computer Corp., 714 F.2d 1240, 1246–47 (3d Cir. 1983) (source code and object code); Williams Elecs., Inc. v. Artic Int'l, Inc., 685 F.2d 870, 875 (3d Cir. 1982) (object code); Stern Elecs., Inc. v. Kaufman, 699 F.2d 852, 855 n.3 (2d Cir. 1982) (source code). Trade secret protection may also be available, generally under state law. *See, e.g.*, Jackson v. Hammer, 653 N.E.2d 809 (Ill. App. Ct. 1995); Gonzales v. Zamora, 791 S.W.2d 258 (Tex. App. 1990); Oberg Indus. v. Finney, 555 A.2d 1324 (Pa. Super. Ct. 1989).

innovations. These included: not wanting to disclose information; the cost of getting the patent, including attorneys' fees; that competitors could have easily invented around the patent; that they believed trade secret was adequate protection; the cost of enforcing the patent, including actions in court; that they did not believe the technology was patentable; and that they had no need for legal protection.[182] By constructing our questionnaire in this manner, we were able to collect information on what is essentially a sample of recently un-patented major technologies generated by startup companies.

A. COST CONSIDERATIONS LOOM LARGE FOR STARTUPS IN DECIDING TO FORGO PATENTING

We find that, among technology startups, the cost of getting a patent is the most common reason cited for not patenting a major technology. Figure 4 shows the shares of reasons respondents reported for not patenting their company's last major technology innovation.[183] Cost considerations in patenting loom large, with the cost of prosecuting and the cost of enforcing the patent cited by more respondents than any other reason. These motivations are followed closely by the ease of inventing around the patent.[184] These results are similar to those found in a Small Business Administration survey conducted in 1998 of small firms, which listed these same reasons at the top of small-business motivations for forgoing patenting.[185] In contrast, the difficulties and costs of acquiring and enforcing patents tended not to be salient for larger firms in other surveys.[186] Yet, these same studies show that even in the large firm surveys, relatively smaller firms tend to report a significantly higher sensitivity to the costs of filing and enforcing patents.[187]

182. Our questionnaire allowed respondents to indicate multiple reasons. In a follow-on question, we asked our respondents to report which of these motivations was the most important reason.

183. This question could be answered by respondents regardless of whether they had filed for a patent. The percentages for each reason do not add up to 100% because respondents could check one or more of these selections.

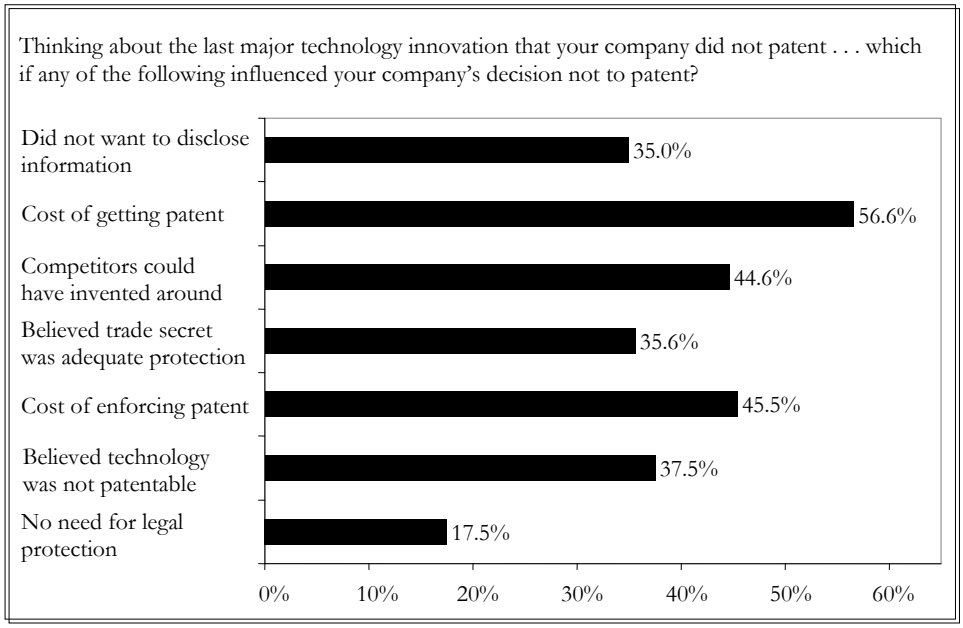
184. The difference between the reported percentage for the costs of acquiring and the costs of enforcing the patent was statistically significant at a 99% confidence level. The costs of enforcing the patent and the ease of inventing around the patent did not show statistically significant differences from one another, but they were statistically significantly different from the next reported reason to a 99% confidence level.

185. JOSEPH J. CORDES, HENRY R. HERTZFELD & NICHOLAS S. VONORTAS, A SURVEY OF HIGH TECHNOLOGY FIRMS 55–58 (1999), *available at* http://sba.gov/advo/research/rs189tot.pdf.

186. *See* Cohen et al., *supra* note 24, at 15–16.

187. *See id.* at 25.

Figure 4: Reasons for Startups to Forgo Patent Protection on Major Technologies



Respondents were asked to indicate all the reasons that applied (share of respondents indicating that the option influenced the decision is reported).

Other evidence from our survey results indicate that the greater sensitivity of smaller firms to costs is not merely due to capital constraints. Another of our survey questions revealed that the average out-of-pocket cost for a respondent firm to acquire its most recent patent was over $38,000.[188] This figure is significantly higher than the averages for patent prosecution reported in the literature, which vary from a low of $10,000 to a high of $30,000.[189] In one of the unstructured hour-long interviews we conducted with respondents, one executive at a venture-backed semiconductor firm stated that startups often pay significantly more than incumbents to their prosecuting attorneys, because startups (1) tend to file for patents on inventions that are more important to the company's core business model than large firms; (2) usually use outside instead of in-house counsel for patent prosecution; and (3) often have difficulty monitoring outside counsel to limit

188. We asked this question only of those respondents who reported that their company had both filed and been granted a U.S. patent.

189. *See* AM. INTELLECTUAL PROP. LAW ASS'N, REPORT OF THE ECONOMIC SURVEY 2007 78–81 (2007) (reporting average attorneys' fees for prosecuting an original patent application, filing one amendment, and issuing an allowed application as between $10,000 and $20,000, depending on the complexity of the technology); Mark A. Lemley, *Rational Ignorance at the Patent Office*, 95 NW. U. L. REV. 1495, 1498 (2001) (estimating the cost of prosecuting a patent to issuance as between $10,000 and $30,000).

overall costs.[190] Indeed, a non-trivial percentage of our respondents—about 10%—listed cost as the only barrier to filing for a patent from among the options we offered. Additionally, when asked to indicate the most important reason for not filing, more than one-third of the respondents selected the costs of acquiring or enforcing the patent.[191]

Fewer respondents, but nevertheless a substantial number, reported that their firms did not seek patent protection because they considered their innovation unpatentable, believed that trade secret protection was adequate, or were reluctant to disclose information.[192] These last two reasons in fact may be complementary, and may be associated with the cost considerations cited earlier. The relative infrequency of startups citing these reasons tends to track the responses by large firms in other surveys,[193] although the reluctance to disclose information appears to be more of a deterrent for large firms than for our early-stage respondents. Part of this difference reflects our survey's heavy focus on software firms, whose executives less frequently cited "reluctance to disclose information" than executives of startups in other sectors. Finally, we note that about 18% of our respondents declared "no need for legal protection" as a motive for not filing a patent for their last major technological innovation.

B. STARTUPS' REASONS TO FORGO PATENTING DIFFER BY INDUSTRY

A major finding of our study is that the most important reason that biotechnology companies cited for not pursuing a patent on their last innovation was a reluctance to disclose information, while software companies most frequently cited patent prosecution costs.

1. Biotechnology Startups Most Commonly Cite Reluctance to Disclose Information as a Reason Not to Patent

In an effort to better understand the drivers of startups' choices to forgo patenting on their major innovations, we segmented the responses by technology and report the results in Table 2. We find that the most marked

190. Interview with anonymous semiconductor company executive (Feb. 20, 2009).

191. Specifically, respondents identified the following as the most important reasons for not patenting their last innovation: cost of acquiring the patent (26.00%); did not believe technology was patentable (20.92%); did not want to disclose information (15.84%); ease of inventing around the patent (12.41%); cost of enforcing the patent (10.52%); no need for legal protection (7.33%); and believed that trade secret protection was adequate (6.97%). *See infra* Section VI.A tbl.2.

192. The responses for these reasons were not statistically significantly different from one another, but they were from the next reason—no need for legal protection—to a 99% confidence level.

193. *See, e.g.*, Cohen et al., *supra* note 24; Levin et al., *supra* note 24.

divergence occurs between biotechnology and software companies, and we note the differences in the share of companies that reported each motivation (along with statistical significance, in the rightmost column of the table). Biotechnology firms are more than twice as likely to cite "disclosing information" as a reason to forgo patenting as are software firms (59% and 25%, respectively). Biotechnology firms are also more likely to believe that trade secret is an adequate means of protecting their innovations than software firms (49% and 29%, respectively), although this difference may be a consequence of the differences in the likelihood of an unpatented invention being a process technology, a possibility that we explore below.

Table 2: Reasons for Not Seeking Patent Protection – Selected Industries

Thinking about the last major technology innovation that your company did not patent . . . which if any of the following influenced your company's decision not to patent?

Category / Reason	All respondents	Biotechnology	Software	Difference	Test of difference
Did not want to disclose	35%	59%	25%	+ 34%	**
Cost of filing	55%	43%	64%	- 21%	**
Ease of inventing around	44%	42%	46%	- 4%	
Trade secret was adequate	36%	49%	29%	+ 20%	**
Cost of enforcing	44%	36%	52%	- 16%	**
Did not believe patentable	38%	28%	42%	- 14%	**
Did not need protection	17%	17%	20%	- 3%	
Total responses	1,057	136	589		

** Differences noted, significant at the 95% confidence intervals. Tests for differences in means were conducted between columns, within rows.

2. Software and Internet Startups Most Commonly Cite Cost Considerations as a Reason Not to Patent

Patenting and enforcement costs are cited much more frequently by software and Internet firms as motives for not patenting, a finding that is consistent with the lower significance (as we report above) that software firms ascribe to patents as a means of securing competitive advantage.[194] We

194. If the asset is considered "less valuable," sensitivity to cost may be expected to be higher.

note that software firms were also substantially more likely to say that they did not believe that the technology was patentable than were biotechnology companies (42% and 28%, respectively), although we are unsure whether the term "unpatentable" triggered a subjective belief in the respondent about the requirements of the patent laws or a more philosophical belief about what *ought* to be patentable subject matter.[195]

3. The Most Important Reason Not to Patent Also Differs by Industry

To add even greater specificity to our results, we asked our respondents to report which of the several reasons for choosing not to patent the technology was the *most* important. The results show that, among all respondents, the cost of filing and the belief that the technology was not patentable are most highly cited (25% and 21%, respectively). As with many of our other results, however, there are substantial industry-specific differences.

Among biotechnology firms, over one-third (34%) cite a reluctance to disclose information as the most important consideration, a level that is interestingly nearly matched among venture-backed IT hardware firms (32%). These industry differences notwithstanding, the cost of filing remains a significant impediment for these startup firms, with greater than 20% of companies in each sector citing this reason as the most important determinant of forgoing patenting. This finding adds weight to other results in this survey suggesting that these technology startups are sensitive to the costs of patenting, even when patenting is seen by the executives as an important factor in their commercial and entrepreneurial success.

C. STARTUPS' MOTIVES TO FORGO PATENTING DIFFER BY INNOVATION TYPE

We further explore the differences in the reasons to forgo patenting by examining whether the innovation is a product or process technology. The results suggest that the cost of applying is a particularly salient factor for those companies that chose to forgo patenting on a product technology, with nearly one-third (32%) citing this reason as the most important factor. When firms choose to leave process technologies unpatented, they are most likely to cite three reasons about equally: reluctance to disclose information (24%), a belief that the technology was unpatentable (22%), and, again, the cost of

195. *See generally* Pamela Samuelson & Robert J. Glushko, *What the User Interface Field Thinks of the Software Copyright "Look and Feel" Lawsuits (and What the Law Ought to Do About It)*, 22 ACM SIGCHI BULL. 13 (1990) (reporting survey results and finding that software engineers preferred copyright protection over patent protection).

filing (21%). As such, because biotechnology companies were more likely to report that their last unpatented innovation was a process innovation, part of the inter-industry differences described earlier may be explained by underlying differences in the types of technologies respondents were contemplating patenting.

VII. TECHNOLOGY ENTREPRENEURS MUST RECKON WITH PATENTS HELD BY OTHERS

We were not only interested in determining how entrepreneurial companies use (or choose not to use) the patent system, but also how they deal with patents held by others in their market environment. When viewed as barriers to innovation, patents may create a minefield for various innovative activities, from invention, to development, to commercialization. This minefield analogy may be particularly relevant in the information technologies, to the extent that these arts are more likely characterized by innovation that is both cumulative (building on earlier generations) and complex (requiring more than one patentable technology for commercialization). Nevertheless, relying on interviews and anecdotes, Ronald Mann has suggested that patents in the competitive environment are often ignored, at least by startup software firms.[196]

Another downside of patents in a startup's competitive environment is the threat of patent disputes and, when negotiation fails, costly litigation. Startups may be particularly sensitive to accusations of infringement because they are likely to experience resource constraints when faced with the costs of funding a suit, estimated for most suits to be between $3 million and $6 million per litigant through appeal.[197] Suits may come in the form of "bullying" by larger competitors trying to put the startup out of business,[198] or even from "trolls" or non-practicing entities seeking royalty payments. These accusations of infringement are particularly problematic when the underlying patent being wielded against the startup is more likely than not invalid:[199] the resource-constrained startup may find that its least-costly alternative is simply to pay licensing fees, thus allowing the firm to avoid suffering the huge costs of litigation—even though the patent is on its face

196. *See* Mann, *supra* note 92, at 977.

197. *See* AM. INTELLECTUAL PROP. LAW ASS'N, REPORT OF THE ECONOMIC SURVEY 2009, at 29 (2009) (reporting that patent litigation suits with over $1 million at stake cost roughly between $3 million and $6 million).

198. *See* Graham & Sichelman, *supra* note 5, at 1080–81.

199. All patents are probabilistic rights until the last court has spoken. *See* Mark A. Lemley & Carl Shapiro, *Probabilistic Patents*, 19 J. ECON. PERSP. 75, 75–76 (2005).

invalid. If the litigation is launched by a "bully" attempting to put the startup out of business, the suit may indeed yield such a result if the startup is unable or unwilling to tap the capital markets to fund lengthy and expensive litigation.[200]

But patents in the market environment may be viewed positively by the startup, too. Patents can serve as mechanisms by which markets for technology develop, allowing some firms to specialize in a technology market in a manner analogous to Adam Smith's division of labor.[201] Patents may therefore serve as a means by which technology, information, and know-how can be more easily transmitted, since there are a host of problems associated with transacting in intangible knowledge assets that patents-as-definable-chits help to solve.[202] Given these discordant, but not necessarily mutually exclusive, views of the roles played by patents held by parties external to the startup, we fashioned a series of questions aimed at better understanding how early-stage companies react to patents in their competitive environment.

A. LICENSING-IN PATENTS: (SOME) STARTUPS LICENSE FROM OTHER PATENT-HOLDERS

We were interested in discovering how commonly startups license patents from third parties. Moreover, given the theories elucidated above, we wanted to ascertain whether, and under what circumstances, patent licenses were taken to gain technology, information or know-how, or to settle disputes. We find that among D&B companies, 15% of respondents licensed-in a patent. This figure is significantly higher, over one-third (37%), among the venture-backed sample.[203] In addition to these differences among companies based on their funding source, we again find divergence based on the sector in which the company operates: in the D&B sample, biotechnology companies are significantly more likely to license-in (37%) than are software firms (8%).

200. *See generally* CONSTANCE E. BAGLEY & DAVID LANE, X-IT AND KIDDE (A) (2003) (detailing a startup's experience with infringement by a larger competitor, and the difficulty experienced by the startup in securing funding to pursue litigation).

201. ADAM SMITH, AN INQUIRY INTO THE NATURE AND CAUSES OF THE WEALTH OF NATIONS 12–13 (1776). Instead of some specialized employees in a pin factory making the shafts and others making the heads, the analogy would be that some firms make technology inputs like software while other firms make the technologies to which those inputs are applied, such as hardware.

202. *See* Ashish Arora & Robert P. Merges, *Specialized Supply Firms, Property Rights and Firm Boundaries*, 13 INDUS. & CORP. CHANGE 451, 454 (2004).

203. Difference significant at the 99% confidence interval.

1. Startups License Both to Gain Knowledge and to Settle Disputes

At least in part, the rationale for having a patent system is that by offering a limited monopoly on inventions, society will acquire more invention.[204] This property right is limited in both scope and time, with the understanding that after a period of years, the patented invention will fall into the public domain and be available for all to use. In the meantime, many of the patent law's provisions are tailored to require adequate disclosure of the invention, so that other innovators may learn from the disclosure, and possibly improve upon it.[205] The operation of markets for technology are also believed to help in this dissemination function—specifically, by relying on the markets to exchange patented knowledge, and thus permitting these intangible chunks of creativity to be propertized and transacted over.[206]

Startup firms reported that they often seek to gain knowledge, information, or know-how by licensing patents from others, yet they also indicated that they frequently—and sometimes *only*—license to avoid a patent dispute. When asked why their company took their last patent license, a majority of all biotechnology company respondents (81%) indicated that they intended to gain (at least in part) technology, information, or know-how, while three in ten responded that they licensed-in the last patent (at least in part) to avoid a patent dispute, or for other defensive or freedom-to-operate considerations.[207] Among the (fewer) software firms that took patent licenses, 79% reported doing so at least in part to gain information, technology, or know-how. Among both the biotechnology and software D&B companies, less than one in ten reported that the only reason they took a patent license was to avoid or settle a patent dispute.

Among the companies receiving venture funding, biotechnology firms were again significantly more likely to in-license a patent (72%) than were software startups (13%). We find that nearly nine in ten of patent-licensing, venture-backed biotechnology startups declared that the last patent licensed was taken to gain (at least in part) technology, information, or know-how. Nearly two in ten of these licenses were taken (at least in part) to settle a dispute. Among the same firms, only two of sixty-three (3%) said that the sole reason they took their last license was to avoid or settle a patent dispute. As such, our results suggest that, on average, early-stage biotechnology

204. *See* Arrow, *supra* note 4.

205. ROBERT P. MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 259 (3d ed. 2002).

206. *See* ARORA ET AL., *supra* note 9, at 116–17 (suggesting that intellectual property protection supports more efficient markets for technology).

207. These latter categories were self-defined by the respondents in an "other" category.

companies are not facing demands for licensing payments on patents that are unrelated to technology that they are working upon or can foresee working upon. In general, these firms appear to be taking licenses to patents that bring some knowledge capability into the startup.

2. *Venture-Backed IT Startups Take More Patent Licenses*

We uncover that some early-stage IT companies may be more likely to face nuisance patent disputes than has been commonly reported. Among the venture-backed IT companies responding to our questions, nearly four in ten (39%) IT hardware firms (semiconductor, communications, and computer hardware) reported taking a patent license, compared with just over one in ten (12%) among venture-backed software firms. These patent-licensing IT hardware and software firms are almost equally likely (about seven in ten) to have taken their last patent license to (at least in part) gain technology, information, or know-how.

In a surprising result, both these types of venture-backed IT startups report that, among those that took a patent license, approximately one in four (25% and 22%, respectively) of their last patent licenses were taken to avoid or settle a patent dispute, but *not* to gain technology, information, or know-how. This share is several times higher than the same figure reported by D&B software companies (8%) and by biotechnology companies (6%) regardless of their funding characteristics.

We find this result intriguing—that a quarter of venture-backed IT firms that have taken patent licenses report that the last license was taken solely to settle a dispute, and not to gain technology, information, or know-how. In fact, these respondents did not suggest any other reasons for taking these licenses,[208] but there are some caveats we must offer. While one-quarter of venture-backed software firms suggest that ending a dispute was the sole reason for taking their last patent license, the number of firms taking any patent license is quite small (12%). As a result, only 3% of all software firms in our sample report taking their last license solely to avoid or settle a patent dispute.

But we can apply the same calculation to the hardware firms in our sample, all of which are venture-backed, and find that nearly four in ten report taking a patent license. This result shows that one in ten (10%) of the venture-backed IT hardware startups take a patent license solely to avoid or to settle a patent dispute. We note that in contrast to the pattern exhibited by

208. Respondents were given the following choices: "For the last patent that your company licensed in, was the license taken (mark all that apply): (a) to gain technology, information, or knowhow; (b) to settle a legal dispute; (c) other [specify]."

biotechnology firms, these figures associated with settling a patent dispute are much higher for venture-backed IT firms than among their (generally non-venture-backed) D&B software counterparts.[209]

What is behind this increased likelihood of venture-backed IT firms facing threats of litigation? The unstructured hour-long interviews that we conducted with a sample of our respondents produced one story of just such a patent dispute. One executive of a venture-backed IT company informed us that his firm had been the target of a cease-and-desist letter during the company's initial SEC Registration prior to its IPO. He believed the patent was invalid, and that it was simply a nuisance suit aimed specifically at the firm at a time when they were most vulnerable—precisely when the executives were trying to convince investors in their IPO "road show" that the company offered a solid investment opportunity.[210]

Anecdotes like this one offer a possible explanation for the pattern that we find. Specifically, that one-tenth of venture-backed IT hardware firms may be paying royalties for a patent license that provides no beneficial knowledge or information raises questions about the operation of the "market for patents" as a knowledge exchange mechanism. We surmise that the higher share of licensing merely to avoid suits among IT firms may be the consequence of several influences: a combination of the type of innovation and the nature of the firm's funding.

It is well accepted that information technologies are different from biochemistry innovations.[211] While the latter are termed "discrete" technologies for which a single patent can often adequately protect the entire invention, the former are characterized as "complex" technologies in which many separately patentable inventions are commonly needed to commercialize a product.[212] Complex technologies have been theorized to increase the transaction costs associated with commercialization,[213] and the problems associated with clearing patent rights are well-documented.[214] There

209. This difference between venture-backed and D&B IT company responses is significant at the 95% confidence interval. Statistically, there is no difference between the biotechnology results for the D&B and venture-backed firms (8% and 4%, respectively).

210. Interview with anonymous executive (Apr. 2009).

211. *See* Cohen et al., *supra* note 24, at 19. *But cf.* Arti Rai & James Boyle, *Synthetic Biology: Caught Between Property Rights, the Public Domain, and the Commons*, 5 PLOS BIOLOGY 389 (2007), *available at* http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050058 (discussing a convergence of these technologies).

212. Cohen et al., *supra* note 24, at 19.

213. *See* Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCI. 698, 699 (1998).

214. *See* JAMES BESSEN & MICHAEL J. MEURER, PATENT FAILURE: HOW JUDGES, BUREAUCRATS, AND LAWYERS PUT INNOVATORS AT RISK 69–70 (2008).

are also claims from engineers that, at least in the software fields, many of the patents issued since the 1980s have been on trivial inventions. Patented technology considered trivial coupled with the complex nature of the technology (and a liberal injunction rule) lends itself to hold-up and could result in the pattern we observe.

But the "venture funded" characteristic of these firms also appears to be playing a role in this difference. We note that the types of software companies we surveyed from D&B are able to operate generally free of the threats that appear to be at the heart of our finding. The venture-funding event is an important one in the life of a startup, and is often publicized—commonly by the startup itself because it provides a positive signal of quality to other investors, competitors, and consumers alike. Moreover, commercial research firms (such as Thomson) collect and report on these investments. Therefore, a startup's venture-funding event is news, and it may raise the profile of the company sufficiently to make it a target for holdup of the type we describe above. Companies that are venture funded are also more likely to have other significant liquidity events like IPOs or high-value acquisitions and mergers: these too may increase the potential pay-off to patent holders external to the firm when threatening these startups with infringement. While this hypothesis is intriguing, a fuller understanding must await further research.

B. CHECKING THE PATENT LITERATURE BY STARTUPS: MIXED FINDINGS

Another little-understood aspect of operating in an environment with patents concerns the propensity and timing of patent searching by innovators. When we consider the goals of the innovator and society, there are substantial benefits that may flow from searching the patent literature. Because innovation is often cumulative, the cost of innovation for both the innovator and society may be reduced when greater information about the current state of the technology is easily discoverable by the innovator. Moreover, to the extent that the underlying subject matter is a complex technology that is "modular," searching the patent literature may allow the innovator to find existing inputs to its ongoing creative process, thus preventing duplicate innovation that tends to squander society's resources.[215] By searching the patent literature, innovators may also avoid the risk of investing in R&D ex ante only to discover that they cannot practice the inventions ex post due to existing and blocking patent rights.

215. We speak here of duplicate efforts on the same composition, method, or process—it may be that different means could be used to reach the same outcome.

While these benefits suggest that searching the patent literature may be efficient and desirable, it has been suggested that entities may actually shy away from conducting searches.[216] Specifically, the patent system may have unintentionally created disincentives to patent searching, because doing so can expose the searcher to the risk of liability for willful infringement.[217] Under the existing patent laws, entities that are aware of existing patent rights and engage in willful infringement can suffer treble damages.[218] This rule has led to advice from attorneys that companies should not do searches at all. (However, the continuing viability of such advice has been cast into doubt by recent case law.[219]) Moreover, searches may be expensive in both money and time, and in new and embryonic technology areas, or ones that are very general in application, there may be added difficulty in discovering the proper scope for patent searching.[220]

Accordingly, we were interested in discovering which way these conflicting incentives might cut. We therefore inquired whether our respondents' companies regularly check the patent literature to determine if someone else has a U.S. patent that covered what they were doing or were considering. For those respondents answering "yes," we also asked at what stage of commercialization the company usually conducts that search: prior to design; during design and development; after launch; or (notably out of temporal order) only when the firm was planning to apply for a patent.

A substantial share of the respondents to this question reported regularly doing patent searches. Among D&B respondents who answered, slightly more than one-third reported conducting these searches. This likelihood was particularly high for biotechnology (nearly seven in ten) and medical device (over half) companies, while slightly less than one-quarter of software companies reported doing regular patent searches.

Among the venture-backed sample, searching was substantially more common. Among all the respondents to this question, nearly six in ten venture-backed firms reported that they regularly searched the patent literature. Again, this propensity was particularly high among biotechnology (nearly nine in ten) and medical device (over nine in ten) firms. Nevertheless, nearly three in ten venture-backed software startups and over six in ten similarly funded IT hardware companies reported doing so. This finding is

216. *See* BESSEN & MEURER, *supra* note 214, at 70.

217. *See id.*

218. *See* 35 U.S.C. § 284 (2006).

219. *See In re* Seagate Tech., L.L.C., 497 F.3d 1360 (Fed. Cir. 2007) (en banc) (adopting a rule making it more difficult for plaintiffs to show "willful infringement").

220. *See* BESSEN & MEURER, *supra* note 214, at 69–72.

notable, because it casts doubt on previous anecdotal evidence that venture-backed software companies, by and large, do not perform patent searches.[221]

Our findings on the timing of searches are also noteworthy given that companies may face conflicting incentives. Because these searches are costly (both in terms of money and technology employees' time), and given that these small startups are resource constrained (both in terms of money and time), it would stand to reason that at the margin, companies would tend to delay these searches, if not put them off altogether. Conversely, to the extent that effective patent search is an information input into the innovation and technology planning processes, early search might yield greater benefits, especially to *technology* startups for which technical information is critical for both ongoing innovation and effective competition.

We find that those startups that do patent searches tend to conduct them relatively early in the commercialization process. Among the D&B population of companies, 65% report usually doing searches prior to product or process design, and 70% report that these searches are usually done during design and development.[222] We may have also uncovered an element of patent strategy associated with this searching among our respondents: 14% report usually conducting examinations only when the company is planning to apply for its own patent (an event that may occur early or late in the commercialization process).

It is also noteworthy that one-third of the companies that conduct searches report usually doing so only after commercial launch. This result raises the possibility, consistent with our finding on patent licensing, that it is not uncommon for threats or other competitive information to arrive only after product launch. Since the company's technology is likely to be most notorious after product launch, it is conceivable that the company's competitors, or those interested in collecting royalties from the company, would be most active at that point. It is also possible that some companies

221. *See* Mann, *supra* note 92, at 1004. Mann states,

> For one thing, none of the startup firms to which I spoke suggested a practice of doing prior art searches before beginning development of their products. As far as I can tell, the only occasion in the software industry in which a startup is likely to experience such costs is when the startup is founded on a specific piece of existing technology spun off from an existing company or from a university laboratory.

Id.

222. These percentages can sum to more than 100% because respondents were allowed to mark more than one category when identifying at which stage the company "usually" conducted the patent search.

put off costly searching until they are more certain of the economic value of a technology—a certainty that may only come to light after product launch.

A similar pattern surfaces for the respondents in our venture-backed sample. For these companies, nearly six in ten relate usually doing searches prior to product or process design, with three-quarters reporting that these searches are usually done during design and development. Like the D&B companies, one-third note that the company usually searches the patent literature only after commercial launch, while nearly one-quarter report that they usually conduct examinations only when planning to apply for their own patents. But, because venture-backed firms are more likely to hold patents,[223] it may be that this increased share influences the higher likelihood among venture-backed firms to do searching at all when compared with the D&B respondents, fewer of which hold patents. That there may be a "patent feedback mechanism"—with startups engaged in patenting also being more likely to examine the patent literature—is an interesting possible finding of our study deserving further research.

VIII. IS THE PATENT SYSTEM WORKING FOR U.S. TECHNOLOGY STARTUPS?

Our evidence suggests that for startup companies, by and large, the patent system is neither working well, nor poorly. We asked our respondents one purely attitudinal question, without a specific definition: "Overall, how well is the U.S. patent system working . . . for your company," and "for your industry?" We offered the respondents five options, including "very well," "well," "neither well nor poorly," "poorly," and "very poorly." We had few expectations about what answers this question would yield, although we hypothesized that given the ongoing debate over the utility and validity of many software patents that we would see answers from software companies skewed toward the poorly end of the scale. Moreover, we were particularly interested to see whether the responses from startups would exhibit an essentially "U-shaped" distribution, with answers populating disproportionately the polar ends of the scale, with many answering "very well" and others answering "very poorly," with few in the center. Such a bi-modal distribution of answers would raise the possibility that our respondents were self-selecting into our survey based on how passionate they felt about the patent system, whether they had a "love" or a "hate" attitude toward it.

223. *See supra* Part III.

In fact, we find that the responses to both questions essentially took a classic bell-shaped form, with a mean and mode around the central answer. This result suggests that executives in early-stage technology startups in the United States find that the patent system, whether they are considering their company or their industry, is working neither well nor poorly. In essence, they appear in general to believe that the patent system is "muddling through."

There are some inter-industry differences in these responses, but the deviations are not as great as we surmised they would be at the project's inception. For instance, over 35% of respondents from the biotechnology industry answered that the patent system is working "well" or "very well" for their company, but the most common answer (the mode of the distribution) is still the central answer, "neither well nor poorly." Among software companies, the most common answer (the modal answer) is again the central answer, with over 55% of the respondents selecting this middle choice.[224]

Surprising to us was the opinion among these startups that the patent system is generally working less well for their industries than for their companies. Among the biotechnology companies for example, the pattern for "how well is the patent system working for your industry" is the mirror image of the answers to the "company" question. The mode was the same (the central answer), but over 35% of the respondents said they believe the patent system is working "poorly" or "very poorly" for their industry. Given that patenting in the "discrete" technology of biotechnology is usually held out as an example of how well the U.S. patent system is operating, this shared opinion among these biotechnology executives raises more questions than it answers. Software and Internet executives' answers are virtually identical on the "industry" question to those of the biotechnology responses, with the most common answer in the center of the choices, but over 35% of respondents indicating that the patent system is working "poorly" or "very poorly" for their industry.

Therefore, executives from both the D&B sample (essentially drawn from the population of early-stage technology companies in the U.S.) and the *VentureXpert* sample of high-quality, venture-backed firms are saying, at best, that the patent system is neither working poorly nor well. At worst, they

224. It is worth speculating whether these respondents were prone to the "central answer" bias reported elsewhere in the literature on surveys, although this same phenomenon did not seem to dominate their answers on other questions. *See generally* Eric A. Greenleaf, *Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles*, 29 J. MARKETING RES. 176 (1992) (finding bias when using certain configurations of response scales).

sense it is working poorly for their companies and their competitors. Such a finding does not bode well for our system. In the opinion of this "entrepreneurial class"—arguably one of the key drivers of innovation in the United States—the patent system does not appear to be functioning particularly well.

IX. CONCLUSION

By conducting a reasonably comprehensive survey on the relationship of the U.S. patent system to technology entrepreneurship, we have attempted to fill an important gap in the considerable body of work exploring intellectual property and more generally, innovation. Notably, our detailed results do not offer simple answers to this inquiry. For instance, we discover that technology startups are generally more likely to file for patents, as well as hold greater numbers of patents, than was previously believed. But we also reveal that these same companies report that patents provide mixed to relatively weak incentives for core innovative activities, such as invention, development, and commercialization. Part of the foregoing analysis attempts to uncover the motives for technology entrepreneurs acquiring and filing patents when those patents—according to the judgment of those same entrepreneurs—do not offer particularly strong innovation incentives.

Our response to the questions raised by these discordant streams has been both nuanced and multifaceted. We report that a large share of startups, especially in the software industry, opt out of patenting altogether. Although startups appear to be more aggressively accessing the patent system than previously reported, we find that these effects reflect our survey's ability to measure difficult-to-capture patent holdings. Specifically, our survey includes the reported number of patents originating from founders, acquisitions, and filed patent applications, rather than merely the number of patents recorded for a given firm name at the Patent Office. As such, we do not believe our larger numbers reflect an upward shift in filing rates among startups in the past several years.

We also report that for many startup companies, patents are an important part of the mix of strategies used by them to capture competitive advantage from their technology innovations. But this important role tends to be much more pronounced among biotechnology and "hardware" companies (including both medical hardware such as surgical devices, and IT hardware, such as computers and semiconductors) than among software and Internet startups. In fact, we find that for software and Internet companies, patents generally serve a much less important function in almost all of the entrepreneurial activities about which we surveyed.

These disparate findings trumpet a major result of our study: the industry in which a startup operates tends to exert a strong influence on the role that patents play in the firm's entrepreneurial activities. This finding may be driven by underlying technology differences in these sectors, variation in industrial organization, or other unseen factors. The deep differences we find in the use and utility of patents by startups across industries tracks many of the patent reform debates of recent years, and while some of the understanding that flows from those debates is useful in explaining our results, we find that startups can be quite different from their larger counterparts in their patenting behavior. As such, more research into the drivers of these industry differences at the level of the startup is needed.

This Article also demonstrates profound differences in the manner and extent to which patents are used by venture-backed companies. In a related series of findings, we report that patents are useful to startups in attracting entrepreneurial capital and for improving the likelihood of a successful exit event (such as being acquired or going public). In an important new showing, we demonstrate that patenting may play a previously underappreciated and important role in helping startups to secure investment from various sources, including "friends and family" and commercial banks, as well as angel investors and venture capital—although these results, too, are context-specific. These findings may partly explain why venture-backed startups are generally more apt to use patents, and consider them important, since our evidence suggests that these companies have been "selected" upon by VC investors for their patents. But here too, more study is needed to uncover the dynamics of the patent-financing relationship in technology entrepreneurship.

Our survey also reports why technology entrepreneurs use the patent system. Contrary to some previous anecdotal accounts, those firms that file for patents report that they do so primarily to prevent others from copying their products and services. Additionally, our study is the first to suggest that startups use patents in strategic ways, such as to improve bargaining positions in cross-licensing deals and defend against infringement suits. This finding undermines the previously held belief that the strategic uses of patents are significant motivations only for larger-firm patenting.

Finally, we find that our respondents believe, by and large, that the patent system is not working particularly well for their companies or their industries, even in the fields of biotechnology, medical devices, and computer hardware, where our survey suggests patents are considered to be more useful for a range of reasons. Because many studies have shown that entrepreneurs and their firms play a substantial role in driving innovation, employment, and

economic growth, we strongly believe that further study is warranted to determine how the system might be reformed to better serve the needs of startups.

In the meantime, following on this Article, we plan to further explore our dataset in a series of more focused articles. These foci include the drivers of patenting by entrepreneurs, patenting among software and Internet startups, patents in entrepreneurial "markets for technology," the effects of patents in allowing startups to enter new markets, and the relationship of patenting to market valuation and exit, among others. In sum, we plan to use our dataset to study a host of topics shedding additional light upon the role that patenting is playing in technology entrepreneurship.

We thank Robert Barr, Executive Director of the Berkeley Center for Law and Technology (BCLT), for his consistently helpful comments and unflagging support in making this survey project a success. We received beneficial insights from our project's Advisory Board (Hank Barry, John Barton, James Bessen, Tom Ciotti, Wesley Cohen, John Duffy, Rebecca Eisenberg, Brad Feld, John Freeman, Richard Gilbert, Michael Goldberg, Bob Gunderson, Bronwyn Hall, Mitchell Kapor, Peter Menell, Mark Myers, James Pooley, Walter Powell, Arti Rai, AnnaLee Saxenian, Carl Shapiro, Robert Strom, Lee Van Pelt, and David Yoffie), and from Peter Ang, Brian Broughman, Colleen Chien, Jesse Fried, Dietmar Harhoff, Bob Litan, Gladys Monroe, Bob O'Connor, Sheridan Paulker, E.J. Reedy, Alicia Robb, F.M. Scherer, and John Walsh. We also thank participants at the 2008 Berkeley Technology Law Journal Symposium on Patenting and Entrepreneurship, the 2008 Kauffman Foundation Symposium on Entrepreneurship and Innovation Data, the 2008 OECD-DIME Patent Statistics for Managers Conference, the 2009 OECD-Kauffman Workshop on Entrepreneurship and Innovation, the National Bureau of Economic Research's Summer Institute 2009, the 2009 Conference on Empirical Legal Studies, the 2009 IP Scholars Conference, the 2009 Creativity, Law, and Entrepreneurship Workshop at the University of Wisconsin Law School, the 2009 Kauffman Foundation Summer Institute, the 2009 Northwestern Law School Searle Center Conference on the Law and Economics of the Entrepreneur, the 2009 West Coast Research Symposium, the 2009 Works in Progress in Intellectual Property Conference, and faculty workshop participants at the Bar-Ilan School of Law, DePaul Law School, Georgia Tech College of Management, Harvard Business School, Stanford University Economics Department, and the University of California, Berkeley, Haas School of Business and School of Law. We thank Galen Hancock for working closely with us as our chief research assistant since the inception of the project, and David Grady and Louise Lee for administrative support. We acknowledge helpful advice from Bob Lee and UC Berkeley's Survey Research Center, and research assistance from Peter Dimitrov, Andrew Harris, Deepak Hegde, Christine Rhyu, Erin Trimble, Stephen Ullmer, and Brian Weissenberg. This research was funded by the Ewing Marion Kauffman Foundation. The contents of this article are solely the responsibility of the grantees/authors. All errors and omissions are our own.

CONTROLLING THE "PLAGUE": REFORMING THE DOCTRINE OF INEQUITABLE CONDUCT

Christian E. Mammen[†]

ABSTRACT

Following its characterization of inequitable conduct as a "plague" in 1988, the Federal Circuit took steps to narrow the defense, and the United States Patent and Trademark Office (PTO) amended its rules to more clearly define the duty of disclosure. However, proliferation of the defense has proved difficult to control. In recent years, the Federal Circuit has issued a number of decisions that expand the defense, leading one judge, in dissent, to recently proclaim that the court had "return[ed] to the 'plague' of encouraging unwarranted charges of inequitable conduct." This recent expansion, in turn, has fueled a series of patent reform proposals. Reform of the inequitable conduct doctrine is necessary to rein in the assertion of inequitable conduct as a litigation tactic, and to stem a growing tide of inequitable conduct cases and establish an appropriate and sensible role for the doctrine within patent litigation. This Article advocates four reforms to the doctrine of inequitable conduct:

- Materiality should be explicitly linked to the PTO's 1992 definition of materiality in its regulations;
- The standards for proving intent should be clearly defined;
- The step of "balancing" materiality and intent should be clarified and codified, making clear that even if thresholds of materiality and intent exist, the court retains equitable discretion to decline to find inequitable conduct;
- There should be a single, narrowed remedy. Only the claims directly affected by the inequitable conduct (rather than the entire patent) should be deemed unenforceable.

---

† Resident Scholar, University of California, Hastings College of the Law; Law Office of Christian Mammen (www.mammenlaw.com); J.D. *magna cum laude*, Cornell Law School; D.Phil. (Law), Oxford University. The author wishes to thank the editors of the Berkeley Technology Law Journal, Stephen Kunin, Mark Lemley, Michael Vitiello, Lloyd Jobe, and Andy Chan for their valuable contributions and insights.

TABLE OF CONTENTS

I. INTRODUCTION

Twenty years ago, the Federal Circuit expressed its displeasure with the proliferation of inequitable conduct claims in patent litigation, famously calling it a "plague":

> [T]he habit of charging inequitable conduct in almost every major patent case has become an absolute plague. Reputable lawyers seem to feel compelled to make the charge against other reputable lawyers on the slenderest grounds, to represent their client's interests adequately, perhaps. They get anywhere with the accusation in but a small percentage of the cases, but such charges are not inconsequential on that account.[1]

Starting with the *Burlington* and *Kingsdown* decisions in 1988, the Federal Circuit took steps to narrow the defense, and the United States Patent and Trademark Office (PTO or USPTO) subsequently amended its rules to define more clearly the duty of disclosure. However, proliferation of the defense has proven difficult to control.

In recent years, the Federal Circuit has issued a number of decisions that have arguably expanded the defense, while also clouding its boundaries. For example, in one recent case, the trial court granted summary judgment of inequitable conduct on the basis that third-party experts who submitted declarations during prosecution failed to disclose material information when those declarations omitted past financial ties with the patentee.[2] The court held the experts had intended to deceive the PTO because they "should have known" that such financial ties were highly material, even without any suggestion that the scientific information they provided was inaccurate or misleading.[3] The doctrine's expansion led one judge to proclaim in dissent that the court had "return[ed] to the 'plague' of encouraging unwarranted charges of inequitable conduct . . . ."[4] This recent expansion has also fueled a series of patent reform proposals.[5]

This Article argues that reform of the inequitable conduct doctrine is necessary. Although there has been a recent, renewed trend toward its expan-

---

1. Burlington Indus., Inc. v. Dayco Corp., 849 F.2d 1418, 1422 (Fed. Cir. 1988); *see also* Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867, 876 n.15 (Fed. Cir. 1988) (citing *Burlington* and alluding to the "present proliferation of inequitable conduct charges").

2. Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181 (Fed. Cir. 2006).

3. *Id.* at 1192.

4. McKesson Info. Solutions, Inc. v. Bridge Med., Inc., 487 F.3d 897, 926 (Fed. Cir. 2007) (Newman, J., dissenting).

5. The recent round of patent reform efforts was launched in 2003–2004 with the release of two major reports, one from the Federal Trade Commission and one from the National Academy of Sciences. FEDERAL TRADE COMMISSION, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW AND POLICY (2003), *available at* http://www.ftc.gov/os/2003/10/innovationrpt.pdf [hereinafter FTC REPORT]; COMM. ON INTELLECTUAL PROP. RIGHTS IN THE KNOWLEDGE-BASED ECON., NAT'L RESEARCH COUNCIL, A PATENT SYSTEM FOR THE 21ST CENTURY (Stephen, A. Merrill, Richard C. Levin & Mark B. Myers eds., 2004) [hereinafter NAS REPORT].

sion, the rate at which inequitable conduct is found in Federal Circuit case law has remained somewhat below the level that existed just prior to *Burlington* and *Kingsdown*. Nonetheless, this Article argues that some critical clarifications and restrictions on the doctrine are necessary to rein in the assertion of inequitable conduct as a litigation tactic, to stem a growing tide of inequitable conduct cases, and to establish an appropriate and sensible role for the doctrine within patent litigation.

As one commentator noted twenty years ago, "[t]he strategic and technical advantages that the inequitable conduct defense offers the accused infringer make it almost too attractive to ignore."[6] One advantage is the possibility of a broad remedy—a finding of inequitable conduct will render the entire patent unenforceable, rather than just the particular affected claims. In certain cases, related patents may be held unenforceable as well. A second advantage is an asymmetrical discovery burden that exerts pressure on the boundaries of the attorney-client privilege, since most relevant documents will come from the files of the inventor and the patent attorney who prosecuted the patent, and those individuals will likely be subject to deposition. A third advantage is the opportunity that inequitable conduct provides to impugn the character of the inventor and her counsel, providing a counterbalance to the patentee's likely narrative at trial of the inventor as an idealized genius.

With those advantages, together with an increasing murkiness in the elements and boundaries of the defense, it is little wonder that accused infringers look for any opportunity to inject the inequitable conduct defense into patent litigation, and are doing so with increasing frequency.

Under the basic doctrine, a prima facie claim of inequitable conduct comprises three elements:

1. an affirmative misrepresentation of material fact, a submission of false material information, or a failure to disclose material information;

2. an intent to deceive the Patent Office; and

3. an equitable evaluation, or "balancing," of materiality and intent to determine whether the conduct is sufficiently culpable to warrant a finding of unenforceability.[7]

---

6. John F. Lynch, *An Argument for Eliminating the Defense of Patent Unenforceability Based on Inequitable Conduct*, 16 AIPLA Q.J. 7, 8 (1988).

7. *E.g.*, Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1313 (Fed. Cir.

If these elements are satisfied, the entire patent is rendered unenforceable.[8] The Federal Circuit has, however, been inconsistent and, recently, increasingly broad in how it articulates and applies those broad-brush elements.

To rectify the problems associated with the doctrine's expansion and to reduce the tactical incentives to over-use the defense, this Article advocates four reforms to the doctrine of inequitable conduct:

1. Explicitly link the element of materiality to the definition of materiality in the 1992 revision of the PTO's regulations.

2. Define clear standards for proving intent.

3. Clarify and codify the step of "balancing" materiality and intent, making clear that even if thresholds of materiality and intent exist, the court retains equitable discretion to decline to find inequitable conduct.

4. Create a single, narrowed remedy. Only the claims directly affected by the inequitable conduct (rather than the entire patent) should be deemed unenforceable.

Part II of this Article explains each of the prima facie elements of inequitable conduct. Part III examines the characterization of inequitable conduct claims as a "plague," which connotes two key aspects: nuisance and prevalence. By analyzing data relating to the more than 300 Federal Circuit cases since 1982 that address inequitable conduct, together with selected data from the district court level, this article examines the trends in the case law, and identifies some benchmarks against which the "plague" characterization can be measured. Part IV critically examines a number of recent inequitable conduct cases from the Federal Circuit to identify several key problems leading to ambiguity and expansion of the doctrine. Part V evaluates several recent legislative reform proposals, identifying the strengths and weaknesses of each. Part VI sets out proposals for substantive reform.

II. BACKGROUND: THE LAW OF INEQUITABLE CONDUCT

Inequitable conduct is a judicially created defense to patent infringement,[9] having its origins in the equitable doctrine of unclean hands.[10] That is,

---

2006).

8. *Id.*

9. *Id.* at 1315.

10. *See* Robert J. Goldman, *Evolution of the Inequitable Conduct Defense in Patent Litigation*, 7

a patentee seeking to enforce its patent rights must not come before the court with unclean hands due to his intentional misleading of the PTO in order to obtain the patent.

Under modern doctrine, to prove that a patent is unenforceable due to inequitable conduct, a party must show that an inventor, an inventor's attorney, or another person substantively involved with the application,[11] with intent to mislead or deceive the Patent Office, failed to disclose to the Patent Office material, non-cumulative information known to that person to be material, or submitted materially false information to the Patent Office during prosecution.[12] If the Court determines that the threshold levels of both materiality and intent were achieved, then the Court must balance materiality and intent, "with a greater showing of one factor allowing a lesser showing of the other."[13]

The following Sections address the three elements of materiality, intent, and balancing in greater detail.

A. MATERIALITY

Although the case law concerning inequitable conduct has articulated a number of different standards for materiality,[14] Rule 56 of the Rules of Practice in Patent Cases has long guided the determination of the materiality prong of the inequitable conduct inquiry.[15] From 1977 until 1992, Rule 56 defined materiality in terms of a "reasonable examiner" test, and that was the dominant test throughout that period. In 1992, Rule 56 was amended to provide a more objective test of materiality. Although the 1992 version of Rule

---

HARV. J.L. & TECH. 37, 49–50 (1993); *see also* S. REP. NO. 110-259, at 59 (2008) (citing Keystone Driller Co. v. Gen. Excavator Co., 290 U.S. 240, 244 (1933)).

11. 37 C.F.R. § 1.56 (2008).

12. *Digital Control*, 437 F.3d at 1313; Norian Corp. v. Stryker Corp., 363 F.3d 1321, 1330–31 (Fed. Cir. 2004); 37 C.F.R. § 1.56.

13. *Digital Control*, 437 F.3d at 1313 (citing Union Pac. Res. Co. v. Chesapeake Energy Corp., 236 F.3d 684, 693 (Fed. Cir. 2001)).

14. *See id.* at 1315.

15. 37 C.F.R. § 1.56. Rule 56, also characterized as the "duty of disclosure" requirement, is intended to improve the quality of examination and the validity of patents. *See* Rene D. Tegtmeyer, *The Patent and Trademark Office View of Inequitable Conduct or Attempted Fraud in the Patent and Trademark Office*, 16 AIPLA Q.J. 88, 88 (1988) (noting that Rule 56 is intended "to improve the quality of examination and the validity of patents"). *See also* Am. Hoist & Derrick Co. v. Sowa & Sons, Inc., 725 F.2d 1350, 1362–63 (Fed. Cir. 1984) (describing four standards for determining materiality and holding that "[t]he PTO 'standard' [i.e., Rule 56] is an appropriate starting point for any discussion of materiality, for it appears to be the broadest, thus encompassing the others, and because that materiality boundary most closely aligns with how one ought to conduct business with the PTO"); Christopher A. Cotropia, *Modernizing Patent Law's Inequitable Conduct Doctrine*, 24 BERKELEY TECH. L.J. 723, 733 (2009).

56 is still in effect, the courts have apparently been reluctant to adopt this change, as explained in greater detail below.

When the Federal Circuit first declared a "plague" in 1988, the operative portion of the 1977 version of Rule 56 defined information as material "where there is a substantial likelihood that a reasonable examiner would consider it important in deciding whether to allow the application to issue as a patent."[16] Perhaps unsurprisingly, the courts and the PTO provide differing accounts of the origin of the "reasonable examiner" standard. The Tenth Circuit (in a pre-Federal Circuit case) asserts that the PTO merely codified existing case law.[17] According to the Federal Circuit, the 1977 version of Rule 56 was "a codification of earlier case law."[18] However, the PTO's commentary from the 1977 enactment indicated that Rule 56 as a whole "codifies the existing Office policy on fraud and inequitable conduct, which is believed consistent with the prevailing case law in the federal courts. . . . [T]he section should have a stabilizing effect on future decisions in the Office, and may afford guidance to courts as well."[19] Additionally, the PTO's commentary noted that the "reasonable examiner" definition of materiality was paraphrased from a Supreme Court decision relating to Securities and Exchange Commission rules, but that nonetheless the PTO believed it to be "consistent with the prevailing concept that has been applied by lower courts in recent patent cases."[20]

The PTO identified problems with the "reasonable examiner" standard, including that it was insufficiently objective, unworkable,[21] and too impre-

16. 37 C.F.R. § 1.56 (1977).

17. True Temper Corp. v. CF&I Steel Corp., 601 F.2d 495, 504 n.9 (10th Cir. 1979) (refusing to apply Rule 56 as alternative test for materiality because it "merely represented a codification of existing case law"); *see also Digital Control*, 437 F.3d at 1315.

18. Connell v. Sears, Roebuck & Co., 722 F.2d 1542, 1551 (Fed. Cir. 1983) (citing *True Temper*, 601 F.2d 495).

19. Rules of Practice in Patent Cases, 42 Fed. Reg. 5588, 5589 (Jan. 28, 1977).

20. *Id. Contra* Lynch, *supra* note 6, at 13–15 (arguing that the 1977 version of Rule 56 was not based on existing inequitable conduct case law).

21. Duty of Disclosure and Practitioner Misconduct, 54 Fed. Reg. 11,334 (proposed Mar. 17, 1989) ("These proposed changes are considered desirable in view of the large amount of resources that are being devoted to duty of disclosure issues both within and outside the Office without significantly contributing to the reliability of the patents being issued.").

cise.[22] The PTO also characterized the standard as vague and not correlated to other areas of the patent law.[23]

Accordingly, in 1989, the year after *Burlington* and *Kingsdown* declared a "plague" of inequitable conduct, the PTO proposed amendments to Rule 56 seeking to replace the reasonable examiner standard with a clearer and more objective set of guidelines.[24] In 1992, the PTO adopted a revised version of Rule 56, which largely remains in place today.[25] These revisions were intended to "specify more precisely the information" that should be disclosed to the PTO during prosecution.[26] Moreover, the determination what a reasonable examiner would have considered important is entirely a hypothetical determination left to the courts—testimony by actual examiners as to what they considered (or would have considered) important is prohibited.[27]

The 1992 version of Rule 56 defines material information subject to the duty of disclosure as information that either (1) establishes (alone or in combination with other information) a prima facie case of unpatentability of a claim, or (2) refutes or is inconsistent with a position the applicant took in arguing for patentability or in opposing an argument of unpatentability relied on by the PTO.[28] Additionally, information is material only if it is not cumu-

22. *E.g.*, Harry F. Manbeck, Jr., *Evolution and Future of New Rule 56 and the Duty of Candor: The Evolution and Issue of New Rule 56*, 20 AIPLA Q.J. 136, 139–40 (1992) ("[In 1990–1991] I concluded that existing Rule 56 was indeed too imprecise, and could, and probably was, leading to unjustifiable charges of inequitable conduct in litigation. It should be changed.").

23. Duty of Disclosure, 56 Fed. Reg. 37,321, 37,322 (proposed Aug. 6, 1991) (codified at 37 C.F.R. § 1.56).

24. 56 Fed. Reg. at 37,321, 37,322.

25. *See* 56 Fed. Reg. at 37,321; Duty of Disclosure, 57 Fed. Reg. 2021 (Jan. 17, 1992).

26. 56 Fed. Reg. at 37,321.

27. 37 C.F.R. § 104.22 (2008) (prohibiting testimony by PTO employees without General Counsel's approval); 37 C.F.R. § 104.23 (2008) (prohibiting expert or opinion testimony by PTO employees without General Counsel's approval); U.S. PATENT & TRADEMARK OFFICE, U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 1701 (8th ed., rev. 7, 2008) (prohibiting testimony and opinions concerning, inter alia, patent enforceability); W. Elec. Co. v. Piezo Tech., Inc., 860 F.2d 428, 430, 433 (Fed. Cir. 1988) (reasonable examiner test "is an objective standard," so testimony of examiners who examined patent-in-suit "is irrelevant to the issue of inequitable conduct"); *id.* at 431 (noting "the general rule . . . that a patent examiner cannot be compelled to testify regarding his 'mental processes' in reaching a decision on a patent application").

28. 37 C.F.R. § 1.56(b) (1992). *See also* Rene D. Tegtmeyer, *Evolution and Future of New Rule 56 and the Duty of Candor: A Refocusing on Inequitable Conduct in New Rule 56*, 20 AIPLA Q.J. 191, 194 (1992) (noting that new Rule 56 "recognizes to some degree the unnecessary problems and expenses that are caused when questions of inequitable conduct arise in litigation based on allegedly withheld or misrepresented information not affecting patentability").

lative of information that is already before the PTO as part of the patent application.[29]

For a number of years after the adoption of the 1992 version of Rule 56, and as late as the *Purdue Pharma* decision on February 1, 2006, the Federal Circuit consistently held that the "reasonable examiner" test applied to patents prosecuted before 1992, and the new version of Rule 56 applied to applications pending or filed after the rule's March 16, 1992 effective date.[30] These cases imply that the "reasonable examiner" test would gradually fade into irrelevance as the last of the pre-1992 patents expired.

However, on February 8, 2006, just a week after it decided *Purdue Pharma*, the Federal Circuit decided *Digital Control Inc. v. Charles Machine Works*,[31] which breathed new life into the "reasonable examiner" test for patent applications pending or filed after March 16, 1992.[32] *Digital Control* reached back to the 1984 *American Hoist* case,[33] decided four years before *Burlington* and *Kingsdown*, to revive a list of four historically accepted and judicially adopted standards of materiality.[34] The *Digital Control* court reasoned that the 1992 version of Rule 56 was "not intended to replace or supplant the 'reasonable examiner' standard," and that the "reasonable examiner" standard should continue to exist as one of the tests for materiality.[35]

Some uncertainty, however, remained. A week after *Digital Control*, the Federal Circuit decided *Ferring B.V. v. Barr Laboratories, Inc.*,[36] which reiterated

29. 37 C.F.R. § 1.56(b) (1992).

30. *See* Purdue Pharma. L.P. v. Endo Pharms. Inc., 438 F.3d 1123, 1129 (Fed. Cir. 2006) ("Because all of the patent applications at issue in this case were pending on or filed after March 16, 1992, we look to the current version of Rule 56, rather than the pre-1992 version of the rule."); Bruno Indep. Living Aids, Inc. v. Acorn Mobility Servs. Ltd., 394 F.3d 1348, 1352–53 (Fed. Cir. 2005) ("According to the PTO's notice of final rulemaking, the rule change applied to all applications pending or filed after March 16, 1992.") (citation omitted); Dayco Prods., Inc. v. Total Containment, Inc., 329 F.3d 1358, 1364 (Fed. Cir. 2003) ("Since the time of the 1992 amendment we have continued to apply the reasonable examiner standard, *but only* as to cases that were prosecuted under the earlier version of Rule 56.") (emphasis added) (citations omitted).

31. 437 F.3d 1309 (Fed. Cir. 2006).

32. *Digital Control* addressed the issue of inequitable conduct as it related to three patents: U.S. Patent No. 5,767,678; U.S. Patent No. 6,008,651; and U.S. Patent No. 6,232,780. *Id.* at 1310. Each of these patents was based on applications filed after March 16, 1992 but all three could be traced back to a common ancestor application that was filed on March 1, 1991. *See* U.S. Patent No. 5,767,678 (filed Oct. 9, 1996); U.S. Patent No. 6,008,651 (filed Sept. 18, 1998); U.S. Patent No. 6,232,780 (filed Mar. 3, 2000). Thus, each of these three patent applications was filed or pending after March 16, 1992.

33. Am. Hoist & Derrick Co. v. Sowa & Sons, Inc., 725 F.2d 1350 (Fed. Cir. 1984).

34. *Digital Control*, 437 F.3d at 1315.

35. *Id.* at 1316.

36. 437 F.3d 1181 (Fed. Cir. 2006).

the position that the pre-1992 version of Rule 56 applied to patents prosecuted before the amendment.[37] The implication in *Ferring* was that the 1992 version of Rule 56 should apply to patents prosecuted after the 1992 amendment. By 2008, any such uncertainty appears to have dissipated, when the Federal Circuit again endorsed the "reasonable examiner" test as the controlling standard.[38]

B. INTENT

The intent prong of inequitable conduct has also posed significant challenges. There are virtually no cases in which there is direct evidence of intent, such as a smoking-gun document or flat-out witness admission that someone concealed information from the PTO for the purpose of misleading the examiner. Thus, courts have repeatedly stated that intent to deceive must often be proven by circumstantial evidence because direct evidence of intent is rarely, if ever, available.[39] But determining precisely what kind of evidence will suffice has been exceedingly difficult.

The *Kingsdown* and *Burlington* cases, decided at the height of the "plague," each address the intent issue. In both cases, the district court granted summary judgment of unenforceability due to inequitable conduct based on actions that may fairly be characterized as sloppy or imprecise work by the patent lawyer during prosecution. As explained in more detail below, the prosecuting attorney in *Burlington* failed to perceive the distinction between two phrases he used interchangeably;[40] and in *Kingsdown*, the prosecuting attorney copied the wrong version of a claim into a continuation application.[41]

In *Burlington*, the Federal Circuit noted that "the nondisclosure of facts of which the applicant *should have known* the materiality may justify an inference of intent to mislead in appropriate cases."[42] The district court had granted

---

37. *Id.* at 1187 n.6 ("[W]e have continued to use the pre-1992 language regarding materiality for evaluating patents that were prosecuted before the amendment.").

38. *See, e.g.*, Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357, 1367 (Fed. Cir. 2008) (reciting only the "reasonable examiner" standard for materiality).

39. *E.g.*, *Ferring*, 437 F.3d at 1191.

40. *See* Burlington Indus., Inc. v. Dayco Corp., 849 F.2d 1418 (Fed. Cir. 1988).

41. *See* Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867 (Fed. Cir. 1988).

42. *Burlington*, 849 F.2d at 1421 (emphasis added). The court also stated that, under 37 C.F.R. § 1.56(d), the PTO would not grant a patent when, inter alia, "the duty of disclosure was violated through bad faith or gross negligence." *Id.*

The Federal Circuit noted in *Burlington* that, in addition to any adjudications of inequitable conduct that the courts might engage in, the Patent Office also had procedures for determining inequitable conduct. *See* U.S. PATENT & TRADEMARK OFFICE, U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 2010 (5th ed., rev. 7, 1986);

summary judgment of unenforceability based on the patent's interchangeable usage of two phrases: one characterizing the invention as "impregnat[ion of] *individual fibers* in the yarn bundle," and the other characterizing the invention as "impregnat[ion of] the *fiber bundles* and encapsulat[ion of] the individual fibers."[43] The distinction allegedly made a difference as to whether the invention read onto two prior art patents. In response, the attorney who prosecuted the patents testified that he had not perceived a distinction between the two phrases, and that he had used one as "shorthand" for the other.[44] The Federal Circuit vacated the summary judgment and remanded, stating that it was error to reject the attorney's explanation out of hand in favor of "a less plausible sinister interpretation."[45] Although the Federal Circuit closed its opinion in *Burlington* with its now-famous paragraph about the "absolute plague" of inequitable conduct, it did not provide affirmative guidance on the standard of proof for intent, other than to note that failure to disclose information the party "should have known" to be material can form a basis for intent.

In *Kingsdown*, the Federal Circuit convened en banc to address whether "gross negligence" could support a finding of intent to deceive the Patent Office,[46] and concluded that "a finding that particular conduct amounts to 'gross negligence' does not of itself justify an inference of intent to deceive;

---

id. § 2020.03 ("As soon as an issue of 'fraud,' 'inequitable conduct' or 'violation of the duty of disclosure' is identified in, or with regard to, an application, the application should be forwarded to the Office of the Assistant Commissioner for Patents.").

Shortly thereafter, however, the PTO ceased making determinations of inequitable conduct. In revisions of the Manual of Patent Examining Procedure (MPEP) published between 1992 and 1995, sections 2010 and 2020 were deleted. Starting in 1995, the Patent Office explicitly stated its policy of not becoming involved in adjudicating issues of intent to deceive, due in part to its lack of expertise and resources to make the fact- and evidence-intensive assessments of intent. U.S. PATENT & TRADEMARK OFFICE, U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 2010 (6th ed. Jan. 1995) ("The Office is not the best forum in which to determine whether there was an 'intent to mislead' . . . . Accordingly, the Office does not investigate and reject original or reissue applications under 37 C.F.R. [§] 1.56."); *see also* Manbeck, *supra* note 22, at 139–140 (noting that the PTO disbanded the "fraud squad" in 1988); Tegtmeyer, *supra* note 28, at 193 (noting that 1992 version of Rule 56 confirms that PTO no longer investigates inequitable conduct issues).

43. *Burlington*, 849 F.2d at 1419 (emphasis added).

44. *Id.* at 1420.

45. *Id.* at 1421.

46. *See* 37 C.F.R. § 1.56(d) (1977) ("An application shall be stricken from the files if it is established by clear and convincing evidence that any fraud was practiced or attempted on the Office in connection with it or that there was any violation of the duty of disclosure through bad faith or *gross negligence*.") (emphasis added); Rules of Practice in Patent Cases, 42 Fed. Reg. 5588, 5594 (Jan. 28, 1977).

the involved conduct, viewed in light of all the evidence, including evidence indicative of good faith, must indicate sufficient culpability to *require* a finding of intent to deceive."[47] In so ruling, the Federal Circuit again reversed a grant of summary judgment of inequitable conduct, ruling that the district court's finding of intent was clearly erroneous. During prosecution, the examiner rejected as indefinite the application's claim 50. Following amendments to overcome the indefiniteness rejection, the examiner indicated that an amended claim 50 would be allowable.[48] Subsequently, a continuation application included the disallowed, pre-amendment version of claim 50, not the allowable, amended version.[49] Hollister, the accused infringer, alleged that this change was made with gross negligence or intentionally (e.g., in order to strengthen Kingsdown's infringement claim against Hollister), and therefore constituted inequitable conduct. The Federal Circuit disagreed, ruling that this error was "insufficient to warrant a finding of an intent to deceive the PTO."[50] The Federal Circuit emphasized that the error would have been an easy one to make.[51] In fact, the prosecuting attorney testified that he was unaware of the error until Hollister raised it.[52] The Federal Circuit also rejected Hollister's contentions that intent should be inferred from Kingsdown's actions, holding that there was nothing improper about trying to write otherwise patentable claims to cover a competitor's known product,[53] and that Kingsdown's failure to abandon its suit, or to file a disclaimer or reissue application, was irrelevant to the issue of intent.[54]

The *Burlington* and *Kingsdown* opinions differed in at least one key respect in their treatment of how intent can be proven. *Burlington* held that intent can be shown if the applicant knew of the undisclosed prior art and "should have known" of its materiality. However, *Kingsdown* held that even gross negligence cannot alone support a finding of intent. At the time of *Burlington* and *Kingsdown*, the "should have known" test and gross negligence were related, as explained in the 1984 *J.P. Stevens* case. In *J.P. Stevens*, the Federal Circuit held that gross negligence is sufficient to prove intent, and that gross negligence is

47. Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867, 876 (Fed. Cir. 1988) (en banc) (emphasis added).

48. *Id.* at 874–875.

49. *Id.* at 869–71.

50. *Id.* at 873.

51. *Id.*

52. *Id.* at 872.

53. *Id.* at 874.

54. *Id.* at 875–76. It was in response to this last issue that the Federal Circuit invoked *Burlington*'s "plague" rhetoric, stating that "[a] requirement for disclaimer or reissue to avoid adverse inferences would merely encourage the present proliferation of inequitable conduct charges." *Id.* at 876, 876 n.15.

shown where the applicant reasonably should have known of the materiality of the undisclosed information.[55]

Consistent with *J.P. Stevens*, *Burlington* held that intent could be inferred from the nondisclosure of facts that the applicant should have known were material.[56] The en banc Federal Circuit in *Kingsdown* then ruled that "a finding that particular conduct amounts to 'gross negligence' does not of itself justify an inference of intent to deceive; the involved conduct, viewed in light of all the evidence, including evidence indicative of good faith, must indicate sufficient culpability to require a finding of intent to deceive."[57]

Viewed through the prism of *J.P. Stevens*, it would have been reasonable to infer that *Kingsdown*'s holding also meant that intent could not be inferred under the "should have known" test. The Federal Circuit's 1990 decision in *Hoffmann-La Roche v. Lemmon*[58] supports such an inference. In that case, the district court ruled that during prosecution of a reissue application the applicant was "grossly negligent since he should have known of the materiality of the withheld information. The intent to deceive can be inferred from this gross negligence."[59] The Federal Circuit reversed because, under *Kingsdown*, gross negligence alone cannot support a finding of intent.[60]

In the 2001 *Brasseler* case, the Federal Circuit provided an indication that *Kingsdown* had not definitively resolved the status of the "should have known" test, holding that gross negligence in the avoidance of learning the materiality of information can support a finding that the applicant should have known of the materiality of that information.[61] The *Brasseler* ruling effectively inverted *J.P. Stevens*: The Federal Circuit used gross negligence to prove that the applicant should have known about the information's materiality and therefore intended to deceive the examiner. In contrast, the Federal Circuit in *J.P. Stevens* used the "should have known" test to prove gross negligence to support an inference of intent to deceive.

---

55. J.P. Stevens & Co. v. Lex Tex Ltd., 747 F.2d 1553, 1560 (Fed. Cir. 1984).
56. Burlington Indus., Inc. v. Dayco Corp., 849 F.2d 1418, 1421 (Fed. Cir. 1988).
57. *Kingsdown*, 863 F.2d at 876.
58. Hoffmann-La Roche Inc. v. Lemmon Co., 906 F.2d 684 (Fed. Cir. 1990).
59. *Id.* at 687 (citation omitted).
60. *Id.* at 688.
61. Brasseler, U.S.A. I, L.P. v. Stryker Sales Corp., 267 F.3d 1370, 1380 (Fed. Cir. 2001); Nordberg, Inc. v. Telsmith, Inc., 82 F.3d 394, 397 (Fed. Cir. 1996) ("[A]n applicant who knew of the art or information cannot intentionally avoid learning of its materiality through gross negligence, i.e., it may be found that the applicant 'should have known' of that materiality.") (citing FMC Corp. v. Manitowoc Co., 835 F.2d 1411, 1415 (Fed. Cir. 1987)).

Finally, in 2006, the Federal Circuit revitalized *Burlington*'s "should have known" test, sidestepping *Kingsdown*'s prohibition on proving intent through gross negligence. In *Ferring*,[62] the Federal Circuit held that in certain cases,

> [S]ummary judgment is appropriate on the issue of intent if there has been a failure to supply highly material information and if the summary judgment record establishes that (1) the applicant knew of the information; (2) the applicant knew or *should have known* of the materiality of the information; and (3) the applicant has not provided a credible explanation for the withholding.[63]

C. BALANCING

The third prong of the inequitable conduct inquiry enters into consideration once minimum thresholds of both materiality and intent have been satisfied. In such cases, the materiality and intent are equitably "balanced" to determine whether the misconduct is sufficiently culpable to support a holding of unenforceability.[64]

The term "balancing" can be somewhat confusing as a characterization of the manner in which the court evaluates materiality and intent. Typically, the term describes competing values, metaphorically arrayed at opposite ends of a simple lever, where the "weightier" value prevails over the lesser value. This ordinary conception might lead some to the conclusion that the balancing that is conducted for inequitable conduct has materiality and intent on opposite ends of the lever.

However, such a conception fails to account for the fact that a finding of inequitable conduct requires both materiality and intent, and that the existence of equally high amounts of both materiality and intent leads not to equipoise, but to a permissible finding of inequitable conduct. Moreover, a finding of high materiality and low intent on the one hand, and a finding of low materiality and high intent on the other hand, may both be sufficient to establish that there has been inequitable conduct. Under the simple-lever model, these alternative findings would lead to opposite or competing outcomes. A different depiction of the "balancing" test is therefore necessary.

In *Star Scientific, Inc. v. R.J. Reynolds Tobacco Co.*, the Federal Circuit provided a more detailed explanation of the manner in which materiality and in-

---

62. 437 F.3d 1181 (Fed. Cir. 2006).

63. *Id.* at 1191 (emphasis added). *See also* James E. Hanft & Stacey S. Kerns, *The Return of the Inequitable Conduct Plague: When "I Did Not Know" Unexpectedly Becomes "You Should Have Known"*, INTELL. PROP. & TECH. L.J., Feb. 2007, at 1, 3–4.

64. *E.g.*, Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1313 (Fed. Cir. 2006); Purdue Pharma L.P. v. Boehringer Ingelheim GMBH, 237 F.3d 1359, 1366 (Fed. Cir. 2001).

tent are balanced.[65] First, the court must find by clear and convincing evidence that threshold levels of both materiality and intent independently exist.[66]

If both thresholds have been met, the court can then engage in an equitable assessment, or balancing, of the merits to determine whether a finding of unenforceability is appropriate:

> At this second stage, . . . the question is no longer whether materiality and/or intent to deceive were proven with evidence that is sufficiently clear and convincing. . . . [T]he district court must balance the *substance* of those now-proven facts and all the equities of the case to determine whether the severe penalty of unenforceability should be imposed.[67]

The balancing process, as explained by *Star Scientific*, may be depicted as shown in Figure 1, below. The thresholds of materiality and intent are represented by the dashed lines. If either materiality or intent (or both) fall below those thresholds, there can be no inequitable conduct.[68] If both thresholds have been met, the court must then engage in discretionary balancing to determine whether the conduct warrants a finding of unenforceability. If the levels of materiality, intent, or both are low, the court may conclude that no inequitable conduct should be found, as shown by the striped area below and to the left of the solid curve. If the levels of materiality, intent, or both are high, the court may reach a finding of inequitable conduct, as shown by the area above and to the right of the solid curve. The solid curve approximately represents the continuum between the endpoints described by the Federal Circuit as "[t]he more material the omission or the misrepresentation, the lower [the] level of intent [is] required to establish inequitable conduct, and vice versa."[69]

65. 537 F.3d 1357, 1367 (Fed. Cir. 2008).

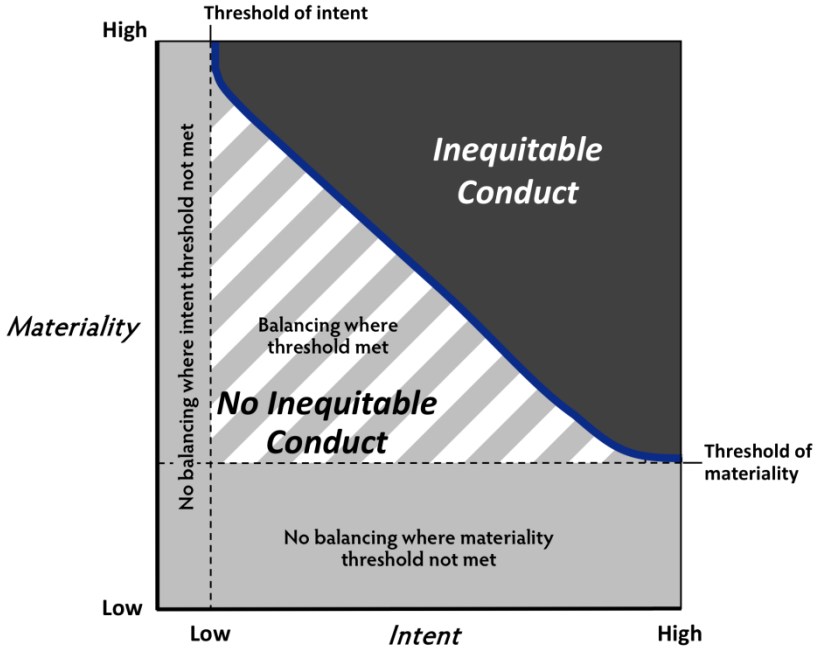
66. *Id.* at 1367.

67. *Id.*

68. *E.g.*, Nordberg, Inc. v. Telsmith, Inc., 82 F.3d 394, 398 (Fed. Cir. 1996) (declining to balance materiality and intent because Telsmith failed to make threshold showing of intent).

69. *Star Scientific*, 537 F.3d at 1367 (quoting Critikon, Inc. v. Becton Dickinson Vascular Access, Inc., 120 F.3d 1253, 1256 (Fed. Cir. 1997)).

Figure 1: Balancing Materiality and Intent



Thus, as Figure 1 illustrates, not every finding of (greater-than-threshold) intentional withholding of (greater-than-threshold) material information requires a finding of inequitable conduct. There are, however, few reported cases in which a court has found both thresholds were satisfied, but nonetheless exercised its discretion at the balancing stage to refuse to find inequitable conduct.[70]

III. IS INEQUITABLE CONDUCT A "PLAGUE"?

At the time of the *Burlington* and *Kingsdown* decisions, was there a "plague" of inequitable conduct claims? If so, does it exist today—either continuously since 1988, or after an interval of quiescence? The "plague" label suggests a thing that is both undesirable and prevalent. This Part addresses both of these concepts. Section A considers the circumstances under which the asser-

70. *See, e.g.*, Rentrop v. Spectranetics Corp., 550 F.3d 1112, 1120 (Fed. Cir. 2008); Informatica Corp. v. Bus. Objects Data Integration, Inc., 489 F. Supp. 2d 1060, 1075 (N.D. Cal. 2007) ("[T]he Court determines that the very minimal showing of materiality, if any, balanced against at most a very weak inference of intent to withhold that prior art, weighs against a determination that [patentee] is culpable of inequitable conduct as to those products.").

tion of inequitable conduct is undesirable. Section B evaluates the extent to which inequitable conduct is actually used in the case law.

A. UNDESIRABILITY OF INEQUITABLE CONDUCT AS A LITIGATION TACTIC

The suggestion that inequitable conduct is undesirable is inherently a subjective assessment. This undesirability assessment focuses on the extent to which the doctrine provides incentives to assert inequitable conduct merely for its value as a litigation tactic. For accused infringers, inequitable conduct has two principal values as a litigation tactic, particularly when liability issues (e.g., infringement, validity, and enforceability) are tried together in an unbifurcated jury trial.[71]

One incentive to assert the inequitable conduct defense is its effect of rendering the entire patent, and possibly any descendant patents,[72] unenforceable. Under current doctrine the entire patent is unenforceable, even if inequitable conduct is proven as to just a single claim and the misrepresentation

71. Because inequitable conduct is an issue arising in equity, there is no Seventh Amendment right to jury trial. Paragon Podiatry Lab., Inc. v. KLM Labs., Inc., 984 F.2d 1182, 1190 (Fed. Cir. 1993). Nonetheless, the issue is frequently tried to a jury, either with the jury as the fact finder or, more frequently, with the jury acting in an advisory capacity for the court. *See, e.g.*, Kemin Foods, L.C. v. Pigmentos Vegetales del Centro S.A. de C.V., 464 F.3d 1339 (Fed. Cir. 2006); ISCO Int'l., Inc. v. Conductus, Inc., 123 F. App'x 974 (Fed. Cir. 2005); Specialty Rental Tools & Supply, Inc. v. Boyd's Bit Serv., Inc., 84 F. App'x 90 (Fed. Cir. 2003); Duro-Last, Inc. v. Custom Seal, Inc., 321 F.3d 1098 (Fed. Cir. 2003); Juicy Whip, Inc. v. Orange Bang, Inc., 292 F.3d 728 (Fed. Cir. 2002); Mentor H/S, Inc. v. Med. Device Alliance, Inc., 244 F.3d 1365 (Fed. Cir. 2001); Upjohn Co. v. Mova Pharm. Corp., 225 F.3d 1306 (Fed. Cir. 2000); Insituform Techs., Inc. v. Cat Contracting, Inc., 161 F.3d 688 (Fed. Cir. 1998); Hupp v. Siroflex of Am., Inc., 122 F.3d 1456 (Fed. Cir. 1997); Cargill, Inc. v. Sears Petroleum & Transp. Corp., 388 F. Supp. 2d 37, 62 (N.D.N.Y. 2005); Echometer Co. v. Lufkin Indus., Inc., No. 7:00-CV-0101-N, 2004 U.S. Dist. LEXIS 30583, at *1–2 (N.D. Tex. Mar. 22, 2004); ISCO Int'l, Inc. v. Conductus, Inc., 279 F. Supp. 2d 489, 499–500 (D. Del. 2003); Transclean Corp. v. Bridgewood Servs., Inc., 101 F. Supp. 2d 788, 792 (D. Minn. 2000). Due to the technical complexity of patent cases, together with the potential for certain issues to be case-dispositive, and the risk that evidence pertinent only to some issues will "infect" decision-making on other issues, courts have frequently experimented with various forms of bifurcation or phasing in patent jury trials. *Cf.* J. Howard T. Markey, *On Simplifying Patent Trials*, 116 F.R.D. 369 (1987). *But see* Rita Mankovich Irani, *The New Skirmish in Patent Cases: Who Goes First at Trial and with What Evidence?*, 17 AIPLA Q.J. 364 (1989).

72. The doctrine of declaring related patents unenforceable is known as infectious unenforceability. Nilssen v. Osram Sylvania, Inc., 440 F. Supp. 2d 884, 902 (N.D. Ill. 2006) (discussing doctrine of infectious unenforceability), *aff'd*, 504 F.3d 1223 (Fed. Cir. 2007). The requirements for infectious unenforceability are not clearly defined, but center on an "immediate and necessary relation" between the claims tainted by inequitable conduct and claims in related patents sought to be declared unenforceable through infectious unenforceability. Consol. Aluminum Corp. v. Foseco Int'l Ltd., 910 F.2d 804, 810–11 (Fed. Cir. 1990).

or omission does not impact the other claims of the patent.[73] Likewise, because of its capacity to take down the entire patent, the target claim for an inequitable conduct defense need not even be one of the claims that the plaintiff has asserted in its infringement action.

Additionally, an allegation of inequitable conduct introduces a pro-defendant narrative of human drama into a proceeding that a jury might otherwise find to be dry and technical. Infringement analysis requires comparing the accused products to the asserted claims on a limitation by limitation basis,[74] and invalidity analysis requires comparing the asserted claims to the prior art.[75] Patent owners frequently tell an invention story to introduce a spark of human drama in patent cases. In the archetypical invention story, the inventor is a kind of hero, toiling away in obscurity on a problem that most of her contemporaries thought could not be solved. The invention is conceived in a "Eureka!" moment that goes on to change the lives of millions of Americans. Inequitable conduct provides the accused infringer a counter-narrative with which to impugn the character of the inventor. The inequitable conduct narrative presents the inventor and the lawyer who prosecuted the patent application as scoundrels who lied, cheated, and misled the Patent Office to obtain issuance of the patent. The power of this counter-narrative provides accused infringers a strong incentive to tell it whenever possible.

The question of undesirability of inequitable conduct as a litigation tactic may thus be summarized in this fashion. The doctrine of inequitable conduct, as it currently stands, is overbroad. The standards for materiality, intent, and their equitable balancing are vaguely and inconsistently defined in the case law. The sole remedy, unenforceability of the entire patent, provides a complete victory to the accused infringer who successfully asserts the inequitable conduct defense in litigation. This combination of vague standards and a powerful remedy incentivizes accused infringers to assert the defense whenever possible, strategically choosing articulations of the standards that are most favorable to the particular facts of the case. Moreover, because inequitable conduct renders the entire patent unenforceable, a party has further incentives to allege the defense as to claims that it has not been accused of infringing.[76] Viewed from a perspective of valuing the integrity of the judicial

73. Fox Indus., Inc. v. Structural Pres. Sys., Inc., 922 F.2d 801, 804 (Fed. Cir. 1990).

74. *E.g.*, *In re* Gabapentin Patent Litig., 503 F.3d 1254, 1259 (Fed. Cir. 2007).

75. *E.g.*, *In re* Crish, 393 F.3d 1253, 1256 (Fed. Cir. 2004).

76. *See* Lynch, *supra* note 6, at 8, 15–18 (summarizing tactical reasons to assert inequitable conduct in litigation).

process (and certainly from the perspective of patentee-plaintiffs[77]), such a broad, litigation-strategy-motivated use of the inequitable conduct doctrine may indeed be undesirable.[78]

This is not to say that the inequitable conduct defense is undesirable as a whole. Patentees who intentionally withhold information or make misleading statements to the Patent Office, with the intent to deceive the Patent Office, should be held accountable for that conduct when it is sufficiently egregious. Thus, it is not the defense itself that presents problems; rather, the boundaries of the doctrine as it has evolved have become vague and malleable, inviting the over-use of the inequitable conduct defense as a litigation tactic, even in cases where the factual basis for asserting the defense is weak.

---

77. The author of this Article claims no affiliation with either a pro-patentee or pro-defendant point of view. However, it is important to keep in mind that some of those who advocate for or against particular reform proposals do likely have a self-interested bias concerning the desirability or undesirability of a broad inequitable conduct doctrine.

78. Aventis Pharma S.A. v. Amphastar Pharm., Inc., 525 F.3d 1334, 1349–50 (Fed. Cir. 2008) (Rader, J., dissenting). Judge Rader wrote,

> Although designed to facilitate USPTO examination, inequitable conduct has taken on a new life as a litigation tactic. The allegation of inequitable conduct opens new avenues of discovery; impugns the integrity of patentee, its counsel, and the patent itself; excludes the prosecuting attorney from trial participation (other than as a witness); and even offers the trial court a way to dispose of a case without the rigors of claim construction and other complex patent doctrines. This court has even observed a number of cases, such as this one, that arrive on appeal solely on the basis of inequitable conduct where the trial court has apparently elected to try this issue in advance of the issues of infringement and validity.

Id.; *see also* S. REP. NO. 110-259, at 3 (2008) (characterizing inequitable conduct's defects as "troubling, plaintiff-focused litigation concerns"); *id.* at 3–4 n.11 (citing testimony from congressional hearings); *id.* at 33 n.156 ("The Committee heard some concerns that inequitable conduct is 'over plead' and a tool of harassment."); *id.* at 59–62. The Committee Report noted,

> [S]hifting standards [of inequitable conduct] . . . encourage improper challenges to patents . . . giv[ing] rise to significant litigation costs and uncertainty about patent rights . . . . The inequitable conduct defense today has become a convenient and frequently raised litigation tactic that is overpled and a quick route to taking down otherwise valid and commercially valuable patents . . . . The defense has proven to be irresistible for litigants—if proven, it allows an infringer to escape any liability for infringing a valid patent. This powerful incentive leads defendants to raise even the most questionable inequitable conduct challenges on the remote chance that they will prevail.

Id.; H.R. REP. NO. 110-314, at 21, 43 (2007) (explaining how inequitable conduct is overused, leading to complexity, expense and uncertainty in patent litigation, with the burden falling disproportionately on the patent owner).

B. THE PREVALENCE OF INEQUITABLE CONDUCT

Having thus considered the undesirability of inequitable conduct as a litigation tactic, this Section examines empirical data in order to assess the prevalence of the inequitable conduct defense.[79]

1. Methodology and Data

In particular, this Section analyzes data from the Federal Circuit cases decided between 1982 and 2008[80] that address inequitable conduct, together with selected data from the district courts.

The Federal Circuit has issued over 600 cases since 1983 that mention "inequitable conduct."[81] Over 300 of those cases substantively address, and contain a ruling on, an issue of inequitable conduct. In most cases, that ruling was made directly on the merits of a defense of inequitable conduct. Some-

---

79. Chief Judge Michel has called for the use of more empirical data, particularly in relation to Congress' patent reform efforts. Paul R. Michel, Chief Judge, U.S. Court of Appeals for the Fed. Circuit, A Strong Patent System, As Prepared For Delivery to the Association of Corporate Patent Counsel 2 (Jan. 28, 2009), *available at* http://www.cafc.uscourts.gov/pdf/1-28-09_CJMACPC_Speech.pdf (referring to "claims . . . made in vague, general terms, devoid of statistics" and criticizing the absence of "comprehensive data about real cases").

80. The data set for this analysis includes only Federal Circuit cases that use the phrase "inequitable conduct" (as identified by a Westlaw search of the CTAF database using the search term "inequitable conduct") and that substantively address an issue of inequitable conduct (as identified by a manual review of the cases). *See* Shashank Upadhye, *Liar Liar Pants on Fire: Towards a Narrow Construction for Inequitable Conduct As Applied to the Prosecution of Medical Device and Drug Patent Applications*, 72 UMKC L. REV. 669, 727, 734 (2004) (providing a selective listing and categorization of cases). Additionally, there are a handful of early Federal Circuit cases that address an early iteration of the defense as "fraud on [or in] the Patent Office." Sun-Tek Indus., Inc. v. Kennedy Sky Lites, Inc., 848 F.2d 179 (Fed. Cir. 1988); Concrete Unlimited Inc. v. Cementcraft, Inc., 776 F.2d 1537 (Fed. Cir. 1985); USM Corp. v. SPS Tech., Inc., 770 F.2d 1035 (Fed. Cir. 1985); Kimberly-Clark Corp. v. Johnson & Johnson, 745 F.2d 1437 (Fed. Cir. 1984); Hycor Corp. v. Schlueter Co., 740 F.2d 1529 (Fed. Cir. 1984); Preemption Devices, Inc. v. Minn. Mining & Mfg. Co., 732 F.2d 903 (Fed. Cir. 1984); Studiengesellschaft Kohle, M.B.H. v. Dart Indus., Inc., 726 F.2d 724 (Fed. Cir. 1984); Am. Hoist & Derrick Co. v. Sowa & Sons, Inc., 725 F.2d 1350 (Fed. Cir. 1984); Connell v. Sears, Roebuck & Co., 722 F.2d 1542 (Fed. Cir. 1983); Envtl. Designs, Ltd. v. Union Oil Co. of Calif., 713 F.2d 693 (Fed. Cir. 1983). However, because the doctrine of inequitable conduct is broader than fraud, see J.P. Stevens & Co. v. LexTex Ltd., Inc., 747 F.2d 1553, 1559 (Fed. Cir. 1984), and because none of these cases found fraud to exist (i.e., they affirmed district court findings of no fraud, or vacated or reversed district court findings of fraud), they have not been included in the data set. In essence, the no-fraud rulings cannot necessarily be equated with no-inequitable-conduct rulings because conduct that is found to be short of the narrow definition of fraud could have nonetheless satisfied the elements of the more expansive definition of inequitable conduct.

81. Westlaw, CTAF Database (providing search results following search term "inequitable conduct").

times, however, the ruling was made in the context of some other issue, such as a request for attorneys' fees. The data set includes all rulings that substantively address inequitable conduct. Additionally, as an approximate measure of the Federal Circuit's overall load of patent cases, this Section examines the number of reported cases in the Federal Circuit database in Westlaw that use the term "patent."[82]

Insofar as Federal Circuit decisions effectively represent the final terminus of the litigation cycle for inequitable conduct cases,[83] pleadings at the district court stage represent the initial terminus. However, with nearly 50,000 patent cases filed in the same 26-year period (1983–2008),[84] it was beyond the scope of this Article to manually compile and examine the pleadings in those cases. Accordingly, this Article compiled data concerning pleadings alleging inequitable conduct from two sources, Westlaw's Federal Filings database (Westlaw) and the Stanford Intellectual Property Litigation Clearinghouse (IPLC).[85] These sources include data on the extent to which inequitable conduct is asserted in pleadings since the year 2000. The Westlaw search[86] identifies the number of answers or other responsive pleadings in Westlaw's database of federal court filings[87] that include the terms "inequitable conduct"

82. Westlaw, CTAF Database (providing sources from search conducted on Oct. 12, 2009, using search terms "patent" and da(=yyyy)).

83. The Supreme Court has not heard a case involving inequitable conduct or related doctrines in over sixty years. The last such case was *Precision Instrument Mfg. Co. v. Automotive Maint. Mach. Co.*, 324 U.S. 806 (1945).

84. *See infra* Table 2, col. A.

85. Stanford IP Litigation Clearinghouse, http://lexmachina.stanford.edu (last visited Feb. 25, 2009) (providing search results for an advanced search conducted on February 25, 2009. Search results restricted to patent cases, restricted to answers, using search term "inequitable conduct," with a date restriction of cases docketed between 1/1/yyyy and 12/31/yyyy for 2000–2008).

86. Westlaw, FED-FILING-ALL Database (limiting Oct. 12, 2009 search to "Pleadings" and "Answers and Counterclaims," and utilizing the search terms "inequitable conduct" & patent & da(=yyyy)). The number of documents resulting from this search may be somewhat over-inclusive as an indicator of the number of cases in which inequitable conduct is pled, to the extent multiple answers (e.g., amended pleadings or replies to counterclaims) allege inequitable conduct in same case. However, assuming the rate of multiple pleadings remains constant from year to year, the rate of change from year to year would not be impacted by any such over-inclusiveness. The data may also be somewhat underinclusive, to the extent inequitable conduct is alleged in a complaint (e.g., unfair competition or declaratory relief claim) and inequitable conduct is not specifically mentioned in the answer.

87. Westlaw, FED-FILING-ALL Database. According to Westlaw's online database summary, as of January 2009, this database provides coverage starting in 2000. As of October 2009, the same summary stated that the database provided coverage starting in 1995. Although searches for the years 1991–1999 provided non-zero results, those results cannot be compared with results from the IPSC. Those results have been presented in italics in Table

and "patent." This search provides an approximation of the number of cases in which inequitable conduct is pled.

Data from both the Federal Circuit and district court searches are summarized below in Table 1. Table 1 also includes the number of patent cases filed[88] in U.S. District Courts each year.[89]

1, but are not otherwise addressed in this Article.

88. U.S. Courts, Judicial Business of the United States Courts, www.uscourts.gov/judbususc/judbus.html (last visited Oct. 22, 2009) (hosting annual reports for 1993–2008). *See also* Patstats.com, Historical Filings, http://www.patstats.org/Historical_Filings_Patent Suits_OtherSuits.doc (providing Administrative Office of U.S. Courts annual report statistics for 1983–2007). *But see* Amanda Bronstad, *Patent Infringement Filings Take a Nosedive*, NAT'L L.J., Jan. 19, 2009, *available at* http://www.law.com/jsp/nlj/PubArticleNLJ.jsp?id=1202427537496 (citing data from Stanford IP Litigation Clearinghouse and indicating that only 2605 patent suits were filed in 2008).

89. It does not include, for example, patent infringement cases initiated in the ITC; however, the number of such cases, or investigations, is relatively minimal. For a list of all ITC Section 337 investigations, which number fewer than 700 instituted between 1976 and 2008, See Listing of U.S. International Trade Commission Section 337 Investigations, http://info.usitc.gov/ouii/public/337inv.nsf/All?OpenView (last visited Feb. 16, 2010). *See also* RUSSELL E. LEVINE, THE PRO'S AND CON'S OF PATENT LITIGATION BEFORE THE INTERNATIONAL TRADE COMMISSION 1 (2006), http://www.abanet.org/litigation/committees/intellectual/roundtables/1106_outline.pdf (stating that from 1995 to 2000, an average of 12 investigations were instituted annually, and that the number of investigations increased to 34 in 2006).

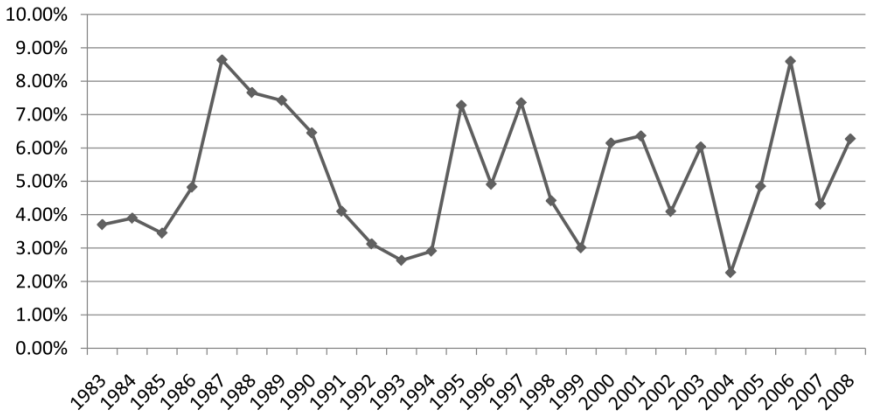
Table 1: Inequitable Conduct Data

Year	Patent Cases Filed	District Court: IC Pled		Federal Circuit:			
		West	IPSC	IC	No IC	DCT Vacated	All Patent Cases
1983	1017			1	3	0	108
1984	1057			3	3	0	154
1985	1155			2	4	1	203
1986	1105			2	9	0	228
1987	1129			5	14	0	220
1988	1226			0	15	3	235
1989	1155			1	12	4	229
1990	1238			4	10	0	217
1991	1171	*2*		1	5	2	195
1992	1474	*2*		1	5	0	192
1993	1553	*5*		3	2	0	190
1994	1617	*3*		1	2	2	172
1995	1723	*10*		4	10	2	220
1996	1840	*7*		4	7	0	224
1997	2112	*17*		4	14	2	272
1998	2218	*33*		1	10	0	249
1999	2318	*40*		1	7	0	266
2000	2484	110	75	3	11	1	244
2001	2520	200	186	2	16	0	283
2002	2700	362	335	2	8	2	293
2003	2814	565	473	2	13	2	282
2004	3075	759	671	0	4	2	265
2005	2720	827	944	5	6	3	289
2006	2830	926	1087	3	16	3	256
2007	2896	1148	1472	5	6	1	278
2008	2909	1157	1631	4	14	1	319

2. Preliminary Analysis of Federal Circuit Data

To test the proposition that the inequitable conduct allegations are growing, this Article first looked at the percentage of Federal Circuit cases per year that address inequitable conduct, regardless of the outcome the Federal Circuit reached on the inequitable conduct issue. This is illustrated in Figure 2.

Figure 2: Percentage of Federal Circuit Patent Cases Addressing Inequitable Conduct



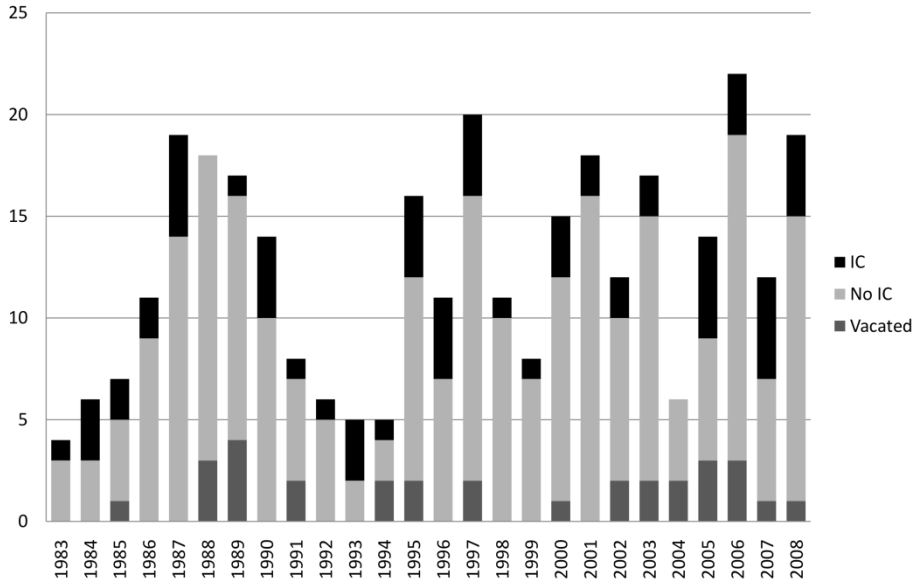
The total body of Federal Circuit inequitable conduct decisions shows a large spike in the number of cases in 1987, the year before *Kingsdown* and *Burlington* were decided, with a marked drop in the number of cases ruling on inequitable conduct during the period 1988 through 1994. Since then, the results at the Federal Circuit level have been mixed. Because the Federal Circuit's overall load of patent cases has fluctuated and generally increased over time, it would be reasonable to expect a corresponding variation in the absolute number of inequitable conduct rulings by the Federal Circuit. To account for this variation in the Federal Circuit's patent case load, the graph in Figure 2 provides a year-by-year detail of the percentage of Federal Circuit patent cases that address inequitable conduct (measured as the ratio of the number of Federal Circuit cases that address inequitable conduct to the number of Federal Circuit cases using the term "patent").

In 2006, the year of the *Ferring* and *Digital Control* decisions, the percentage of cases addressing inequitable conduct nearly reached the level seen in 1987. Otherwise, the wide fluctuation in the percentages since 1994 does not provide conclusive insights.

In addition to analyzing the overall frequency of all inequitable conduct rulings, it may be useful to assess how the Federal Circuit actually ruled on the inequitable conduct issue. This Article separates the rulings into three categories: "IC" when the Federal Circuit affirmatively ruled there was inequitable conduct, "No IC" when the Federal Circuit affirmatively ruled there was no inequitable conduct, and "Vacated" when the Federal Circuit vacated a district court ruling on inequitable conduct. The results of this analysis are

in Figure 3, which shows the absolute numbers (rather than percentages) of the various types of inequitable conduct rulings.

Figure 3: Federal Circuit Cases Addressing Inequitable Conduct



The graph in Figure 3 shows several noteworthy data points. In 1987, the year before *Kingsdown* and *Burlington*, the Federal Circuit found inequitable conduct to exist in five cases. Over the next several years, although the total count of inequitable conduct rulings remained relatively high, the number of "IC" cases dropped significantly. There was not another year with five "IC" rulings until 2005. This can be seen in Figure 4, which shows only the cases where the Federal Circuit found inequitable conduct.

Figure 4: Number of CAFC Cases Finding Inequitable Conduct

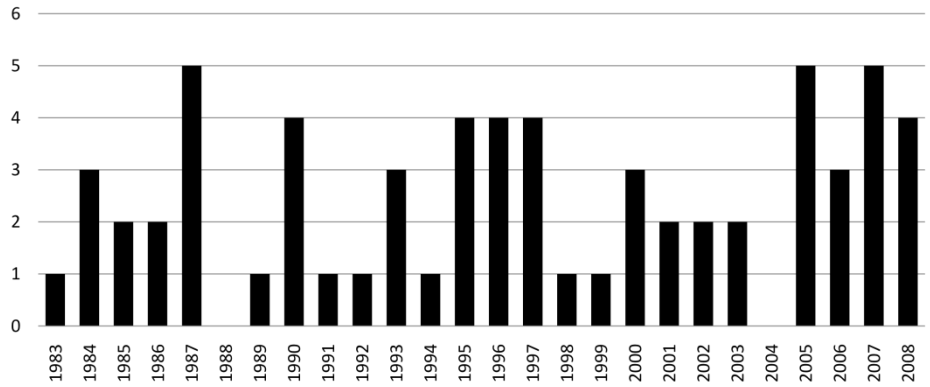


Figure 4 shows a cluster of three to five "IC" rulings per year over the past four years. This is above the average of just over 2.5 "IC" rulings per year. In absolute numbers, that sustained above-average trend is unmatched in the history of the Federal Circuit. But is it appropriate to declare the recent trend a "plague," relative either to the 1987–1988 period, or to the overall 25-year pattern of inequitable conduct rulings? Given the overall growth in the Federal Circuit's case load over that period, this conspicuous cluster may not be sufficiently decisive to warrant declaration of a "plague."

3. Further Analysis of Federal Circuit Data

Several additional patterns emerge when the Federal Circuit data is examined more closely. Virtually all of the Federal Circuit's findings of inequitable conduct affirm lower-court findings of inequitable conduct. Federal Circuit courts affirm lower court "No IC" rulings 92% of the time. By contrast, Federal Circuit courts affirm "IC" findings only 41% of the time. These patterns are illustrated in Figures 5 and 6.

Figure 5: Federal Circuit Dispositions of "No IC" Ruling by District Court

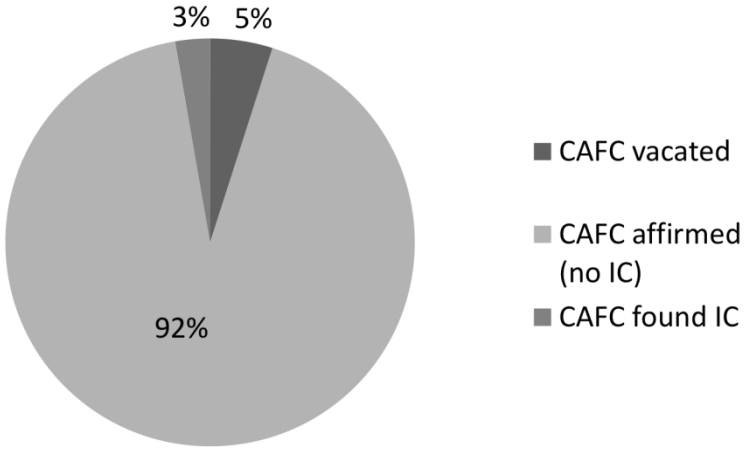
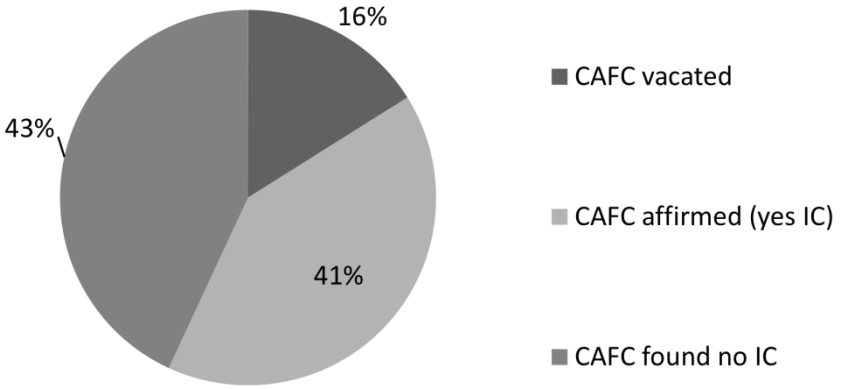


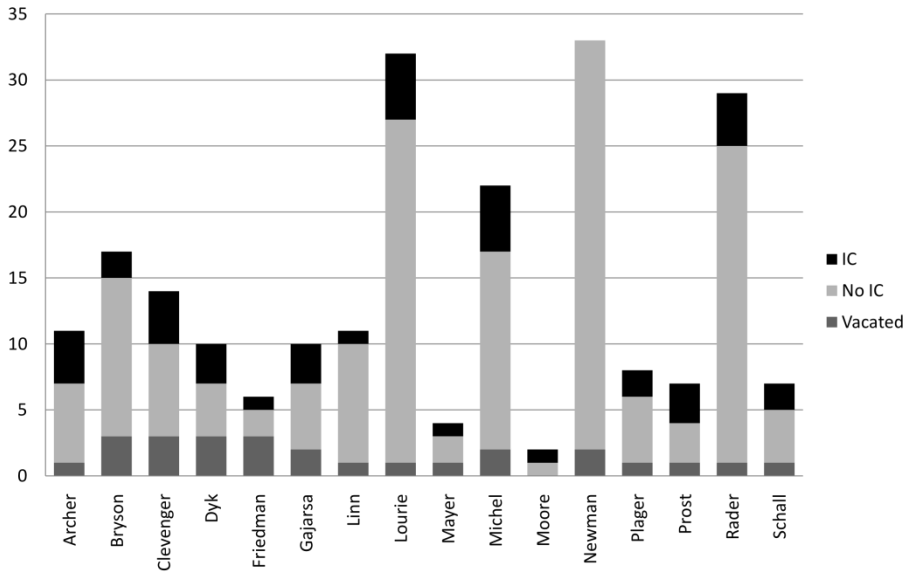
Figure 6: Dispositions of IC Findings by District Court



Another way to examine the Federal Circuit data is to consider whether the cases characterizing inequitable conduct as a "plague" reflect the perspectives of particular judges. In fact, not all judges are equally disposed to support or reject inequitable conduct claims. The next chart, Figure 7, identifies the number of majority opinions with inequitable conduct rulings written by each sitting judge on the Federal Circuit.[90]

90. This chart has been limited to sitting judges, and has excluded former judges and judges from other courts who authored an opinion while sitting by designation on the Federal Circuit. *See* United States Court of Appeals for the Federal Circuit, Judicial Biographies, http://www.cafc.uscourts.gov/judgbios.html (last visited Oct. 22, 2009) (providing a list of the current and former judges of the Federal Circuit).

Figure 7: Authors of Majority Opinions Concerning Inequitable Conduct



The height of the bar for each judge correlates approximately with his or her tenure on the court. Thus, for example, Judges Newman (appointed 1984), Michel (appointed 1988), Rader (appointed 1989), and Lourie (appointed 1990) have the longest tenure.[91] Perhaps the most striking feature of this graph is that, despite her long tenure and many opinions, Judge Newman has *never* written a majority opinion supporting a finding of inequitable conduct. When this fact is considered, it should come as no surprise that, although she did not originally coin the "plague" label, she is responsible for nine of the twelve subsequent characterizations of inequitable conduct as a "plague" in Federal Circuit opinions.[92] Judges Lourie and Rader together are responsible for the remaining three opinions. The six most recent invoca-

91. One curious exception to this pattern appears to be Judge Mayer (appointed in 1987). Also, the pattern does not hold for the judges who have taken "senior" status (Judges Friedman, Archer, Plager and Clevenger).

92. McKesson Info. Solutions, Inc. v. Bridge Med., Inc., 487 F.3d 897, 926 (Fed. Cir. 2007); Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181, 1196–97 (Fed. Cir. 2006), *cert. denied*, 549 U.S. 1015 (2006); Ulead Sys., Inc. v. Lex Computer & Mgmt. Corp., 351 F.3d 1139, 1155 (Fed. Cir. 2003); Hoffmann-La Roche, Inc. v. Promega Corp., 323 F.3d 1354, 1372 (Fed. Cir. 2003); Ohio Cellular Prods. Corp. v. Adams USA, Inc., 175 F.3d 1343, 1355 (Fed. Cir. 1999); Litton Sys., Inc. v. Honeywell, Inc., 140 F.3d 1449, 1469 (Fed. Cir. 1998); Multiform Desiccants, Inc. v. Medzam, Ltd., 133 F.3d 1473, 1482 (Fed. Cir. 1998); Magnivision, Inc. v. Bonneau Co., 115 F.3d 956, 960 (Fed. Cir. 1998); Allied Colloids Inc. v. American Cyanamid Co., 64 F.3d 1570, 1578 (Fed. Cir. 1995).

tions of "plague" have appeared in dissenting opinions, five of which were written by Judge Newman.[93]

4. Analysis of District Court Data

It appears that the patterns of inequitable conduct rulings at the Federal Circuit are not strongly indicative of a "plague." While there is an identifiable cluster of inequitable conduct findings in the past several years, the overall pattern of the Federal Circuit is to affirm most findings of "No IC" and to vacate or reverse over half of the lower court findings of inequitable conduct that it addresses on appeal.

The data concerning the frequency with which inequitable conduct is pled in the district courts tells a different story. There has been a strong upward trend. Table 2 analyzes some of the data identified in Table 1.[94]

93. *McKesson*, 487 F.3d at 926; *Ferring*, 437 F.3d at 1196–97; *Ulead*, 351 F.3d at 1155; *Hoffmann-La Roche*, 323 F.3d at 1372; *Ohio Cellular*, 175 F.3d at 1355.

94. Table 2 does not include data concerning district court adjudications of inequitable conduct. *See* Katherine Nolan-Stevaux, Note, *Inequitable Conduct Claims in the 21st Century: Combating the Plague*, 20 BERKELEY TECH. L.J. 147, 163–64 (2005) (providing data on summary judgment and post-trial adjudications of inequitable conduct by district courts from 1995–2004); Patstats Home Page, http://www.patstats.org/patstats2.html (last visited Nov. 18, 2009) (compiling statistics on patent cases, including district court rulings on inequitable conduct).

Table 2: Selected Data from District Court Filings and Federal Circuit Outcomes

	A	B	C	D	E	F	G	H
	District Courts			Federal Circuit			IC @ CAFC ÷ IC Pled	IC @ CAFC ÷ Patent Cases Filed
Year	Patent Cases Filed	IC Pled	IC Pled ÷ Cases Filed	IC	No IC	DCT Vacated		
1983	1017			1	3	0		0.10%
1984	1057			3	3	0		0.28%
1985	1155			2	4	1		0.17%
1986	1105			2	9	0		0.18%
1987	1129			5	14	0		0.44%
1988	1226			0	15	3		0.00%
1989	1155			1	12	4		0.09%
1990	1238			4	10	0		0.32%
1991	1171	*2*		1	5	2		0.09%
1992	1474	*2*		1	5	0		0.07%
1993	1553	*5*		3	2	0		0.19%
1994	1617	*3*		1	2	2		0.06%
1995	1723	*10*		4	10	2		0.23%
1996	1840	*7*		4	7	0		0.22%
1997	2112	*17*		4	14	2		0.19%
1998	2218	*33*		1	10	0		0.05%
1999	2318	*40*		1	7	0		0.04%
2000	2484	110	4%	3	11	1	2.73%	0.12%
2001	2520	200	8%	2	16	0	1.00%	0.08%
2002	2700	362	13%	2	8	2	0.55%	0.07%
2003	2814	565	20%	2	13	2	0.35%	0.07%
2004	3075	759	25%	0	4	2	0.00%	0.00%
2005	2720	827	30%	5	6	3	0.60%	0.18%
2006	2830	926	33%	3	16	3	0.32%	0.11%
2007	2896	1148	40%	5	6	1	0.44%	0.17%
2008	2909	1157	40%	4	14	1	0.35%	0.14%

Column A lists the number of patent cases filed in U.S. District Courts each year.

Column B identifies the number of answers or other responsive pleadings in Westlaw's database of federal court filings that include the terms "inequitable conduct" and "patent." This roughly approximates the number of cases in which inequitable conduct is pled. While Table 1 also includes simi-

lar data from the Stanford IP Litigation Clearinghouse, this Article elects to focus primarily on the Westlaw data, which shows a slightly less extreme (but still dramatic) trend at the district court level.

Column C calculates the number of cases in which inequitable conduct is pled (Column B) divided by the total number of patent cases filed (Column A). Although the Table presents this calculation as a percentage, it is appropriate to note that this calculation may not accurately represent a true percentage, for two reasons. First, the responsive pleadings asserting inequitable conduct may not have been filed in the same year as the case was filed. Second, the numbers of inequitable conduct pleadings may be both under-counted (i.e., pleadings not included in the database) and over-counted (i.e., multiple such pleadings in the same case). Nonetheless, since each of those reasons may fairly be expected to remain constant over time, this ratio does represent a useful depiction of the trend.

Columns D–F[95] contain the specific Federal Circuit data compiled by the author, as described above in connection with Table 1.

Column G calculates the number of Federal Circuit findings of inequitable conduct (Column D) divided by the number of cases in which inequitable conduct is pled at the district court level (Column B). Again, this ratio is expressed as a percentage, and represents an approximation of the percentage of patent cases in which inequitable conduct is pled that ultimately result in a ruling of inequitable conduct by the Federal Circuit.[96] This value could be considered an "ultimate success rate" for inequitable conduct allegations. However, it does not account for settled cases (in which the inequitable conduct defense may or may not have been a factor), cases in which the inequitable conduct defense was abandoned, and cases in which the district court ruled on inequitable conduct but either the issue was not appealed or the Federal Circuit did not address the issue on appeal.

Column G indicates that, over the past 7 years (2002–2008), a pleading of the inequitable conduct defense results in a Federal Circuit finding of inequitable conduct in only approximately 1 of 250 cases. Yet, as Column C indicates, it has been pled in about 3 of 10 cases during that same period, with a strong upward trend in the pleading frequency. This 75-fold differential,

95. Westlaw, CTAF Database (providing data compiled by author based on review of cases including phrase "inequitable conduct"). *Cf.* Patstats, http://www.patstats.org/patstats2.html (last visited Nov. 18, 2009) (offering various full-year statistics involving issues argued in patent cases, including inequitable conduct).

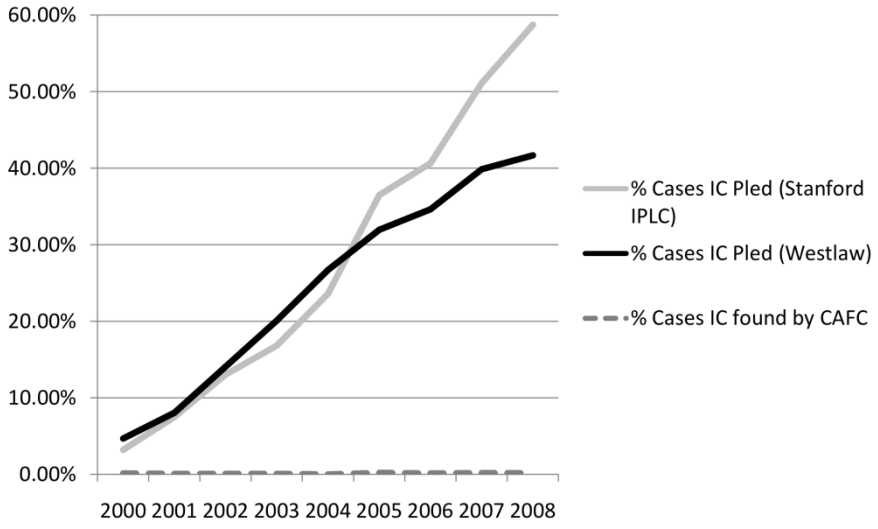
96. The pleadings and Federal Circuit rulings referred to in this ratio are almost certainly never from the same cases; it would be exceedingly rare for a case to proceed from pleading to Federal Circuit ruling within a single calendar year.

between pleading and ultimate success on appeal, supports the widely repeated belief that inequitable conduct is overpled and a "plague."

Finally, Column H calculates the Federal Circuit's findings of inequitable conduct (Column D) divided by the total number of patent cases filed in a particular year (Column A), and expresses this ratio as a percentage. With the same caveat noted above that the cases filed and cases decided on appeal within a single calendar year will not be the same cases, Column H approximates the percentage of patent cases filed that ultimately result in a Federal Circuit ruling of inequitable conduct, regardless of the frequency with which inequitable conduct is pled.

The arithmetic mean of Column H is 0.14% (or roughly 1 Federal Circuit finding of inequitable conduct per 700 patent cases filed in the district courts), with a standard deviation of 0.10%. Only a few years (1984, 1987, 1988, 1990, and 2004) fall more than one standard deviation away from the mean, suggesting that the rate at which the Federal Circuit ultimately finds inequitable conduct in patent cases is relatively stable over time, regardless of the frequency with which the defense is pled. Similarly, the 7-year average (2002–2008) 210-fold differential between Column H and Column C (which increased to more than 260-fold in 2007–2008) also supports the "plague" conclusion. Figure 8, below, depicts this growing gap in graphic form. Both the Westlaw and IPLC data are included for emphasis.

Figure 8: Growing Gap Between IC Pleadings and Federal Circuit Rulings of IC



It bears emphasizing that the calculations in Columns G and H cannot represent event-history (actual case-outcome) data. First, no effort has been made to track the history of particular cases by correlating specific filings and

pleadings with Federal Circuit rulings. Second, a case rarely, if ever, progresses from the filing of an answer in the district court to a final disposition by the Federal Circuit within a single year. Typically, two to five years might elapse between those two events in a particular case. In the absence of actual case-history, it would be arbitrary and potentially confusing to correlate appellate outcome data with case-filing data from an earlier year, rather than the same year. Additionally, there is a countervailing factor at play—the impact of recent Federal Circuit rulings on whether inequitable conduct is pled in new filings. That is, it may be appropriate to consider whether a particular Federal Circuit case or trend of cases signals, induces, or encourages parties to plead relatively more or fewer inequitable conduct allegations in subsequently-filed cases. Thus, in the absence of more detailed data about the actual procedural histories of cases or the impact of specific Federal Circuit rulings on subsequent pleadings, and to capture and reflect overall trends, Tables 1 and 2 and Figure 8 simply report the data in the year in which it was filed, and perform calculations on the data as it appears within a particular year.

5. Conclusions

To summarize: The overall volume of inequitable conduct cases at the Federal Circuit, both as a percentage of the Federal Circuit's patent case load, and in terms of absolute numbers of cases finding inequitable conduct, has trended slightly upward in the past several years. But, particularly in light of the Federal Circuit's pattern of affirming most "No IC" findings and vacating or reversing a majority of lower-court "IC" findings, this trend does not, by itself, appear to be a sufficiently dramatic change to warrant a declaration that a "plague" of inequitable conduct has returned. Among those judges who have expressed concern about inequitable conduct as a "plague," Judge Newman stands out as particularly vocal on the issue. But are her protestations warranted? The statistical data is suggestive that the prevalence of inequitable conduct cases is expanding, especially at the pleading stage.

Another possible way to address this question is to review some of the Federal Circuit's recent inequitable conduct decisions and their potential impact on the scope of the inequitable conduct doctrine. The implication may be that recent cases have expanded the doctrine, thus potentially expanding the incentives for accused infringers to allege inequitable conduct as a defense. The next Section of this Article analyzes several recent Federal Circuit inequitable conduct decisions.

IV. RECENT CASES

The Federal Circuit decided six cases during the period 2006–2008 that are of particular importance to the development of the inequitable conduct doctrine. As this Part explains, the *Ferring*, *Digital Control*, *McKesson*, *ESpeed*, *Star Scientific*, and *Praxair* cases have each contributed in important ways to the development of the inequitable conduct doctrine.

A. *FERRING*

Ferring[97] represents a return to the "should have known" standard of intent that was rebuked in *Kingsdown*. Like *Burlington* and *Kingsdown*, the appeal in the 2006 *Ferring* case arose from a district court's grant of summary judgment of inequitable conduct. Ferring's patent related to medicine administered "orally."[98] During prosecution, the examiner considered prior art relating to "peroral" administration of the compound. The applicants argued that "oral" administration meant swallowing for absorption through the gastrointestinal tract, and that "peroral" meant absorption through the walls of the mouth. The examiner suggested that the applicants "obtain evidence from a 'non-inventor' " to support this asserted distinction.[99] The applicants submitted four such declarations. However, the declarations did not disclose that Ferring had previously employed or granted funding to three of the four declarants.[100] The Federal Circuit held that these non-disclosed prior relationships were material information, particularly where the declarants' neutrality was relevant to the credibility of their assertions.[101] The Federal Circuit did not address the substantive truth or accuracy of the declarations, or the fact that one declarant apparently had no such prior relationship with Ferring.

Having determined that this omission was material, the Federal Circuit then turned to the intent prong. Although the court recited the rule that "materiality does not presume intent, which is a separate and essential component of inequitable conduct,"[102] it then disregarded the rule and inferred intent from materiality. Specifically, the court announced a new rule stating that where the undisclosed information is highly material, summary judgment on the issue of intent is appropriate if "(1) the applicant knew of the informa-

97. Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181 (Fed. Cir. 2006).

98. *Id.* at 1184.

99. *Id.* at 1183–84.

100. The decision does not mention whether the fourth outside declarant, Miller, had any such connections. *Id.* at 1185.

101. *Id.* at 1188.

102. *Id.* at 1190–91 (quoting Manville Sales Corp. v. Paramount Sys., Inc., 917 F.2d 544, 552 (Fed. Cir. 1990)).

tion; (2) the applicant knew or should have known of the materiality of the information; and (3) the applicant has not provided a credible explanation for the withholding."[103]

Judge Newman filed a sharp and lengthy dissent, arguing that the majority decision in *Ferring* "resurrects the plague of the past, ignoring the *Kingsdown* requirements of clear and convincing evidence of a misrepresentation or omission material to patentability, made intentionally and for the purpose of deception."[104] Judge Newman singled out the majority's intent ruling, stating: "The panel majority's holding that deceptive intent is established as a matter of law if the applicant 'should have known' that information might be material to patentability, further revives the 'plague' of the past, with burdens that far outweigh any conceivable benefits."[105]

Implications: Although subsequent cases have given *Ferring* a broad scope, it may be interpreted quite narrowly as part of a line of cases including *Refac*[106]and *Paragon*.[107] The *Ferring* majority quotes *Refac* to suggest that factual affidavits present special cases of heightened scrutiny.[108] As *Refac* indicates, affidavits are "inherently material," even if cumulative, precisely because they are "intended to be relied upon."[109] This suggests that affidavits generally lie at the higher end of the continuum of materiality, particularly if there are indicia that the examiner actually relied on the contents of the affidavits.[110] If so, then a relatively lower degree of intent may be required to support a finding of inequitable conduct arising out of an affidavit. *Ferring* stands for the proposition that the failure to disclose the existence of prior dealings or relationships between an affiant and a patent applicant is inexcusable, as the applicant or the affiant "should have known" that the existence of any such relationships would be pertinent. This use of the "should have known" standard in the summary judgment context comes perilously close to outright abrogation of the intent element of inequitable conduct. Whatever the doctrinal merit of this ruling, one pragmatic response to *Ferring* for patent prosecutors and inventors is to ensure that any affidavits submitted during prose-

103. *Id.* at 1191.

104. *Id.* at 1197.

105. *Id.* at 1202; *accord* Lynn C. Tyler, *Kingsdown Fifteen Years Later: What Does it Take to Prove Inequitable Conduct?*, 13 FED. CIR. B.J. 267, 268–69 (2003) (arguing that courts should not infer intent solely from failure to disclose a known reference that is material).

106. Refac Int'l, Ltd. v. Lotus Dev. Corp., 81 F.3d 1576 (Fed. Cir. 1996).

107. Paragon Podiatry Lab., Inc. v. KLM Labs., Inc., 984 F.2d 1182 (Fed. Cir. 1993).

108. *Ferring*, 437 F.3d at 1189 n.9 (quoting *Refac*, 81 F.3d at 1583).

109. *Refac*, 81 F.3d at 1583.

110. Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1315 (Fed. Cir. 2006).

cution scrupulously disclose all possible interests or relationships with the patent applicants.[111]

B. *DIGITAL CONTROL*

In February 2006, the same month it decided *Ferring*, the Federal Circuit reaffirmed the vitality of the "reasonable examiner" test in *Digital Control v. Charles Machine Works*,[112] ruling that the test represented the broadest (and lowest) threshold of materiality in the doctrine of inequitable conduct. The Federal Circuit affirmed the district court's ruling that misstatements in the applicant's Rule 131 declaration (i.e., a declaration to establish an earlier invention date, in order to "swear behind" prior art) were material, but reversed summary judgment on the materiality of undisclosed prior art, holding that there were fact issues concerning the prior art's cumulativeness.[113]

On the issue of intent, the Federal Circuit held that the district court had intertwined its findings on the Rule 131 declaration and the undisclosed prior art, requiring remand to separate out the intent analysis.[114]

In reaching this ruling, the Federal Circuit summarized the history of the materiality prong, explaining that several standards for materiality had been applied throughout the history of the doctrine of inequitable conduct.[115] After reviewing these various standards, the court ruled that the 1977 "reasonable examiner" test remained applicable, even as to patents prosecuted entirely after 1992.[116] The court found that the 1992 standard did not supplant the "reasonable examiner" standard, and that it therefore remained viable as the broadest threshold level of materiality.[117] This ruling effectively relegated to the sidelines the more objective definition of materiality introduced by the PTO's 1992 revision of Rule 56.

---

111. Senators Specter and Hatch have specifically criticized *Ferring* for finding "an applicant's failure to adequately disclose its relationship with an expert to be material even though the expert's views were accurate and true." S. REP. NO. 110-259, at 60 (2008) (noting additional views of Senators Specter and Hatch).

112. *Digital Control*, 437 F.3d 1309.

113. *Id.* at 1319, 1321–22.

114. *Id.* at 1321.

115. *Id.* at 1314–16.

116. *Id.* at 1316 ("[T]he 'reasonable examiner' standard and our case law interpreting that standard were not supplanted by the PTO's adoption of a new Rule 56."). *Compare* Elizabeth Peters, *Are We Living in a Material World?: An Analysis of the Federal Circuit's Materiality Standard Under the Patent Doctrine of Inequitable Conduct*, 93 IOWA L. REV. 1519, 1557–64 (2008) (arguing for application of the 1992 Rule 56 standard of materiality), *with* James Cronin, *Inequitable Conduct and the Standard of Materiality: Why the Federal Circuit Should Use the Reasonable Patent Examiner Standard*, 50 ST. LOUIS U. L.J. 1327, 1328–29 (2006) (arguing for application of the reasonable examiner standard).

117. *Id.*

Judge Newman repeatedly has disapproved of this ruling. In *Ferring*, Judge Newman argued in dissent that "[t]he court in *Digital Control* holds, in contradiction of precedent, that it will hold practitioners to the standard of the pre-1992 version of Rule 56 for patents prosecuted after 1992, even though that standard no longer exists."[118]

The enacting history accompanying the 1992 revisions, which refers to the "plague" of inequitable conduct claims as a motivation for the amendment, supports Judge Newman's view.[119] The 1992 amendments were intended to provide a clearer and more objective rule, which was less vague than the "reasonable examiner" standard.[120] The 1992 rule was proposed to ameliorate uncertainty in the "reasonable examiner" standard, with aspirations to minimize litigation while still providing the PTO the information it needed for efficient and effective examination.[121] Former PTO Commissioner Gerald Mossinghoff has also opined that the 1992 version of Rule 56 should govern post-1992 patents.[122]

Separation-of-Powers Issue: The debate between the majority and dissent in *Digital Control* exposes a deeper issue concerning separation of powers and the origins of the inequitable conduct doctrine. Under the majority's approach, inequitable conduct is a judicially-created doctrine.[123] As such, no deference is owed to the PTO's determinations of what information is material to patent prosecution. The *Digital Control* majority noted that the Federal Circuit has articulated several different tests for materiality, which corroborates this view.[124] However, the court has pervasively referred to PTO Rule 56 for

---

118. Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181, 1202 n.3 (Fed. Cir. 2006) (Newman, J., dissenting). *See also* Agfa Corp. v. Creo Prods. Inc., 451 F.3d 1366, 1384 (Fed. Cir. 2006) (Newman, J., dissenting).

119. Duty of Disclosure and Practitioner Misconduct, 54 Fed. Reg. 11,334 (proposed Mar. 17, 1989) ("These proposed changes are considered desirable in view of the large amount of resources that are being devoted to duty of disclosure issues both within and outside the Office without significantly contributing to the reliability of the patents being issued.").

120. Duty of Disclosure, 56 Fed. Reg. 37,321, 37,322 (proposed Aug. 6, 1991) (to be codified at 37 C.F.R. § 1.56).

121. Duty of Disclosure, 57 Fed. Reg. 2021, 2023 (Jan. 17, 1992).

122. Gerald S. Mossinghoff, The Duty of Candor and Good Faith to the United States Patent and Trademark Office, Remarks to the American Bar Association, Intellectual Property Law Section at the 17th Annual Intellectual Property Law Conference (Apr. 12, 2002) (transcript on file with Oblon, Spivak, McClelland, Maier & Neustadt, LLP), *available at* http://www.oblon.com/media/index.php?id=44 ("My own view is that the courts should apply the version (or versions) of Rule 56 that was (were) in effect at the time the conduct objected to occurred.").

123. Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1315 (Fed. Cir. 2006).

124. *Id.*

the definition of materiality, most commonly using the "reasonable examiner" standard. If the PTO was the originator of the "reasonable examiner" standard, with the courts following suit, it would stand to reason that the courts would owe some deference to any PTO modifications of the standard.

In contrast, under the approach Judge Newman takes in dissent, the PTO ought to be vested with the determination of what information is material to patent examination. Because it conducts such examination, it is therefore entitled to deference, perhaps under a principle analogous to the rule stated by the Federal Circuit in *Garner*: "An agency's interpretation of its own regulations is entitled to substantial deference and will be accepted unless it is plainly erroneous or inconsistent with the regulation."[125] This analogy is imperfect, however, since the concept being interpreted is "materiality" as used in the judicially-created doctrine of inequitable conduct, not "materiality" as used in Rule 56. Moreover, the 1992 amendment is not, strictly speaking, an interpretation of the 1977 rule—though it could be argued that the 1992 amendments may be viewed as the PTO's interpretation of what a reasonable examiner would want.

In ruling that the 1992 revision to Rule 56 did not supplant the "reasonable examiner" standard, the Federal Circuit disregarded its own precedents and the clearly articulated purpose in Rule 56's enacting history. The effect of this ruling is that a standard that the PTO has criticized as "vague" could render patents unenforceable for failure to disclose information that the PTO's own regulations would not require "for effective and efficient examination."

Implications: Despite this unsettling revival of the "reasonable examiner" standard, the court in *Digital Control* also left two openings for advocates to narrow the scope of inequitable conduct. First, by describing a continuum among the various standards of materiality, the court made clear that the "reasonable examiner" standard occupies the broadest (and therefore lowest) threshold of materiality. Thus, in balancing materiality and intent, "the requisite finding of intent must be high" where only this low threshold of materiality has been met.[126] *Digital Control*'s quantification of materiality should provide an effective counterpoint to the invariable assertion in litigation that the nondisclosed information in that case is highly material (even if it merely satisfies the "reasonable examiner" threshold) and therefore requires only a low showing of intent. Second, although the Federal Circuit carefully pro-

125. In re Garner, 508 F.3d 1376, 1378–79 (Fed. Cir. 2007) (quoting Star Fruits S.N.C. v. United States, 393 F.3d 1277, 1282 (Fed. Cir. 2005)).

126. Digital Control, Inc. v. Charles Mach. Works, 437 F.3d 1309, 1315–16 (Fed. Cir. 2006).

tected its precedents interpreting the "reasonable examiner" standard, it did not rule that the "reasonable examiner" standard and the 1992 version of Rule 56 are substantively different standards. Therefore, a fair and persuasive argument is that the PTO, through the rulemaking process of the 1992 amendment to Rule 56, has defined what information a reasonable examiner (i.e., a PTO employee) would find important in determining patentability. If successful, this argument would establish that the "reasonable examiner" standard—even if not abrogated—is coextensive with the 1992 version of Rule 56.

C. *MCKESSON*

In 2007, the Federal Circuit decided *McKesson Information Solutions, Inc. v. Bridge Medical, Inc.*[127] In *McKesson*, the Federal Circuit ruled that the same test for determining the materiality of prior art should apply to determining materiality of rejections in copending applications.[128] In other words, a rejection of claims in an inventor's copending application may be material information, and failure to disclose such a rejection can be a basis for finding inequitable conduct.

The district court found U.S. Patent No. 4,857,716 to be unenforceable due to inequitable conduct, and the Federal Circuit affirmed.[129] *McKesson* involved three lines of patent applications, all prosecuted during approximately the same time period in 1986–1989:

- Application no. 06/862,278 (the '278 application) led to continuation application no. 07/205,527, which led to U.S. Patent No. 4,857,716.
- The '278 application also led to continuation-in-part application no. 07/078,195-a, which led to U.S. Patent No. 4,835,372.
- A separate application, no. 06/862,149 (the '149 application), led to U.S. Patent No. 4,850,009.[130]

Examiner Trafton examined the applications descending from the '278 application,[131] while Examiner Lev examined the '149 application.[132] The same attorney, Schumann, prosecuted all three lines of applications.[133] Dur-

127. 487 F.3d 897 (Fed. Cir. 2007).

128. *See id.* at 920–921 (evaluating materiality of examiner's rejections under reasonable examiner test).

129. *Id.* at 902, 926.

130. *Id.* at 902–07.

131. *Id.* at 903–04, 906–07.

132. *Id.* at 904–06.

133. *Id.* at 902–07.

ing prosecution of the '149 application, Examiner Lev rejected claims on two occasions, and attorney Schumann did not disclose those rejections to Examiner Trafton in the '527 application.

The Federal Circuit affirmed the district court's ruling that attorney Schumann had committed three separate acts of inequitable conduct, including failing to tell Examiner Trafton about Examiner Lev's rejections of copending claims. The patentee argued, based on the *Dayco* case,[134] that there was no duty to disclose the rejection to Examiner Trafton because the rejected claims were not "substantially similar" to claims pending in the '527 application. In response, the Federal Circuit held that *Dayco* was not binding on the court's materiality inquiry:

> Under *Dayco*, [materiality under the reasonable examiner] standard is satisfied in the rejected-claims setting if the rejected claims are substantially similar to the claims at issue. In other words, a showing of substantial similarity is *sufficient* to prove materiality. It does not necessarily follow, however, that a showing of substantial similarity is *necessary* to prove materiality.[135]

This ruling strikes down the bright-line *Dayco* rule that the duty to disclose copending rejections exists only where the claims in the two applications are substantially similar (i.e., "could have conceivably served as the basis of a double patenting rejection"[136]).

The *McKesson* court was also significantly deferential to the district court on the issue of intent, leading Judge Newman to argue in dissent that the "plague" had returned:

> To avoid the inequity resulting from litigation-driven distortion of the complex procedures of patent prosecution, precedent firmly requires that the intent element of inequitable conduct must be established by clear and convincing evidence of deceptive intent—not of mistake, if there were such, but of culpable intent. . . . That standard was not met here. This court returns to the "plague" of encouraging unwarranted charges of inequitable conduct, spawning the opportunistic litigation that here succeeded despite consistently contrary precedent.[137]

---

134. Dayco Prods., Inc. v. Total Containment, Inc., 329 F.3d 1358, 1367 (Fed. Cir. 2003) (holding that intent to deceive cannot be inferred simply from the decision to withhold the reference where the reasons given for the withholding are plausible).

135. *McKesson*, 487 F.3d at 919 (Fed. Cir. 2007) (emphasis added) (citation omitted).

136. *Dayco*, 329 F.3d at 1365 (quoting Akron Polymer Container Corp. v. Exxel Container, Inc., 148 F.3d 1380, 1382 (Fed. Cir. 1998)).

137. *McKesson*, 487 F.3d at 926–27 (citation omitted).

Implications: Particularly for large technology companies with many patent applications simultaneously pending before the PTO, the absence of a bright-line rule concerning materiality of rejections in copending applications could impose an enormous burden on both applicants and examiners. The applicants would have to assess whether each office action rejecting claims in any pending application is material to, and therefore must be disclosed in, every single other pending application being prosecuted by that company.

However, as with *Digital Control*, despite the apparent and worrisome broadening of the duty to disclose rejections in copending applications, the court's ruling provides some clues about arguments that may limit the scope of the ruling. On the issue of intent, the court compared the *McKesson* facts favorably with those found in *Li Second Family LP v. Toshiba Corp.*,[138] in particular that "the applicant made statements to the examiner [i.e., Trafton] inconsistent with the other examiner's [i.e., Lev's] decisions."[139] Thus, because Schumann argued to Examiner Trafton that the prior art did not disclose "3-node communication," and Examiner Lev subsequently made a rejection in another application based in part on prior art that in fact disclosed 3-node communication, Schumann's failure to disclose Lev's rejection to Trafton supported an inference that Schumann intended to mislead the PTO via the nondisclosure of information that would have undermined the argument he had made to Trafton. The court's analysis actually seems more germane to materiality, namely that the fact of having made statements to one examiner inconsistent with the other examiner's rejection increases the materiality of that rejection to the other application. This interpretation offers a possible fallback bright-line rule.

Additionally, since the court analyzed three separate theories of inequitable conduct for the same patent, the ruling could be interpreted as either endorsing inequitable conduct-by-multiple-minor-transgressions[140] or, alternatively, as including dicta as to two of the three theories. The court tacitly acknowledged these possibilities when it concluded that "[i]t is not necessary to decide whether any one of the three nondisclosures, standing alone, would have been sufficient to justify a judgment of unenforceability."[141]

138. 231 F.3d 1373 (Fed. Cir. 2000).

139. *McKesson*, 487 F.3d at 924.

140. *See* Nilssen v. Osram Sylvania, Inc., 504 F.3d 1223 (Fed. Cir. 2007) (endorsing a theory of inequitable conduct via a pattern of misconduct).

141. *McKesson*, 487 F.3d at 926.

D. *ESPEED*

In *eSpeed*,[142] the Federal Circuit raised the possibility that disclosing *too much* information to the PTO could result in a finding of inequitable conduct, on the theory that the applicant "buried" material information among irrelevant information. The patent in suit related to a computerized system for trading government securities. Two sets of rules, referred to as the "new rules" and "old rules," governed the process by which customers could increase their securities purchase volume.[143] The patentee had a prior art software program called the Super System.[144] During prosecution, three declarations were submitted to the examiner describing the Super System, but indicating that it did not include the "new rules." Over one thousand pages of exhibits accompanied the declarations, including portions of the Super System source code. The source code submitted in those exhibits demonstrated that the Super System did in fact accommodate the "new rules," contradicting the assertions made in the declarations. The applicant did not specifically point out these portions of the source code to the examiner.[145]

Although the Federal Circuit affirmed the district court's inequitable conduct ruling, the two courts differed in their approaches. In the *intent* portion of its inequitable conduct ruling, the district court ruled that submitting a "blizzard of paper" without pointing out the references to the "new rules" was "more consistent with an intent to hide than to disclose."[146] In the *materiality* portion of its opinion, the Federal Circuit "agree[d] with the district court that the 'blizzard of paper' submitted to the PTO . . . 'left the examiner with the impression that the examiner did not need to conduct any further . . . investigation.' "[147]

Implications: The Federal Circuit's shift of the "blizzard of paper" discussion from the analysis of intent to the analysis of materiality has led some to speculate that the Federal Circuit was signaling a revival of the doctrine of inequitable conduct by "burying." In other words, even though an applicant actually submitted a material reference to the examiner, the applicant could nonetheless be found to have committed inequitable conduct by "burying" the reference in a "blizzard of paper" in an effort to prevent the examiner from duly considering it.

---

142. eSpeed, Inc. v. Brokertec USA, L.L.C., 480 F.3d 1129 (Fed. Cir. 2007).
143. *Id.* at 1131–32.
144. *Id.* at 1132.
145. *Id.* at 1132–33.
146. eSpeed, Inc. v. Brokertec USA, L.L.C., 417 F. Supp. 2d 580, 598 (D. Del. 2006).
147. *eSpeed*, 480 F.3d at 1137 (citing Semiconductor Energy Lab. Co. v. Samsung Elecs. Co., 204 F.3d 1368, 1377 (Fed. Cir. 2000)).

Although some historical support exists for the "burying" doctrine,[148] the Federal Circuit rejected that doctrine in *Molins PLC v. Textron, Inc.*[149] Instead of a revival of "burying," the *eSpeed* ruling is better understood as a straightforward application of the holding in *Rohm & Haas*,[150] that a misstatement to the examiner cannot be "cured" simply by disclosing the correct information to the examiner without comment, and instead can be "cured" only by specifically identifying the misstatement and pointing out how the newly submitted information contradicts the prior misstatement. Moreover, by citing *Semiconductor Energy Laboratory Co. v. Samsung Electronics Co.*,[151] the court evinced some concern that, like references in foreign languages, the relevance of source code may not be readily accessible or apparent to an examiner without some additional explanation. Thus, an alternative explanation of *eSpeed* is that it requires source code submitted to the examiner to be treated like foreign language references, and is not an all-purpose revival of the "burying" doctrine.

E. *STAR SCIENTIFIC*

In *Star Scientific*,[152] decided in August 2008, the Federal Circuit provided a renewed focus on a narrow and clearly defined articulation of the intent element. At the same time, though, it continued to support the "reasonable examiner" standard for materiality. The court reversed a district court finding of inequitable conduct as to two patents.[153] For one patent, the Federal Circuit ruled that the threshold of materiality was not met.[154] For the other patent, the court ruled that the threshold of intent was not satisfied.[155] The pa-

148. *E.g.*, Rules of Practice in Patent Cases, 42 Fed. Reg. 5588, 5590 (Jan. 28, 1977) (analogizing Rule 56 to a Supreme Court decision relating to the SEC, in which "[t]he Court noted that the standard of materiality should not be so low that . . . the fear of liability would cause management 'simply to bury the shareholder in an avalanche of trivial information—a result that is hardly conducive to informed decision making.' "); U.S. PATENT & TRADEMARK OFFICE, U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 2002.03 (5th ed., rev. 3, 1986) ("[N]on-identification of an especially relevant passage buried in an otherwise less or non-relevant text could result in a holding of 'violation of duty of disclosure.' "). *See also* Glenn E. Von Tersch, *Curing the Inequitable Conduct Plague in Patent Litigation*, 20 HASTINGS COMM. & ENT. L.J. 421, 430–31 (1998) (discussing "burying" issue); Upadhye, *supra* note 80, at 715–17.

149. 48 F.3d 1172, 1183–84 (Fed. Cir. 1995). *But see* Cotropia, *supra* note 15, at 768 (advocating expanding inequitable conduct to include intentional "burying").

150. Rohm & Haas Co. v. Crystal Chem. Co., 722 F.2d 1556, 1572–73 (Fed. Cir. 1983).

151. 204 F.3d 1368, 1372, 1377 (Fed. Cir. 2000).

152. Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357 (Fed. Cir. 2008).

153. *Id.* at 1360.

154. *Id.* at 1370.

155. *Id.* at 1367–68.

tents in *Star Scientific* related to methods for curing tobacco that would reduce the amount of carcinogens known as tobacco specific nitrosamines (TSNAs) in the tobacco. The predominant method for tobacco-curing in the U.S. from the 1970s to the late 1990s involved burning fuel, typically propane, and blowing the hot exhaust gas directly on the tobacco. Tobacco cured in this manner had higher levels of TSNAs.[156]

Williams, an inventor at Star, developed methods of reducing the level of TSNAs in cured tobacco and engaged attorney Delmendo to prosecute the patents.[157] Burton, a Star consultant, sent Delmendo a letter relating his observations about lower TSNA levels in Chinese tobacco, which was cured using an older (radiant-heat) method.[158] Additionally, Williams sent Delmendo data from two U.S. farms that still used the radiant-heat method (the "Jennings data" and the "Curran data") and had reduced TSNA levels.[159] The Curran data was of a partially cured sample, and Williams's associate finished curing it with a microwave oven.[160] Delmendo filed several patent applications for Williams.[161] Thereafter, Delmendo was replaced as prosecution counsel by Rivard, an attorney from a different firm.[162] In the application that ultimately issued as U.S. Patent No. 6,202,649 (the '649 patent), Rivard filed a petition to make special, and included an Information Disclosure Statement (IDS) that did not disclose the Burton letter.[163] Rivard also filed a continuation application (which ultimately issued as U.S. Patent No. 6,425,401 (the '401 patent)), and filed a petition to make special and IDS in that application, again without disclosing the Burton letter.[164] During prosecution, Rivard became aware of the Burton letter and the Curran data but ultimately did not disclose the letter or the data because he concluded that neither was material.[165]

The district court held a bench trial on inequitable conduct, and ruled that both patents were unenforceable.[166] The Federal Circuit, in an opinion by Chief Judge Michel, provided an extended exposition of the elements of inequitable conduct, emphasizing—

156. *Id.* at 1361.
157. *Id.*
158. *Id.* at 1361–62.
159. *Id.* at 1362.
160. *Id.*
161. *Id.* at 1361–62.
162. *Id.* at 1363.
163. *Id.*
164. *Id.*
165. *Id.*
166. *Id.* at 1365.

- The need for a threshold level of both materiality and intent;[167]
- The court's discretion to balance the equities and determine that there was no inequitable conduct, even if both thresholds have been met;[168]
- The importance of strictly enforcing the clear and convincing burden of proof, since the penalty ("the loss of the entire patent even where every claim clearly meets every requirement of patentability") is so severe;[169]
- The separateness of materiality and intent, and of the proofs of these elements;[170]
- When intent is inferred from circumstantial evidence, it must be "the single most reasonable inference able to be drawn from the evidence to meet the clear and convincing standard;"[171] and
- The reasonable examiner standard for determining materiality, and the immateriality of information that is cumulative of information already before the PTO.[172]

Applying these principles, the Federal Circuit reversed, holding that the district court erred for two reasons in concluding that Williams and Star intended to deceive the PTO during prosecution of the '649 patent. First, defendant R.J. Reynolds (RJR) failed to adduce evidence sufficient to infer intent, and could not carry its burden by relying on the absence of a credible explanation by Star.[173] Second, the district court clearly erred in finding that the Burton letter and Curran data were material to the prosecution of the '401 patent, due to their cumulativeness with other information previously disclosed to the examiner.[174]

---

167. *Id.*

168. *Id.* at 1365, 1367.

169. *Id.* at 1365–66 (noting that the doctrine had its origins in fraud but has subsequently become more broad, covering lesser wrongful conduct, without any expansion of remedies to include lesser penalties for less culpable conduct).

170. *Id.* at 1366.

171. *Id.* This articulation accords with the holding in *Kingsdown* that circumstantial evidence must indicate sufficient culpability to "require"—not merely permit—a finding of intent. Kingsdown Med. Consultants, Ltd. v. Hollister Inc., 863 F.2d 867, 876 (Fed. Cir. 1988).

172. *Star Scientific*, 537 F.3d at 1367.

173. *Id.* at 1365, 1368.

174. *Id.* at 1365, 1370.

Implications: The extensive exposition of the inequitable conduct doctrine in *Star Scientific*, together with the opinion's admonitions about the severity of the penalty, the high burden of proof (particularly its comments about proof of intent), and the court's discretion to deny inequitable conduct even where both thresholds have been met, appears to be an effort to limit the influence of various earlier cases that may have stated the standards more loosely. Perhaps inconsistently with that purpose, the opinion again elects the "reasonable examiner" standard as the test for materiality—without any mention of the 1992 version of Rule 56 and its less-malleable, more objective definition of materiality. Also, it bears noting that the opinion is not an en banc ruling and therefore lacks the precedential authority to override inconsistent panel decisions from the Federal Circuit. This limited influence became apparent a month later, when the Federal Circuit issued the *Praxair* decision.

F. *PRAXAIR*

In September 2008, a month after *Star Scientific*, the Federal Circuit decided *Praxair, Inc. v. ATMI, Inc.*,[175] in which it apparently applied a test for materiality even lower and more malleable than the "reasonable examiner" standard, and also reiterated the *Ferring* "should have known" test for intent. The district court ruled that "the level of materiality of the [restricted flow orifice (RFO)] art is sufficiently high so as to support an ultimate finding of inequitable conduct."[176]

The Federal Circuit affirmed the unenforceability of one patent-in-suit. To support its materiality determination, the Federal Circuit did not articulate any of the established standards, instead holding merely that the "overall degree of similarity between the omitted reference and the claimed invention" could be used to determine materiality.[177] On this standard, the nondisclosed art was material.[178] The Federal Circuit then declined to consider Praxair's argument that the nondisclosed art was cumulative, concluding on a narrow reading of the record that Praxair had failed to raise the argument below.[179]

As for intent, the Federal Circuit specifically recited *Ferring*'s ruling that intent can be proven when the applicant should have known of the materiality of known, material art. The court then combined *Ferring* with the holding

175. 543 F.3d 1306 (Fed. Cir. 2008). Judge Dyk, who was on the *Star Scientific* panel and did not dissent, wrote the *Praxair* decision. The differences between the legal standards applied in these two cases cannot therefore be explained away as reflecting the views of two entirely different panels of the Federal Circuit.

176. *Id.* at 1313.

177. *Id.* at 1314–15.

178. *Id.* at 1314.

179. *Id.* at 1315.

in *Critikon*[180] that an applicant who should have known of the materiality of nondisclosed art "can expect to find it difficult to establish 'subjective good faith' sufficient to prevent" a finding of intent.[181] On that basis, the court rejected the prosecuting attorney's testimony and attempted good faith explanation for not having disclosed the RFO art, affirming the finding of inequitable conduct.[182]

Implications: *Star Scientific* arguably represented an effort by Chief Judge Michel to re-establish doctrinal and analytical rigor to inequitable conduct cases. A month after the *Star Scientific* opinion was issued, one of its panel members authored the *Praxair* opinion, which clearly embraces—and expands—the *Ferring* ruling as to intent, while ignoring the *Star Scientific* decision for both materiality and intent. The contrast between these two cases is strongly suggestive that, without some kind of authoritative restatement of the doctrine of inequitable conduct, various panels of the Federal Circuit will continue to offer up a smorgasbord of inequitable conduct rulings and doctrinal articulations to suit any taste.

These disparate rulings have wrought doctrinal uncertainty that will inexorably lead to over-assertion of the inequitable conduct defense. The implications of many of these cases that mere threshold levels of materiality and intent can regularly support determinations of unenforceability suggest that even if the statistical data does not clearly show a resurgent "plague" of inequitable conduct, at least at the Federal Circuit level, the combined enticements of a strong remedy and an uncertain (and perhaps expanding) legal standard mean the defense could grow even more popular.

It is unclear when the Federal Circuit will undertake another en banc review of the inequitable conduct doctrine—it has been over twenty years since *Kingsdown*. Therefore, perhaps, it is time to consider statutory reform of this judicially-created doctrine.

---

180. Critikon, Inc. v. Becton Dickinson Vascular Access, Inc., 120 F.3d 1253, 1257 (Fed. Cir. 1997).

181. *Praxair*, 543 F.3d at 1313–14.

182. *Id.* at 1317–18. Noting that a notice of allowability had already issued for the '609 patent at the time statements inconsistent with RFO art were made during prosecution of the '115 patent, the Federal Circuit reversed as to the '609 patent, finding neither materiality nor intent. *Id.* at 1318–19.

V. LEGISLATIVE PROPOSALS TO REFORM INEQUITABLE CONDUCT

In recent years, Congress has seen several proposals for legislative reform of the inequitable conduct doctrine.[183] The inequitable conduct provisions in these reform proposals have varied considerably in scope. In the 2007–2008 session of Congress, three patent reform bills were introduced, each of which included proposals to reform inequitable conduct.

Congress' proposed reform measures in all three bills generally fall within one of four broad categories: (1) specifying the prima facie elements of inequitable conduct, specifically materiality and intent; (2) specifying the standard of pleading or proof in litigation in the courts; (3) changing the range of remedies available in the courts; and (4) providing a forum within the PTO (rather than the courts) for adjudication of inequitable conduct allegations. This Article argues that reform should focus only on the first and third items, clarifying the prima facie elements and reforming the remedy. The second item, standards of pleading[184] or proof, is unambiguously established; however, the Federal Circuit's August 2009 ruling in *Exergen* may significantly increase the Rule 9(b) scrutiny given to inequitable conduct pleadings.[185] The fourth item, a new PTO forum, would cause unwarranted complexity, would

183. The current round of reform efforts started in earnest with the publication of the FTC and NAS reports in 2003 and 2004. *See supra* note 5. Patent reform legislation has been introduced into Congress more or less annually since 2005. Cotropia, *supra* note 15, at 737–41 (summarizing inequitable conduct reform proposals in 2005 and 2006 patent reform legislation); Kevin Mack, Note, *Reforming Inequitable Conduct to Improve Patent Quality: Cleansing Unclean Hands*, 21 BERKELEY TECH. L.J. 147, 156–61 (2006) (summarizing 2005 patent reform legislation).

184. A search of the Westlaw DCT database, conducted January 17, 2009, with the search terms (9(b) /s "inequitable conduct") yielded 140 district court cases, dating mainly from 1988 to the present. *E.g.*, Venetec Int'l, Inc. v. Nexus Medical, LLC, 541 F. Supp. 2d 612 (D. Del. 2008); Solarex Corp. v. Arco Solar, Inc., 121 F.R.D. 163 (E.D.N.Y. 1988). However, the Federal Circuit did not address the applicability of Rule 9(b) until 2003, when it indicated its applicability to inequitable conduct in dicta. Ferguson Beauregard/Logic Controls, Division of Dover Resources, Inc. v. Mega Systems, LLC, 350 F.3d 1327, 1344 (Fed. Cir. 2003). In 2007, the Federal Circuit squarely ruled that Rule 9(b)'s heightened pleading standard governs inequitable conduct. Central Admixture Pharmacy Servs., Inc. v. Advanced Cardiac Solutions, P.C., 482 F.3d 1347, 1356–57 (Fed. Cir. 2007) (affirming dismissal of inequitable conduct pleading due to insufficient particularity). In August 2009, the Federal Circuit again held that Rule 9(b) governs pleadings of inequitable conduct, and also set forth greater detail concerning what Rule 9(b) requires. Exergen Corp. v. Wal-Mart Stores, Inc., 575 F.3d 1312, 1326–29 (Fed. Cir. 2009). *But see* David Hricik, *Wrong About Everything: The Application by the District Courts of Rule 9(b) to Inequitable Conduct*, 86 MARQ. L. REV. 895 (2003) (arguing that Rule 9(b) should not apply to inequitable conduct).

185. *See supra* note 184.

increase administrative costs, and (as the MPEP acknowledges[186]) would require the PTO to investigate allegations of intent to deceive, which lies outside its area of expertise.

The inequitable conduct provisions of each of the recent patent reform bills are discussed in detail below.

A. LEAHY BILL, SENATE BILL 1145

In April 2007, Senators Leahy, Hatch, and eight other senators introduced Senate Bill 1145 (the Leahy bill).[187] Versions of the Leahy bill had been introduced in several previous sessions of Congress.[188] The Senate Judiciary Committee Report supporting the Leahy bill acknowledges that earlier iterations of the bill did not include provisions relating to inequitable conduct,[189] but that three concerns prompted its inclusion: (1) the absence of a clear standard of materiality in the Federal Circuit,[190] (2) the collapse of the intent element into materiality,[191] and (3) the courts' lack of discretion in selecting a remedy.[192]

To remedy these three problems, the Leahy bill proposed to add a new § 298 to the Patent Act. The Leahy bill's proposed § 298 consists of five subsections. Subsection (a) would codify the prima facie elements of inequitable conduct and the requirement that inequitable conduct be proved by clear and convincing evidence; subsection (b) would define materiality using the "reasonable examiner" standard; subsection (c) would permit intent to be inferred, but prohibit proof of intent via "gross negligence" or materiality; subsection (d) would codify the requirement that inequitable conduct be pled with the particularity of Rule 9(b) of the Federal Rules of Civil Procedure;

---

186. *See supra* note 42 and accompanying text.

187. S. 1145, 110th Cong. (as reported in Senate, Apr. 18, 2007).

188. *E.g.*, S. 3818, 109th Cong. (2006).

189. S. REP. NO. 110-259, at 32 (2008).

190. The Senate Report states,

> First, the Federal Circuit has failed to establish one clear standard of materiality for inequitable conduct purposes. Having multiple materiality standards is hardly helpful to the district courts that are charged with making inequitable conduct determinations in the first instance, and patent holders are left with less than clear guidance about what they should disclose to the USPTO.

Id. The committee report specifically calls out *Digital Control* to illustrate the lack of a single clear standard. *Id.* at 32 n.151.

191. *Id.* ("Second, direct evidence of an intent to deceive is uncommon, so some courts collapse the issue of intent into the issue of materiality, so that intent to deceive is often inferred from materiality.").

192. *Id.* ("Third, if inequitable conduct is found, judges have no discretion as to the remedy—no claim of the patent can ever be enforced against anyone.").

and subsection (e) would create several new alternative remedies and grants the district court discretion to select one or more of those remedies.[193]

The Leahy bill proposes no substantive changes to the clear and convincing standard and the requirement for particularized pleading in compliance with Rule 9(b) of the Federal Rules of Civil Procedure,[194] as those principles are well-established in the case law. The Judiciary Committee Report suggests a belief, however, that the heightened pleading requirement and "clear and convincing" proof standard are novel additions to the law of inequitable conduct that "presumably" will ameliorate "concerns that inequitable conduct is 'over plead [sic].' "[195] Contrary to the Judiciary Committee's optimism, it is difficult to conceive how codification of existing law will "ameliorate" existing problems.[196]

Each of the three other provisions of the Leahy bill's proposal for inequitable conduct requires more detailed discussion.

First, the Leahy bill would codify the definition of materiality using the vague and subjective "reasonable examiner" standard that was set forth in the 1977 version of PTO Rule 56,[197] and that the Federal Circuit continues to identify as the lowest threshold for materiality.[198] In selecting this standard, the Judiciary Committee noted that some cases "appear to emphasize improperly the first part of this definition (reasonably important to an examiner) without giving necessary consideration to the latter part of the definition (in deciding whether to allow the patent)."[199] This, the Committee believed, essentially reduced the materiality standard to a "relevancy standard."[200] The Committee expressed the hope that codification of the standard would force

193. *Id.* at 32–33 (summarizing proposed 35 U.S.C. § 298).

194. *See supra* note 184.

195. S. REP. NO. 110-259, at 33 n.156 (2008).

196. Anecdotally, however, the application of Rule 9(b) to inequitable conduct pleadings has been sporadic. Therefore, codification of the heightened pleading requirement may serve to increase its enforcement. In contrast with the un-elaborated Rule 9(b) pleading requirement proposed in the Leahy bill, the Federal Circuit has recently articulated in substantial detail what will be required to plead inequitable conduct under Rule 9(b). *See* Exergen Corp. v. Wal-Mart Stores, Inc., 575 F.3d 1312, 1328–29 (Fed. Cir. 2009).

197. 37 C.F.R. § 1.56 (1977) ("[I]nformation is material where there is a substantial likelihood that a reasonable examiner would consider it important in deciding whether to allow the application to issue as a patent."). The Leahy bill would also codify the requirement that material information be non-cumulative. S. 1145, 110th Cong. § 12 (as reported in Senate, Apr. 18, 2007) (proposing to enact 35 U.S.C. § 298(b)(2)). The non-cumulativeness requirement is a straightforward codification of existing law that has not attracted controversy.

198. Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1314–16 (Fed. Cir. 2006).

199. S. REP. NO. 110-259, at 33 n.155 (citing *Nilssen*, *McKesson*, and *Ferring*).

200. *Id.*

courts to apply the entire definition,[201] thereby tightening the scope of materiality over that found in existing cases. In adopting this definition of materiality, the Committee specifically rejected a proposal by Senator Hatch that would have defined materiality in a manner more similar to the 1992 version of PTO Rule 56.[202] Senator Leahy proposed the "reasonable examiner" standard as a modification of Senator Hatch's amendment, and Senator Leahy's version prevailed.[203] In response, Senators Specter and Hatch registered their disagreement.[204] Senators Specter and Hatch stated that, under the "reasonable examiner" standard, "virtually any information can be characterized as 'material.' "[205] They cite *Nilssen*, *McKesson*, and *Ferring* as examples of this unacceptable result.[206] Accordingly, they stated, "[w]e do not support the ambiguous language reported by the Committee" defining materiality under the "reasonable examiner" standard, adding that it "does not improve current law," and calling the codification of that standard an "unworkable solution."[207] Specifically, they suggested that the "reasonable examiner" standard impedes the patent examination process by encouraging applicants to submit too much information to the examiner.[208] Accordingly, they called for an "objective threshold" of materiality.[209] However, they did not provide proposed language for a competing proposal, other than to suggest that materiality ought to be limited to "information that can affect the validity of a patent claim."[210] In this respect, Senators Specter and Hatch seem to sympathize with the adoption of the 1992 version of Rule 56's definition of materiality.[211]

Second, the Leahy bill's proposed definition of intent states,

---

201. *Id.*

202. S. REP. NO. 110-259, at 40 ("Senator Hatch offered an amendment that would codify and raise the standard to prove inequitable conduct, including defining materiality as information that [if] considered would render a claim of the patent invalid."). Compare this formulation with 37 C.F.R. § 1.56(b) (1992), which states,

> [I]nformation is material to patentability when it is not cumulative to information already of record or being made of record in the application, and (1) It establishes, by itself or in combination with other information, a prima facie case of unpatentability of a claim; or (2) It refutes, or is inconsistent with, a position the applicant takes in: (i) Opposing an argument of unpatentability relied on by the Office, or (ii) Asserting an argument of patentability.

203. S. REP. NO. 110-259, at 32, 40.

204. *Id.* at 59–62.

205. *Id.* at 60.

206. *Id.* at 60, nn. 9, 10.

207. *Id.* at 60–61.

208. *Id.* at 61.

209. *Id.*

210. *Id.*

211. *See* 37 C.F.R. § 1.56 (1992).

> Intent to deceive the Office may be inferred under subsection (a), [which sets forth the prima facie elements of inequitable conduct,] but the inference may not be based solely on the gross negligence of the patent owner or its representative, or on the materiality of the information misrepresented or not disclosed.[212]

While this language essentially tracks the ruling in *Kingsdown*, it also appears to be an attempt to legislatively overrule more recent cases such as *Ferring*[213] and *McKesson*.[214] However, this language does not propose to define what evidence may be used to prove intent, or what evidence would be sufficient to prove intent. Nor does this language address the "should have known" standard, or preclude all use of gross negligence or evidence of materiality to prove intent; rather, intent may not be "solely" based on such evidence. In short, although this definition would potentially override some of the more extreme cases, it would codify in statute many of the problems and ambiguities that are found in the current doctrine. Senators Specter and Hatch, by contrast, would draw a bright-line prohibition on the use of materiality as evidence of intent: "[W]e believe intent must be proven with independent evidence separate from and unrelated to the materiality of the information at issue."[215] The addition of language to the same effect as *Star Scientific* would considerably improve this provision—if circumstantial evidence is used to prove intent, such intent must be the single most compelling inference from all the evidence (including evidence of good faith) to satisfy the clear and convincing burden of proof.

Third, the Leahy bill would empower the courts with the discretion to award one or more of a range of new remedies when inequitable conduct has been proven.[216] In addition to the traditional remedy of holding the entire patent unenforceable, the Leahy bill would permit courts to hold individual patent claims unenforceable, or to limit remedies to a reasonable royalty while denying equitable relief and lost profits damages.[217] The Leahy bill says nothing about the doctrine of infectious unenforceability. Nor does it say anything about the balancing element in the traditional inequitable conduct doc-

---

212. S. 1145, 110th Cong. § 12 (as reported in Senate, Apr. 18, 2007) (proposed 35 U.S.C. § 298(c)).

213. Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181, 1191 (Fed. Cir. 2006) (affirming summary judgment as to intent because applicant "should have known" of materiality of information).

214. McKesson Info. Solutions, Inc. v. Bridge Med., Inc., 487 F.3d 897, 924–25 (Fed. Cir. 2007) (inferring intent from fact of nondisclosure and representations concerning the prior art).

215. S. REP. NO. 110-259, at 61.

216. S. 1145 § 12 (proposing to enact 35 U.S.C. § 298(e)).

217. *Id.*

trine. Although the Report of the Judiciary Committee does not say so, the introduction of a range of alternative remedies could be viewed as an alternative implementation of the balancing prong of the traditional inequitable conduct defense. So viewed, and considered with reference to Figure 1, the Leahy bill would actually expand the conduct for which an inequitable conduct remedy can be awarded, from the dark area above and to the right of the curve in Figure 1, to also include the striped area, between the curve and the dashed lines in the middle of the graph. That is, if the thresholds of materiality and intent are satisfied, the Leahy bill would appear to deem that sufficient for a finding of inequitable conduct, without the additional step of equitable balancing. The possibility of a lesser remedy may actually create *added* incentives for accused infringers to allege inequitable conduct on thin evidence, as infringers could still win a lesser remedy on such evidence, whereas infringers would get no remedy under current standards. So viewed, the expansion of conduct for which some remedy is available (particularly low-materiality, low-intent conduct) will likely *increase* the incentives of litigants to allege inequitable conduct—precisely the opposite of the bill's intent. Senators Specter and Hatch recognized this possibility, questioning whether "expanding the range of available sanctions for inequitable conduct in the absence of other meaningful changes to the doctrine will encourage more, not less, inequitable conduct litigation."[218]

The Leahy bill was not approved in the Senate during the 110th Congress. A version of the Leahy bill was reintroduced in the 111th Congress.[219] However, as introduced and as reported in the Senate on April 2, 2009, it did not include provisions for inequitable conduct.[220] Co-sponsor Senator Hatch has indicated a desire to include inequitable conduct provisions in the bill.[221]

B. BERMAN BILL, HOUSE BILL 1908

Representative Berman introduced a version of the Leahy bill in the House of Representatives in April 2007.[222] The House approved the bill in September 2007. Because the inequitable conduct provisions are similar in structure to the Senate version of the bill, this Article discusses only the differences between the House and Senate versions. The Report of the House

218. S. REP. NO. 110-259, at 61.

219. S. 515, 111th Cong. (as introduced in Senate, Mar. 3, 2009).

220. *Id.*; S. 515, 111th Cong. (as reported in Senate, Apr. 2, 2009).

221. Press Release, Orrin Hatch, Senator, Senators Hatch, Leahy Introduce Patent Reform Act of 2009 (Mar. 3, 2009) (on file with author), *available at* http://hatch.senate.gov/public/index.cfm?FuseAction=PressReleases.Detail&PressRelease_id=ce28c6f0-1b78-be3e-e028-418ea18126e5.

222. H.R. 1908, 110th Cong. (2007).

Committee on the Judiciary stated that a desire to address "abusive" litigation practices motivated the reform of the inequitable conduct doctrine,[223] specifically citing "overuse" of inequitable conduct.[224] The specific defects in existing doctrine, as identified by the House committee report, include the proliferation of standards of materiality[225] and the commingling of the materiality and intent prongs.[226] Because these defects lead to improper overuse of the inequitable conduct defense and introduce uncertainty into patent litigation, the stated purpose of House Bill 1908 is "to provide an increased level of certainty to the defense."[227]

The House version of the bill would add the substantive provisions of inequitable conduct as a new § 282(c) in the Patent Act.[228] Each of the three main components discussed above is different in the House version.

Unlike the Senate version, the House version does not define "material information."[229] Instead, the House version would establish a "but for" test for causation: "in the absence of such deception, the Office, acting reasonably, would, on the record before it, have made a prima facie finding of unpatentability."[230] As the Federal Circuit noted in *Digital Control*, this "but-for" standard is the narrowest test for materiality among the various tests articulated by the courts.[231] It is worth noting that "material information" is not the only term of art that the House leaves undefined in this section. The House version uses, but fails to define the following terms: "duty of disclosure," "person with a duty of disclosure," and "prima facie finding of unpatentability."[232] All three terms are defined in the current version of PTO Rule 56.[233]

---

223. H.R. REP. NO. 110-314, at 20 (2007).

224. *Id.* at 21.

225. *Id.* at 42 (citing Digital Control v. Charles Mach. Works, 437 F.3d 1309 [incorrectly identified as Digital Control v. Merlin Technology]).

226. *Id.* at 42–43. In this regard, however, the House committee report conflates two separate issues. The Senate Judiciary Committee also notes the problem with using materiality to prove intent. The balancing of materiality and intent, once both elements have been found to exist at a threshold level, is an aspect of the inequitable conduct doctrine that may be valuable independently of the identification of evidence that may be used to establish intent. *See supra* Figure 1.

227. *Id.* at 43, 85.

228. *Id.* at 18, 116–17.

229. *Id.* at 116–17 (proposing to enact 35 U.S.C. § 282(c)). Nor does it expressly require that the withheld information be non-cumulative.

230. *Id.* at 117 (proposing to enact 35 U.S.C. § 282(c)(1)(B)).

231. *See* Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1314–16 (Fed. Cir. 2006).

232. *See* H.R. REP. NO. 110-314, at 116–17 (proposing to enact 35 U.S.C. § 282(c)).

233. 37 C.F.R. § 1.56 (2008). Rule 56 specifies that individuals having a duty of disclosure must "disclose to the Office all information known to that individual to be material to

Ordinarily, in matters of statutory interpretation where an administrative agency (here, the PTO) has a substantive connection to the statute (here, the Patent Act), it would be reasonable to infer that Congress intended the PTO's definitions of those terms to govern, and perhaps that is a reasonable way to interpret the statute. However, because the bill largely failed to discuss the current version of Rule 56, and because the Federal Circuit in *Digital Control* essentially overrode any expectation that the 1992 version of Rule 56 would supplant judicial definitions of materiality,[234] it may not be reasonable to expect that the courts would interpret the House version of the bill in accordance with Rule 56, unless expressly directed by Congress to do so.

Concerning intent, the House version mirrors the Senate version in requiring that the facts used to prove intent be "beyond" those used to prove materiality, apparently without prohibiting such facts as a component of the evidence of intent.[235] The House version does not adopt the Senate version's disavowal of "gross negligence" as a method of proving intent, but instead provides an affirmative definition of how to satisfy the intent standard, requiring "conscious or deliberate behavior."[236]

Third, the House version provides a similar menu of remedies for inequitable conduct.[237] The option to hold the entire patent unenforceable is the same. The House version would more narrowly limit the discretion to hold individual claims unenforceable, limiting the discretion to the claims-in-suit or the claims in which the inequitable conduct occurred.[238] The House ver-

---

patentability as defined in this section." 37 C.F.R. § 1.56(a). Individuals having a "duty of disclosure" are enumerated in Rule 56(c). 37 C.F.R. § 1.56(c). Rule 56 defines material information as information that

> is not cumulative to information already of record or being made of record in the application, and [that either] establishes, by itself or in combination with other information, a prima facie case of unpatentability of a claim; or . . . refutes, or is inconsistent with, a position the applicant takes in: opposing an argument of unpatentability relied on by the Office, or [a]sserting an argument of patentability.

37 C.F.R. § 1.56(b). A "prima facie case of unpatentability" exists "when the information compels a conclusion that a claim is unpatentable . . . giving each term in the claim its broadest reasonable construction consistent with the specification, and before any consideration is given to [rebuttal evidence submitted to establish patentability]." *Id.*

234. *See supra* Section IV.B.

235. H.R. REP. NO. 110-314, at 117 (proposing to enact 35 U.S.C. § 282(c)(2)).

236. *Id.* ("Facts support an inference of intent if they show circumstances that indicate conscious or deliberate behavior on the part of the patentee, its agents, or another person with a duty of disclosure to the Office, to not disclose material information or to submit materially false information.").

237. *Id.* at 117 (proposing to enact 35 U.S.C. § 282(c)(3)).

238. As a practical matter, this distinction between the House and Senate versions may not make much difference.

sion would permit the court merely to deny equitable relief, and would not deprive the patentee of the option to seek lost-profits damages.[239]

Fourth, in contrast to the silence of the Senate version, the House version expressly addresses infectious unenforceability, and would authorize the court to hold claims of a related patent unenforceable.[240]

Finally, the House version includes a short section requiring the court to refer inequitable conduct findings involving patent attorneys to the PTO for "appropriate disciplinary action."[241]

A version of the Berman bill was reintroduced by Congressman Conyers in the 111th Congress.[242] However, as introduced, the 2009 version of the House bill did not include provisions for inequitable conduct.[243]

C. KYL BILL, SENATE BILL 3600

In September 2008, Senator Kyl introduced Senate Bill 3600, a patent reform bill intended to compete with Senate Bill 1145, the Leahy bill. Like Senator Leahy's bill, the Kyl proposal includes provisions on inequitable conduct. But the Kyl bill takes a significantly different approach. To handle inequitable conduct, the Kyl bill would create a new judicial procedure and two new administrative forums within the PTO. Courts would be prohibited from addressing inequitable conduct in patent litigation (or other litigation), except to pass on a threshold motion that would trigger the commencement of administrative proceedings. A reissue proceeding would address the impact of possible inequitable conduct on the patent itself, and the perpetrators of the inequitable conduct would be subjected to a separate disciplinary proceeding.

The Kyl bill's inequitable conduct reform provisions are found in section 11. They would add two new sections to the Patent Act: § 298, titled "Inequitable Conduct"; and § 299, titled "Civil sanctions for misconduct before the Office."

Section 298(a) starts with a claim to statutorily occupy the field of inequitable conduct: "Except as provided under this section or section 299, a patent shall not be held invalid or unenforceable based upon misconduct before the Office."[244] This is an evident effort to preempt the Federal Circuit's

239. H.R. REP. NO. 110-314, at 117–18.
240. *Id.* at 117
241. *Id.* (proposing to enact 35 U.S.C. § 282(c)(4)).
242. H.R. 1260, 111th Cong. (as introduced in Congress, Mar. 3, 2009).
243. *Id.*
244. Patent Reform Act of 2008, S. 3600, 110th Cong. § 11 (2008) (as introduced in Senate, Sept. 25, 2008) (proposing to enact 35 U.S.C. § 298(a), first sentence).

position that inequitable conduct, as a judicially created doctrine, exists independently of Patent Office regulations.[245] Section 298(a) continues by eliminating any claim or defense of inequitable conduct in civil litigation: "Nothing in this section shall be construed to create a cause of action or a defense in a civil action."[246] Thus, § 298(a) wipes the slate clean. With this bill, there would be neither a judicially-created doctrine of inequitable conduct nor a claim or defense of inequitable conduct in civil litigation.

Curiously, then, the first part of § 298(b) establishes motion practice in a civil action as a gating event to further inequitable conduct proceedings before the PTO.[247] But if, under § 298(a), there is no claim or defense of inequitable conduct that may be asserted in a civil action, it is not clear whether a court in a civil action could have subject-matter jurisdiction to entertain such a motion. Moreover, § 298(b) contains no suggestions or limitations on the nature of civil action in which such a motion could be brought. This omission implies that such a motion could be brought in any civil action, not just patent infringement actions and related actions for declaratory relief. Indeed, unless the patent issue arises as part of a well-pled complaint, the federal courts' exclusive jurisdiction over patent cases[248] would not necessarily preclude such a motion from being brought in a state court action.[249]

Assuming a court in a civil action has jurisdiction to entertain such a motion, the Kyl bill provides a starkly different process. First, in ruling on the motion, the court would be required to make its findings only by a preponderance of the evidence[250] rather than the clear and convincing evidence required under the current doctrine.[251]

The prima facie elements to be found by the court bear a passing resemblance to the current judicially-created doctrine of inequitable conduct, as well as Rule 56.[252] But the bill contains important differences. In this regard,

245. *See* Digital Control Inc. v. Charles Mach. Works, 437 F.3d 1309, 1315 (Fed. Cir. 2006).

246. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(a), second sentence).

247. *Id.* (proposing to enact § 298(b)(1)(A)) ("If a court in a civil action, upon motion of a party to the action . . .").

248. 28 U.S.C. § 1338 (2006) (stating that federal district courts have exclusive jurisdiction for patent cases).

249. *See, e.g.*, Holmes Group, Inc. v. Vornado Air Circulation Sys., Inc., 535 U.S. 826 (2002); Christianson v. Colt Indus. Operating Corp., 486 U.S. 800 (1988).

250. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(b)(1)(A)) (instructing the court to order the patent to be made the subject of a reissue application if it finds the applicant "more likely than not . . . intentionally deceived the Office . . . .").

251. Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357, 1365 (Fed. Cir. 2008).

252. 37 C.F.R. § 1.56 (2008).

the Kyl bill appears to have drawn inspiration from the PTO's Code of Professional Responsibility.[253] In particular, that code provides that a "practitioner" shall not "[k]nowingly giv[e] false or misleading information or knowingly participat[e] in a material way in giving false or misleading information, to . . . [t]he Office or any employee of the Office."[254] The Kyl bill would require proof that "a person who participated in a matter or proceeding before the Office knowingly and intentionally deceived the Office by concealing material information or by submitting false material information in such matter or proceeding . . . ."[255]

The differences between the Kyl bill and existing doctrine include the following. First, this provision reaches a broader range of people than practitioners covered by 37 C.F.R. § 10.23 and people identified in Rule 56(c) as owing a duty of disclosure to the Patent Office.[256] Second, in contrast with the case law's requirement that the person accused of inequitable conduct act with intent to deceive the Patent Office, the Kyl bill parallels 37 C.F.R. § 10.23 and requires that the person act "knowingly and intentionally."[257] Third, the existing doctrine requires only intent to deceive, while the Kyl bill imposes the additional requirement that the Patent Office in fact was deceived.[258] Fourth, Rule 56 imposes a duty of disclosure, such that a mere failure to disclose material information is sufficient to satisfy the materiality prong;[259] by contrast, the Kyl bill requires "concealing" material information. Fifth, there is no requirement that materiality and intent be balanced; apparently under the Kyl bill (as with the Leahy bill), if both materiality and intent are satisfied at a threshold level, the burden of proof has been met. Finally, the district court's discretion is constrained—if it makes the pertinent find-

253. *See, e.g.*, 37 C.F.R. § 10.23 (2008); 37 C.F.R. § 11.18(b)–(c) (2008); U.S. PATENT & TRADEMARK OFFICE, U.S. DEP'T OF COMMERCE, MANUAL OF PATENT EXAMINING PROCEDURE § 410 (8th ed., rev. 7, 2008).

254. 37 C.F.R. § 10.23(c)(2)(ii).

255. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(b)(1)(A)).

256. *See* 37 C.F.R. § 1.56(c); 37 C.F.R. § 10.23.

257. *See* 37 C.F.R. § 10.23(d) (defining knowledge to include "reckless indifference," "half-truths," and "concealment of material facts").

258. In view of the prohibition on examiner testimony, *supra* note 27, proof of actual deception (other than evidence found in the prosecution history) would be difficult or impossible.

259. 37 C.F.R. § 1.56(b).

ings, it "shall" order the commencement of a reissue proceeding,[260] and the district court's ruling "shall not be subject to appellate review."[261]

There are also some ways in which this part of the Kyl bill is clearly narrower than existing doctrine. For instance, building on the proposition that a claim of inequitable conduct must be pled with particularity,[262] the Kyl bill requires an explanation as to how the withheld information would invalidate one or more claims of the patent, but does not require any particularity in the motion's allegations of materiality, nondisclosure, intent to deceive, and actual deception of the Patent Office. Relatedly, proposed § 298(b)(1)(B) defines materiality narrowly in relation to patentability. It incorporates portions of Rule 56, requiring that the material be noncumulative and not already of record.[263] And it is limited to information that establishes nonpatentability or refutes an argument that an applicant made in support of patentability of a patent claim. This is narrower than the current law because it largely tracks the 1992 version of Rule 56,[264] abrogating the 1977 "reasonable examiner" standard that the Federal Circuit re-endorsed in *Digital Control*, and it is limited to issues of patentability, abrogating such recent cases as *Ferring*,[265] *McKesson*,[266] *Nilssen*,[267] and even *Star Scientific*.[268]

With regard to the Kyl bill's proposed use of a reissue proceeding to determine the impact of inequitable conduct on the patent itself, the bill limits the impact of the reissue proceeding on pending litigation. Once the reissue

260. The enforcement mechanism is in proposed section 298(b)(5). If the patentee fails to "seek reissue within 2 months of the court's order, the court shall enter judgment that the patent is unenforceable." S. 3600, at § 11.

261. *Id.* (proposing to enact 35 U.S.C. § 298(b)(1)(A)).

262. FED. R. CIV. P. 9(b); Central Admixture Pharmacy Servs., Inc. v. Advanced Cardiac Solutions, P.C., 482 F.3d 1347, 1356 (Fed. Cir. 2007) (citing Ferguson Beauregard/Logic Controls, Inc. v. Mega Sys., LLC, 350 F.3d 1327, 1344 (Fed. Cir. 2003)).

263. 37 C.F.R. § 1.56(b).

264. The Kyl bill's wording closely tracks Rule 56, but with some key omissions. While Rule 56 permits the analysis of unpatentability to be made with reference to the omitted information alone or in combination with other information of record in the application, the Kyl bill requires that that the omitted information—apparently alone—establishes unpatentability. *Compare* 37 C.F.R. § 1.56(b)(1), *with* S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(b)(1)(B)). This is, essentially, the difference between the materiality of omitted information for purposes of § 103 obviousness and § 102 anticipation. Also, the Kyl bill omits from its definition of materiality information that is inconsistent with arguments the applicant made in asserting patentability. *Compare* 37 C.F.R. § 1.56(b)(2)(ii), *with* S.3600, at § 11 (proposing to enact 35 U.S.C. § 298(b)(1)(B)).

265. Ferring B.V. v. Barr Labs., Inc., 437 F.3d 1181 (Fed. Cir. 2006).

266. McKesson Info. Solutions, Inc. v. Bridge Med., Inc., 487 F.3d 897 (Fed. Cir. 2007).

267. Nilssen v. Osram Sylvania, Inc., 440 F. Supp. 2d 884 (N.D. Ill. 2006), *aff'd*, 504 F.3d 1223 (Fed. Cir. 2007).

268. Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357 (Fed. Cir. 2008).

proceeding has commenced, the court may not stay the civil action due to the pendency of the reissue, unless and until the Patent Office rejects one or more claims as to which allegations of infringement are still-pending and the court determines that such a stay would be in the interests of justice.[269]

Under proposed § 298(d), the patentee may file a reissue application, omitting one or more claims. This would allow the patentee to "scrub" the offending claim out of the patent.

This procedure will potentially burden the PTO with a new caseload of many hundreds (at least) of additional reissue petitions each year.[270] The current caseload is about one thousand reissue petitions filed each year.[271] Therefore, assuming that the number of inequitable conduct-inspired reissue proceedings parallels the number of district court cases in which inequitable conduct is pled (see Table 1), the Kyl proposal could result in a doubling of the reissue caseload. With the existing caseload, the average time to first action of all patent applications[272] is 25.3 months, and the average time to final resolution is 31.9 months.[273] Yet the Kyl bill would require a first action within three months of the filing for the reissue application,[274] and completion of the reissue proceeding within one year of the initial notification date (i.e., 15 months from commencement).[275] In light of estimates that the PTO's reissue case load may double, it is unrealistic to expect the PTO to cut its time to first action by 88% (from 25.3 months to 3 months) and its average pendency by more than 50% (from 31.9 months to 15 months) without dedicating massive additional examination resources to inequitable conduct-prompted reissue proceedings.

If at least one claim comes out of the reissue proceeding,[276] "no further sanctions may be imposed against the patentee,"[277] except criminal, antitrust, and PTO-imposed sanctions, as described in proposed § 298(h). That is, a district court can hold a patent-in-suit unenforceable only if the patentee fails to initiate reissue proceedings within two months of the court's order to do

269. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(b)(4)).

270. *See supra* Table 1.

271. UNITED STATES PATENT AND TRADEMARK OFFICE, PERFORMANCE AND ACCOUNTABILITY REPORT: FISCAL YEAR 2007, at 110 (2007) (table 2).

272. PTO statistics on reissue applications alone are not available.

273. UNITED STATES PATENT AND TRADEMARK OFFICE, PERFORMANCE AND ACCOUNTABILITY REPORT: FISCAL YEAR 2007, at ii (2007).

274. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(e)(1)).

275. *Id.* (proposing to enact 35 U.S.C. § 298(e)(3)).

276. The procedural details of the Kyl bill's proposed reissue proceeding are not addressed here and are not compared in this Article with preexisting reissue procedures.

277. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 298(g)).

so, and cannot otherwise impose any sanction or remedy for inequitable conduct.[278] As noted below, however, one of the sanctions under the new proposed § 299 is a determination that one or more patent claims is unenforceable;[279] it is unclear whether a successful reissue petition will override this potential sanction.

Section 299(a) instructs the PTO to develop a new administrative disciplinary procedure covering "parties to a matter or proceeding before the Office" that may have engaged in inequitable conduct.[280] This procedure appears to be an attempt to implement a practice that the PTO abandoned in the late 1980s due, at least in part, to the PTO's lack of expertise in determining intent.[281] The administrative procedure in § 299 consists of several steps.

First, the PTO determines that probable cause exists if "[one] or more individuals or parties engaged in misconduct consisting of intentionally deceptive conduct of a material nature in connection with a matter or proceeding before the Office."[282] This determination is unreviewable.[283] The test for probable cause is different from and potentially much broader than the test proposed in § 298: the intent to deceive need not be directed at deceiving the PTO; the test could include conduct other than a failure to disclose material information or a misrepresentation of material information; and the language "material nature" does not track "material information," the term defined in § 298.

Second, there must be notice and an opportunity for a hearing, resulting in a determination within one year of the probable cause finding.[284] If the PTO finds misconduct, it may levy a civil penalty of up to $150,000 for each act of misconduct and up to $1 million for a "pattern of misconduct." To put these amounts in perspective, consider the fact that the typical patent application costs only $8,500 to $15,500 to prosecute, including filing fees and the fees of the prosecuting attorney.[285] Thus, under the Kyl bill, the penalty imposed on a prosecuting attorney for a first offense could be as much as ten to twenty times the fee that lawyer earned for prosecuting the patent. Particularly in the absence of a clear, applicationally self-sufficient standard of conduct,

278. *Id.* (proposing to enact 35 U.S.C. § 298(b)(5)).
279. *Id.* at § 11 (proposing to enact 35 U.S.C. § 299(b)(3)(C)(iii)(I)).
280. *Id.* (proposing to enact 35 U.S.C. § 298(a)).
281. *See supra* note 42.
282. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 299(b)(1)).
283. *Id.*
284. *Id.* (proposing to enact 35 U.S.C. § 299(b)(2)).
285. Oversight Hearing on the United States Patent and Trademark Office: Hearing Before the H. Subcomm. on Courts, the Internet, and Intellectual Property, Comm. on the Judiciary, 110th Cong. (2008) (statement of Alan J. Kasper, First Vice President, AIPLA).

many patent prosecutors could find this peril to be a prohibitive risk. But the Kyl bill provides even more possible fines. If the misconduct is deemed "exceptional" and was practiced by or at the behest of a real party in interest in a patent application, the penalty may include a determination that one or more claims is unenforceable and an additional penalty of up to $10 million may be assessed.[286] Also, anyone found to have engaged in misconduct can be held jointly and severally liable.[287] The Attorney General is authorized to file collection actions to recover penalties in the Eastern District of Virginia.[288]

In connection with both the probable cause phase and the determination phase, the PTO is empowered to gather whatever evidence it "determines pertinent" that is in the possession of "any person."[289] This new, general power of investigation would be a significant expansion of the powers of, and burdens on, the PTO.

The Kyl bill does not specify the standard of proof required to establish such misconduct in the new administrative proceeding. It does, however, provide a right of appeal directly to the Federal Circuit.[290] Determinations of misconduct are reviewed for substantial evidence and legal correctness.[291] Sanctions are reviewed for abuse of discretion.[292]

The Kyl bill was not approved in the Senate during the 110th Congress. A version of the Kyl bill was reintroduced in the 111th Congress.[293] It includes only the civil sanctions provisions for inequitable conduct (proposed § 299) that were included in the 2008 version of the Kyl bill.[294]

VI. REFORM PROPOSALS

This Article advocates four key reforms[295] to the doctrine of inequitable conduct:

---

286. S. 3600, at § 11 (proposing to enact 35 U.S.C. § 299(b)(3)(C)).
287. *Id.*
288. *Id.* (proposing to enact 35 U.S.C. § 299(b)(3)(F)(i)).
289. *Id.* (proposing to enact 35 U.S.C. § 299(c)(2)(A)).
290. *Id.* (proposing to enact 35 U.S.C. § 299(d)(1)).
291. *Id.* (proposing to enact 35 U.S.C. § 299(d)(4)).
292. *Id.*
293. Patent Reform Act of 2009, S. 610, 111th Cong. (as introduced in Senate, Mar. 17, 2009).
294. *Id.* at § 11 (proposing to enact 35 U.S.C. § 299).
295. Others have proposed alternative reforms. *See* Cotropia, *supra* note 15, at 43–57 (advocating (1) a broad definition of materiality, (2) requiring proof of intent to be separate from materiality, (3) discouraging "burying," and (4) a narrower remedy); Lisa A. Dolak, Inequitable Conduct: A Flawed Doctrine Worth Saving at 7–12 (Sept. 23, 2008) (unpublished manuscript, on file with the author), *available at* http://works.bepress.com/lisa_dolak/4/ (advocating (1) adopting the PTO standard of materiality, (2) tighter standards for proof of

1. Materiality should be explicitly linked to the PTO's 1992 definition of materiality in its regulations.

2. The standard for proving intent should be clearly defined.

3. The step of "balancing" materiality and intent should be clarified and codified, making clear that even if thresholds of materiality and intent exist, the court retains equitable discretion to decline to find inequitable conduct.

4. There should be a single, narrowed remedy. Only claims directly affected by the inequitable conduct (rather than the entire patent) should be deemed unenforceable.

First, concerning both materiality and intent, any codification should make clear that, for any particular allegedly wrongful act during prosecution, a single, clear, and objective (i.e., applicationally self-sufficient) standard should apply. The Federal Circuit's two February 2006 decisions in *Digital Control* and *Ferring* have contributed to an unsettled status quo, providing multiple, subjective, and varying standards for both the materiality and intent prongs of the inequitable conduct analysis. A single, clear standard is necessary for both.

For materiality, the 1992 version of Rule 56[296] would suffice. It is more objective than the subjective and malleable "reasonable examiner" standard. However, in deference to the PTO's presumptive expertise concerning the information that it needs to conduct effective and efficient examination, it would be appropriate for a patent reform statute to codify deference to the PTO's then-existing definition of materiality, consistent with former PTO Commissioner Gerald Mossinghoff's view.[297]

For intent, *Ferring* and its progeny should be overruled. A standard that both affirmatively states what evidence of intent is sufficient and prohibits certain methods of proving intent should replace it. Specifically:

---

intent, including abrogation of *Ferring* and adoption of *Star Scientific*, (3) abandoning of the balancing step, (4) judicial discretion to fashion remedies, and (5) awarding attorney fees to patentees who successfully defend against claims of inequitable conduct).

296. 37 C.F.R. § 1.56 (1992).

297. Gerald S. Mossinghoff, The Duty of Candor and Good Faith to the United States Patent and Trademark Office, Remarks to the American Bar Association, Intellectual Property Law Section at the 17th Annual Intellectual Property Law Conference (Apr. 12, 2002) (transcript on file with Oblon, Spivak, McClelland, Maier & Neustadt, LLP), *available at* http://www.oblon.com/media/index.php?id=44 ("My own view is that the courts should apply the version (or versions) of Rule 56 that was (were) in effect at the time the conduct objected to occurred.").

1. Defendants should not be permitted to prove intent from materiality;

2. The concepts of "gross negligence" and "should have known" should be abrogated from the intent determination; and

3. Circumstantial evidence should continue to be permissible in proving intent. But if circumstantial evidence is used to satisfy the clear-and-convincing burden of proof, an intent to deceive the PTO must be the *single most reasonable* inference,[298] not merely a permissible or plausible inference.

Even with improved and clarified standards of materiality and intent, it will be important to stay true to the doctrine's equitable roots and retain some degree of judicial discretion, so the balancing step—as articulated most recently in *Star Scientific*[299]—should be retained.

Finally, the remedy for inequitable conduct should be harmonized with the law of invalidity, such that unenforceability is determined on a claim-by-claim basis, and only as to patent claims for which there is a justiciable case or controversy. Under current rules, inequitable conduct relating to a single patent claim—even a claim that is not asserted in a patent infringement suit—can produce the draconian result that renders unenforceable the entire patent (and, via the doctrine of infectious unenforceability, downstream patents as well). This creates significant incentives to "roll the dice" by asserting an inequitable conduct claim, spawning satellite litigation on non-asserted claims. The proposed narrower remedy would strike a more appropriate balance. It would limit overbroad incentives for accused infringers to allege inequitable conduct while still providing consequences for inappropriate conduct during prosecution. Even this narrowed remedy would be broader than the remedy for invalidity, since material information—under either the "reasonable examiner" standard or the 1992 version of Rule 56—is defined more broadly than information that could ultimately result in a finding of invalidity of a claim. For example, the 1992 version of Rule 56 defines materiality to include information that would tend to show a prima facie case of unpatentability *without regard to rebuttal evidence*.[300] Empowering the courts with a discretionary range of remedies should not be embraced, because such discretion

298. *Supra* note 171 and accompanying text.

299. *Supra* note 168 and accompanying text.

300. Thus, for example, withheld material information may not result in invalidity, once appropriate rebuttal evidence is considered, but could result in unenforceability due to inequitable conduct if the withholding was done with the intent to deceive the PTO.

merely increase incentives for litigants to assert weak claims of inequitable conduct, hoping that the weak claim will produce at least a limited remedy.[301]

These specific reforms could be achieved with the same verbal economy as the inequitable conduct provisions in Senate Bill 1145, the Leahy Bill, as shown in the table below.

Leahy Bill	This Article's Proposal
a) IN GENERAL.—A party advancing the proposition that a patent should be cancelled or held unenforceable due to inequitable conduct in connection with a matter or proceeding before the United States Patent and Trademark Office shall prove independently by clear and convincing evidence that material information was misrepresented or omitted from the patent application of such patent with the intention of deceiving the Office.	(a) IN GENERAL.—A party advancing the proposition that one or more claims of a patent should be cancelled or held unenforceable due to inequitable conduct in connection with a matter or proceeding before the United States Patent and Trademark Office shall prove independently by clear and convincing evidence that a person subject to the duty of disclosure in connection with such matter or proceeding violated the duty of disclosure with the intention of deceiving the Office.
(b) MATERIALITY.—Information shall be considered material for purposes of subsection (a) if— (1) a reasonable patent examiner would consider such information important in deciding whether to allow the patent application; and (2) such information is not cumulative to information already of record in the application.	(b) DUTY OF DISCLOSURE.—For the purposes of this section, the Director shall be authorized to promulgate regulations defining the duty of disclosure. The duty of disclosure governing an allegation of inequitable conduct shall be that which was in effect at the time the alleged inequitable conduct occurred.
(c) INTENT.—Intent to deceive the Office may be inferred under subsection (a), but the inference may not be based solely on the gross negligence of the patent owner or its representative, or on the materiality of the information misrepresented or not disclosed.	(c) INTENT.—Intent to deceive the Office may be proven under subsection (a) by direct or circumstantial evidence. Circumstantial evidence may be found to constitute clear and convincing evidence of an intent to deceive the Office only if an intent to deceive is the single most reasonable inference from all the evidence, including any evidence of good faith. Any such inference may not be based solely on: (a) the gross negligence of the person alleged to have violated the duty of disclosure; (b) the materiality, or degree of materiality, of the information misrepresented or not disclosed; or

[301] *See supra* note 216–18 and accompanying text.

	(c) a finding that the person alleged to have violated the duty of disclosure "should have known" of the materiality of the information misrepresented or not disclosed.
(d) PLEADING.—In actions involving allegations of inequitable conduct before the Office, the party asserting the defense or claim shall comply with the pleading requirements set forth under Federal Rules of Civil Procedure 9(b).	
	(d) EQUITABLE BALANCING.—If both a violation of the duty of disclosure and intent to deceive the Office are proven by clear and convincing evidence, the court shall undertake an equitable determination whether, in view of all the facts, a finding of inequitable conduct is warranted.
(e) REMEDIES.—If the court finds both that material information was misrepresented to, or withheld from, the Office and an intent to deceive, after balancing the equities, the court, using its discretion, shall impose 1 or more of the following remedies as it deems appropriate: (1) Hold the patent unenforceable. (2) Hold 1 or more claims of the patent unenforceable. (3) Order that the patentee is not entitled to equitable relief and that the sole and exclusive remedy for infringement of the patent shall be a reasonable royalty.	(e) REMEDY.—If the court makes a finding of inequitable conduct, it shall hold unenforceable those claims of the patent to which the information misrepresented or not disclosed is material.
	(f) PLEADING.—In actions involving allegations of inequitable conduct before the Office, the party asserting the defense or claim shall comply with the pleading requirements set forth under Federal Rules of Civil Procedure 9(b).

VII. CONCLUSION

In the late 1980s, measured by the number of cases and the deleterious effects of improper use of the defense, there was a plague of inequitable conduct allegations in patent litigation. In response, the Federal Circuit implemented several reforms in the *Kingsdown* case, and the PTO issued an amended Rule 56. Over the following two decades, the prevalence of inequitable conduct rulings in the Federal Circuit ebbed and flowed, and a number of doctrinal ambiguities persisted. Then, in 2006, with its rulings in *Digital Control* and *Ferring*, the Federal Circuit rolled back the reforms of the late

1980s and early 1990s. Subsequent decisions have further muddied the doctrine, with inconsistent and expansive application of the doctrine's general principles. These doctrinal problems invite tactical use and abuse of the doctrine in future cases—which may be viewed as one form of expanding the "plague."

In fact, litigants at the district court level have asserted the defense of inequitable conduct in an ever-increasing proportion of patent cases. The wide and growing disparity between the frequency with which inequitable conduct is pled, on the one hand, and the percentage of all patent cases that ultimately result in a Federal Circuit ruling of inequitable conduct, on the other, is further indicative of a spreading "plague."

Reform is needed. Congress has proposed a number of different statutory revisions. Each of the patent reform bills introduced during the 110th Congress had several good ideas and several ideas that lacked historical context, were insufficient to implement true reform, or, worse, would likely be counterproductive against the stated purpose of reining in abuse of the inequitable conduct defense.

Patent reform legislation was again introduced early in the 111th Congress.[302] However, initial versions of the legislation did not include provisions for inequitable conduct reform.[303] Between April 2, 2009 and October 5, 2009, the Patent Reform Act of 2009 showed little progress in Congress. However, on October 5, 2009, Commerce Secretary Locke sent Senators Leahy and Sessions a letter providing the Obama Administration's views on patent reform,[304] which may signal a resumption in legislative activity.

In the absence of legislative reform, judicial reform of inequitable conduct is also possible. However, there is no prospect for Supreme Court action on the horizon. The Court denied certiorari in 2006 in *Ferring*,[305] and did so again in 2009 in *Aventis v. Amphastar*.[306] In January 2009, patentee Aventis petitioned the Supreme Court for a writ of certiorari[307] from the Federal Cir-

---

302. S. 515, 111th Cong. (as introduced in Senate, Mar. 3, 2009); H.R. 1260, 111th Cong. (as introduced in Congress, Mar. 3, 2009); S. 610, 111th Cong. (as introduced in Senate, Mar. 17, 2009).

303. *Id.* The sole exception is that the 2009 Kyl bill, S. 610, § 299, includes provisions for civil sanctions for misconduct before the PTO.

304. Letter from Gary Locke, Sec'y, Dep't of Commerce, to Patrick Leahy, Chairman, and Jeff Sessions, Ranking Member, Comm. on the Judiciary U.S. Senate (Oct. 5, 2009) (on file with the U.S. Dep't of Commerce), *available at* http://judiciary.senate.gov/resources/documents/111thCongress/upload/100509LockeToLeahySessions.pdf.

305. Ferring B.V. v. Barr Labs., Inc., 549 U.S. 1015 (2006).

306. Aventis Pharma S.A. v. Amphastar Pharm., Inc., 129 S. Ct. 2053 (2009).

307. Petition for Writ of Certiorari, *Aventis*, 129 S. Ct. 2053 (No. 08-937) (arguing for

cuit's decision in *Aventis v. Amphastar*, in which the patent on Aventis' $2 billion-per-year drug was held unenforceable due to an omission in a non-inventor expert's declaration. With echoes of *Ferring*, the expert was deemed to have made this omission intentionally because he should have known of the materiality of the omitted information.[308] Aventis concluded its petition for certiorari with a plea for reform:

> This issue will not benefit from further percolation in the circuits. The split in the lower courts and within the Federal Circuit itself is deep and mature, and the Federal Circuit has exhibited a steadfast unwillingness to revisit the issue en banc. Four decades of confusion are enough. The question presented is ripe—indeed overdue—for this Court's review.[309]

Nonetheless, the Supreme Court denied certiorari.

Absent action by Congress or the Supreme Court, it remains possible that the Federal Circuit will implement doctrinal reform. Given the proliferation of panel decisions from the Federal Circuit, it would take an en banc decision to reform the substantive doctrine. However, assuming the Federal Circuit is actively looking for a suitable inequitable conduct case to take en banc,[310] it is unclear when that might happen; in late 2008, for example, the Federal Circuit declined a rehearing en banc in *Star Scientific*.[311]

In this context, the Federal Circuit's August 2009 decision concerning the Rule 9(b) pleading standard in *Exergen v. Wal-Mart*[312] hints at an interesting possible avenue for reform. The decision goes significantly farther than prior cases in the degree of specificity required to plead inequitable conduct. For example, an inequitable conduct pleading must now name the specific individual(s) alleged to have committed inequitable conduct, identify the specific claim limitations affected, identify the specific passages in the withheld reference that are alleged to be material, establish that those passages are not cu-

---

reform of the intent element of inequitable conduct, abolition of the "sliding scale between intent and materiality," and for a broader discretionary range of remedies).

308. Aventis Pharma S.A. v. Amphastar Pharm., Inc., 525 F.3d 1334, 1344–49 (Fed. Cir. 2008); Petition for Writ of Certiorari, *Aventis*, 129 S. Ct. 2053 (No. 08-937).

309. *Id.* at *30.

310. Chief Judge Michel has called upon practitioners to be "more strategic and more imaginative" in submitting *en banc* petitions to the Federal Circuit on important issues. Chief Judge Paul R. Michel, Remarks at Harvard Law School Conference on Intellectual Property Law 7 (Sept. 9, 2008) (transcript on file with author), *available at* http://www.cafc.uscourts.gov/CJM_Speech_Harv_LS_Conf_9-08.pdf.

311. Star Scientific, Inc. v. R.J. Reynolds Tobacco Co., 537 F.3d 1357, 1357 (Fed. Cir. 2008), *reh'g en banc denied*, 2008 U.S. App. LEXIS 25385 (Fed. Cir., Oct. 22, 2008).

312. Exergen Corp. v. Wal-Mart Stores, Inc., 575 F.3d 1312 (Fed. Cir. 2009).

mulative of all the other references that were cited to the examiner, and plead facts from which a specific intent to deceive the examiner is reasonable (i.e., is plausible and flows logically from the facts alleged).[313]

This degree of required specificity will doubtless chill some proportion of strategically-pled inequitable conduct claims, and time will tell how effective a strengthened Rule 9(b) requirement is at controlling the "plague" of inequitable conduct.

Ultimately, however, the success of a heightened pleading standard in reining in allegations of inequitable conduct will depend on the contours of the underlying doctrine. Consider, for example, the lesser impact of a heightened pleading standard on the intent element if, instead of the *Star Scientific* requirement that intent be the single most reasonable inference, the prevailing standard was *Ferring*'s should-have-known test. Put another way, if the bar for proving inequitable conduct is low, requiring the elements to be pled with specificity does not raise the bar of proof. If, however, the substantive bar is higher (and clearly articulated), then a stringent Rule 9(b) standard will weed out the strategically-asserted claims of inequitable conduct that seek to capitalize on lax pleading standards and doctrinal confusion. Therefore, although *Exergen* can be expected to cause a reduction in the prevalence of inequitable conduct in patent litigation, substantive reform of the kind outlined in this Article will be required to truly control the "plague."

313. *Id.* at 1326–28.

THE EMPEROR HAS NO COPYRIGHT: REGISTRATION, CULTURAL HIERARCHY, AND THE MYTH OF AMERICAN COPYRIGHT MILITANCY

John Tehranian[†]

ABSTRACT

This Article subverts the myth of American copyright militancy by providing a more nuanced view of our enforcement regime and detailing how, in the age of mechanical (and digital) reproduction, procedural nicities establish cultural hierarchy through the selective restoration of Benjaminian 'aura' to creative works. As it turns out, the Emperor has been sold a suit of copyright that leaves a surprising number of authors naked—without sufficiently meaningful remedies for infringements of their creative output. Copyrighted works are effectively placed into a hierarchy of protection that, in many ways, safeguards creators less vigorously than regimes in other countries. Through the use of ostensibly neutral formalities, the current system privileges the interests of repeat, sophisticated rights holders, often at the expense of smaller, less sophisticated creators. Moreover, existing law practically encourages certain kinds of infringement. In the end, sophisticated players enjoy powerful remedies when enforcing their copyrights. They dangle the legal Sword of Damocles—draconian statutory damages—over the heads of accused infringers, threatening to hand defendants their heads on a platter with more fervor than Salomé's dance (to licensed music, of course). By sharp contrast, when they function as users of intellectual property (something all creators do), these same players often face only the most paltry of penalties for unauthorized exploitation—even when they infringe willfully. Our copyright regime therefore beatifies the works of elites—consecrating their cultural production as sacred texts and subjecting

---

† Professor of Law and Director, Entertainment Law Center, Chapman University School of Law. I would like to thank Mark Bartholomew, David Carson, Jonathan Handel, Eric Goldman, Ernesto Hernandez-Lopez, Chris Koras, Tyler Ochoa, Pamela Samuelson, Alfred Yen, and all of the participants of the Conference on the 100th Anniversary of the 1909 Copyright Act, co-sponsored by the Berkeley Center for Law & Technology, University of California, Berkeley School of Law, and the High Technology Law Institute, Santa Clara University School of Law, for their helpful comments and thoughts on this Article. Parts of this article come from my forthcoming book *Infringement Nation: Making the Case for Copyright 2.0* (Oxford University Press).

any use to permission and payment—while rendering the creative output of the rest of society into fodder for unauthorized manipulation and commercialization. The point of this analysis is not to call for even greater copyright protection for all creators. Rather, this Article deconstructs the beneficiaries of the existing regime and highlights the need for holistic reform that equalizes protection among different classes of authors and rights holders while also balancing the interests of copyright users.

TABLE OF CONTENTS

I. INTRODUCTION: ART, AURA, AND AUTHENTICITY

In his seminal meditation on art and technology, Walter Benjamin contemplated the transformative role of mechanical reproduction on society's relationship with creative works.[1] Specifically, he postulated that the increasing ease of replication would destroy art's *aura*—its perceived authenticity and ritualistic value. "For the first time in world history," argued Benjamin, "mechanical reproduction emancipates the work of art from its parasitical dependence on ritual."[2] To Benjamin, mass mechanical reproduction would result in the demystification of art by liberating it from its erstwhile settings. He concluded that "the technique of reproduction detaches the reproduced object from the domain of tradition. . . . [I]n permitting the reproduction to meet the beholder or listener in his own particular situation, it reactivates the object reproduced."[3]

Benjamin's prescient views on art in the postmodern and digital eras have been widely appreciated.[4] But just as nature cannot escape Newton's Third Law of Motion,[5] there have been forces pushing against the inexorable march of technology. Benjamin underestimated the way in which law could emerge as a powerful countervailing force against the demystification of art. Indeed, as mechanical reproduction has flourished—thereby subverting art's *aura*—copyright law has concomitantly grown more robust—thereby policing authenticity and reaffirming the *aura*. In short, copyright law has served as a powerful bulwark against the demystifying tide of mechanical reproduction.

Copyright law, where it is most strictly applied, prohibits any kind of reproduction, whether manual or mechanical.[6] It controls exhibition of

---

1. Walter Benjamin, *The Work of Art in the Age of Mechanical Reproductions*, *in* ILLUMINATIONS (Hannah Arendt ed., Harry Zohn trans. 1968).

2. *Id.* at 224.

3. *Id.* at 221.

4. *See, e.g.*, Robert W. Sweeny, *Three Funerals and a Wedding: Simulation, Art Education, and an Aesthetics of Cloning*, 31 VISUAL ARTS RES. 60 (2005) (discussing Benjamin's theory in the context of mediating the relationship between art and developing technologies); LIZ WELLS, PHOTOGRAPHY: A CRITICAL INTRODUCTION (2004) (discussing Benjamin's argument that changes brought about by mechanical reproduction precipitated a sea change in attitudes toward the arts, especially photography); Najmeh Khalili, *Walter Benjamin Revisited: The Work of Cinema in the Age of Digital (Re)production*, OFFSCREEN, Oct 31, 2003, *available at* http://www.horschamp.qc.ca/new_offscreen/new_media.html (discussing digital media theory in light of Benjamin's work).

5. For every action, there is an equal and opposite reaction. *See* Isaac Newton, *Mathematical Principles of Natural Philosophy*, *in* THE AGE OF REASON 108 (Louise L. Snyder ed., 1955).

6. 17 U.S.C. § 106(1) (2006) (granting copyright holders the exclusive right to make any type of reproduction of a protected work).

works through public performance and display rights.[7] And it carefully patrols a creative work's cultural value through the derivative-works doctrine.[8] Moreover, the decoupling of a copyright from ownership of a physical object enables the exertion of control over a creative work to be distant and omnipresent. As Christian Stallberg points out, "[i]f intellectual works can be used everywhere, then the exclusive protection of those works restricts people everywhere."[9] After all, in the age of mechanical (and digital) reproduction, creative works can be disseminated universally. Yet copyright law imposes an artificial scarcity. It may do so with very good reason, but the consequences of this regulatory power bear careful scrutiny.

To understand the role of copyright law in enforcing artificial scarcity, this Article closely examines the practicalities of infringement litigation, specifically the issue of copyright registration and damages. Perhaps due to its banal technicalities, our registration regime has received little attention from academics, who have eschewed analysis of its various niceties in favor of more substantive aspects of copyright law. However, the registration system is deeply relevant to anyone attempting to enforce his or her copyright and, as such, plays a key role in understanding how copyright law functions in practice. As we shall see, the remedies afforded under the Copyright Act, and the prerequisites for their availability, inextricably affect infringement behavior. Moreover, they determine the types of works that are entitled, in the age of mechanical reproduction, to resurrect the Benjaminian *aura*, and the types of works that are not. As theory begets praxis, these seemingly procedural rules have a profound and substantive impact on the fundamental nature of our copyright regime. Specifically, the registration system reifies the divide between high and low-brow works, sustaining the *aura* of art according to a cultural hierarchy policed by legal formalities.

The implications of this gestalt are significant, contradicting one of the most oft-repeated axioms about our intellectual property laws: that we take copyright seriously and enforce it vigorously with one of the most protective regimes in the world. Academics, politicians, trade representatives, and the content-creation industries alike have reiterated this apparent truism time and time again.[10] But it is not entirely accurate. In a sense, the Emperor has been

7. 17 U.S.C. §§ 106(4)–(5) (2006).

8. *See* 17 U.S.C. §§ 103, 106(2) (2006).

9. Christian G. Stallberg, *Towards a New Paradigm in Justifying Copyright: An Universalistic-Transcendental Approach*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 333, 337 (2008).

10. *See, e.g.*, BÉNÉDICTE CALLAN, PIRATES ON THE HIGH SEAS: THE UNITED STATES AND GLOBAL INTELLECTUAL PROPERTY RIGHTS 1 (1998) (noting that, in recent years, the United States has "cast itself as the great proponent of intellectual property rights worldwide").

sold a suit of copyright that leaves unsophisticated creators naked—that is, without sufficiently meaningful remedies for infringements of their creative works. Copyrighted works are effectively placed into a hierarchy of care that in many ways safeguards creators less vigorously than regimes in other countries. At its core, the current system privileges the interests of repeat, sophisticated, and monied rights holders—rights holders who are invariably also users of content. And it does so at the expense of smaller, less sophisticated creators and authors.[11] Moreover, existing law practically encourages certain kinds of infringement. Specifically, a vast disparity has emerged between sophisticated and unsophisticated creators of copyrightable content—a divide enforced through a single technical feature of our copyright regime: the registration requirement. For the sophisticated creators who timely register their copyrighted content, the inviolable aura of their works is virtually assured. For unsophisticated creators who fail to timely register their copyrighted content, their works enjoy only low-tier protection and remain vulnerable to unauthorized manipulation and appropriation.

By unfurling the unique importance of timely registration in shaping remedies, this Article punctures the myth of American copyright militancy. Part II examines how judicial interpretations of the Copyright Act have narrowed the availability of enhanced damages for continuing infringements, created a one-way risk of attorneys' fees assessments for unsophisticated plaintiffs, foreclosed availability of punitive and reputational damages, and, in short, left most authors at a comparative disadvantage in protecting their intellectual property rights in the United States vis-à-vis the rest of the world. As a result, the registration system has failed to achieve its basic notice function and has potentially shirked our international treaty obligations—a particularly salient problem in light of our efforts to combat lax copyright enforcement in many developing countries.

Part III deconstructs the failure of prior efforts to amend the harsh results of the timely registration requirement. This analysis suggests that certain sophisticated, repeat players in the content industries derive the best of both worlds from the timely registration requirement. On one hand, they enjoy strong rights when seeking to enforce their copyrights, often wielding the threat of disproportional penalties against accused infringers. On the other hand, when they function as users of intellectual property (something all creators do), these same players often face only the most paltry of

11. One might argue that it is unsurprising that any aspect of our legal regime would privilege repeat players, but copyright law does so with extreme vigor. And the particular privileges that copyright law grants go against conventional wisdom on the subject.

penalties, even when they infringe willfully. Finally, Part IV assesses possible avenues for change and offers some caveats regarding outright repeal of the timely registration requirement. Specifically, upon consideration of the consequences of copyright law's remaining technicalities, this Article proposes holistic reform measures that place creators—both sophisticated and unsophisticated alike—on a relatively equal footing while balancing the rights of copyright holders with copyright users.

II. FORM FRUSTRATES FUNCTION? RE-EXAMINING COPYRIGHT'S REGISTRATION REQUIREMENT

A. CHALLENGING THE CONVENTIONAL WISDOM ON AMERICAN COPYRIGHT PROTECTION: CONTENT HIERARCHY AND THE REGISTRATION REQUIREMENT

Conventional wisdom maintains that we enjoy one of the world's most robust intellectual property regimes through our arduous protection of the exclusive right of creators to control the reproduction, distribution, and exploitation of their works. Indeed, we pride ourselves on our respect for creations of the mind, often analogizing the piracy of copyrighted works to the outright theft of tangible property.[12] This view is further buoyed by our reputation on the international scene and the heated rhetoric of federal officials and entertainment industry players in chastising some countries for their more lax intellectual property regimes. Our demands for stronger copyright enforcement abroad have led to high profile clashes with officials in such countries as China[13] and Russia,[14] where loose enforcement and rampant piracy have drawn our ire.

There are two limited exceptions, however, to this general proposition. First, as many observers have noted, our legal regime is less protective of the moral rights of creators than regimes in some other countries, especially

12. *See, e.g.*, Grand Upright Music, Ltd. v. Warner Bros. Records, Inc., 780 F. Supp. 182, 183 (S.D.N.Y. 1991) (concluding, upon considering the propriety of digital sampling without authorization, that "[t]hou shalt not steal").

13. *See, e.g.*, WTO, China - Measures Affecting the Protection and Enforcement of Intellectual Property Rights, http://www.wto.org/english/tratop_e/dispu_e/cases_e/ds362_e.htm (last visited Oct. 25, 2009) (summarizing the conflict between the United States and China over the adequacy of latter country's intellectual property protection and enforcement mechanisms).

14. *See* David E. Miller, *Combating Copyright Infringement in Russia: A Comprehensive Approach for Western Plaintiffs*, 33 VAND. J. TRANSNAT'L L. 1203 (2000) (noting that American corporations lose millions of dollars each year as a result of the illegal reproduction and sale of copyrighted goods in Russia).

those in Western Europe.[15] But, our resistance to moral rights has not been explained as a product of hostility towards copyright protection or even the rights of authors. Instead, it has been rationalized as a product of American capitalism and its desire to maximize the alienability of property rights and to preserve a marketplace for copyrighted works.[16] Moral rights, we are told, may unduly interfere with the disposition of tangible property that incorporates copyrighted content.

Second, observers have pointed out that the hard line that the United States has taken on copyright issues is of relatively recent vintage. Specifically, these critics have questioned the moral undertones of the international North-South discourse on copyright protection by calling attention to the selective historical consciousness at play.[17] Despite efforts by the United States and its copyright allies to pressure some developing countries to increase their enforcement efforts in combating piracy and protecting copyright, it is important to recognize that, at a similar point in our nation's development, we adamantly refused to recognize the copyrights of foreign authors. In short, until the early twentieth century, the United States was the most prominent rogue nation on the international copyright scene.

As law professor Harry G. Henn wrote in 1953, "[t]he United States has been among the most parochial of nations so far as copyright protection for published works is concerned. For over a hundred years, this nation not only denied copyright protection to published works by foreigners . . . but appeared to encourage the piracy of such works."[18] Our nation's first Copyright Act, passed in 1790, explicitly denied protection to any creative work "written, printed or published by any person not a citizen of the United States, in foreign parts or places without the jurisdiction of the United States."[19] Indeed, between 1800 and 1860, nearly fifty percent of the

---

15. Henry Hansmann & Marina Santilli, *Authors' and Artists' Moral Rights: A Comparative Legal and Economic Analysis*, 26 J. LEGAL STUD. 95, 95–97 (1997).

16. *See, e.g.*, Lawrence A. Beyer, *Intentionalism, Art and the Suppression of Innovation: Film Colorization and the Philosophy of Moral Rights*, 82 NW. U. L. REV. 1011, 1047, 1052–54 (1988) (arguing that the moral right of integrity subverts both buyer and seller freedom in market transactions involving copyrighted works); Stephen L. Carter, *Owning What Doesn't Exist*, 13 HARV. J.L. & PUB. POL'Y 99, 101 (1990) (noting that the moral rights doctrine runs counter to traditional property rights notions by telling owners of paintings, films and other works that they "should not have the right to do with their possessions as they wish").

17. Peter K. Yu, *The Copyright Divide*, 25 CARDOZO L. REV. 331 (2003).

18. Harry G. Henn, *The Quest for International Copyright Protection*, 39 CORNELL L. Q. 43, 52 (1953).

19. Act of May 31, 1790, ch. 15, 1 Stat. 124, § 5.

bestsellers in the United States were pirated English novels.[20] It was not until the end of the nineteenth century that things changed, after heavy lobbying by prominent British and American writers. Authors across the pond, such as Charles Dickens, were deprived of royalties for sales in the States. Domestic authors, such as Mark Twain,[21] were being denied foreign royalties since other countries were reciprocally declining to grant copyright to American authors.[22] Not until the passage of the International Copyright Act of 1891, also known as the Chase Act, would foreign authors finally enjoy copyright protection in the United States.[23]

According to the popular narrative, such unabashedly piratical lapses are merely vestiges of a bygone era.[24] As a result, we continue to view our modern copyright regime as muscular and highly protective of creators. Indeed, many observers—myself included—have critiqued the growing magnitude of our copyright monopoly and how it has often come at the

---

20. Yu, *supra* note 17, at 341.

21. *See, e.g.*, R. KENT RASMUSSEN, MARK TWAIN A TO Z: THE ESSENTIAL REFERENCE TO HIS LIFE AND WRITINGS 54 (1995). Rasmussen noted that

> [t]he absence of international copyright laws allowed Canadian publishers to prey on Mark Twain's early books. He was hurt badly in 1876, when the Toronto publisher Charles Belford issued *Tom Sawyer* before the American edition even appeared. To combat this problem, Mark Twain spent several weeks in Montreal in November–December 1881 with James R. Osgood to meet a residency requirement to protect his *The Prince and the Pauper* copyright.

Id.

22. Prior to 1891, some foreign authors circumvented America's refusal to honor copyrights of foreign authors by having an American citizen collaborate in the publishing process. Usually, this would take the form of the American writing a short preface to the book and then registering the work with the U.S. Copyright Office under the collaborator's name. For example, Thomas Henry Huxley took this route to gain protection. *See, e.g.*, Philip V. Allingham, *Nineteenth-Century British and American Copyright Law*, VICTORIANWEB, Jan. 5, 2001, http://www.victorianweb.org/authors/dickens/pva/pva74.html (detailing the technical subterfuge and arduous machinations required of British authors to obtain American copyright protection both before and after 1891, respectively).

23. Chace Act, ch. 565, 26 Stat. 1106, 1110 (1891).

24. According to the Council on Foreign Relations' American Intellectual Property Rights Policy Study Group, the merely "nominal protection" of intellectual property rights and "indifference and resistance from American officials . . . to enforce copyrights for literary works" in the nineteenth century has given way in recent years to a regime of strong enforcement, with the United States "cast[ing] itself as the great proponent of intellectual rights worldwide . . . [by] tak[ing] the moral high ground in the battle against international piracy and counterfeiting, denouncing unfair practices abroad and claiming that strong rights can only help the economy in developing countries." CALLAN, *supra* note 10, at 1.

expense of the public interest and the rights of users of expressive materials.[25]

However, broadly speaking, the copyright regime is not nearly as uniformly protective of copyright holders and creators as we often think. On the surface, we appear to advance the interests of copyright holders with vigor—perhaps too much so. But, the formalities of copyright protection and enforcement reveal a more complex system at operation.

Through formalities, the 1976 Copyright Act actually created two distinct tiers of effective protection for copyrighted works. Sophisticated, routine creators—generally corporations in content-creation industries—timely register their works and therefore enjoy generous remedies against infringers. These remedies include the recovery of reasonable attorneys' fees and the assessment of statutory damages—which can rise to the draconian level of up to $150,000 per willful act of infringement. Absent any proof of actual damages, such plaintiffs can elect statutory damages that quickly create the possibility of a multi-million dollar judgment in their favor. By sharp contrast, unsophisticated creators, like individual artists, typically do not timely register their works and are often left with little except moral force and the uncertain threat of injunctive relief to enforce their intellectual property rights. The dichotomy between sophisticated and unsophisticated creators thereby determines the relative sanctity of copyrighted works.

---

25. *See, e.g.*, James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 34–37, 40–41 (2003) (questioning the recent expansion of intellectual property monopolies by comparing it to the enclosure movement of the eighteenth century); Jessica Littman, *Creative Reading*, 70 LAW & CONTEMP. PROBS. 175, 180 (2007) (criticizing the prevailing position that any use of an existing work constitutes an infringement unless specifically exempted from liability by law); Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law, 1900-2000*, 88 CALIF. L. REV. 2187, 2191 (2000) (characterizing the history of intellectual property rights over the past one-hundred years as a century of "solicitude" by corporate interests bent on maximizing monopoly-like protections for their intellectual properties); John Tehranian, *Et Tu, Fair Use? The Triumph of Natural-Law Copyright*, 38 U.C. DAVIS L. REV. 465, 466 (2005) (arguing that "the fair use doctrine has actually enabled the expansion of the copyright monopoly well beyond its original bounds and has undermined the goals of the copyright system as envisioned by the Framers of the Constitution."); Rebecca Tushnet, *Copy This Essay: How the Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535 (2004) (criticizing existing copyright doctrine for failing to recognize adequately the public interest served through unauthorized non-transformative reproduction of copyrighted works); Pamela Samuelson, *The Copyright Grab*, WIRED.COM, Jan. 1996, http://www.wired.com/wired/archive/4.01/white.paper_pr.html (critiquing proposed expansions in copyright protection by the Clinton administration for their harm to the freedom and privacy of the general public).

B. REGISTRATION AND THE PIVOTAL ROLE OF STATUTORY DAMAGES AND ATTORNEYS' FEES

The registration requirement is a critical aspect of the governing 1976 Copyright Act. Oddly enough, however, the abandonment of key formalities required for copyright protection was a purported hallmark of the Act.[26] Most notably, the Act vested an automatic federal copyright with authors from the moment that they fix an original work in a tangible medium, without the need for registration or any other procedural step.[27] Nevertheless, upon closer examination, the 1976 Act's general reputation for eschewing formalism appears vastly exaggerated.

While formalities for subsistence may have been eliminated under the 1976 Act, formalities for effective enforcement of a copyright actually *increased*. First, the 1976 Act retained its predecessor's requirement of registration prior to the filing of an infringement action.[28] Even more significantly, the 1976 Act dramatically expanded formalities in a key regard.

---

26. *See, e.g.*, Shira Perlmutter, *Freeing Copyright from Formalities*, 13 CARDOZO ARTS & ENT. L.J. 565, 566, 568, 581(2006). Perlmutter notes that

> [f]ormalities have long been a hallmark of the American copyright system. . . . In the 1976 Act, Congress began the journey toward eliminating formalities from our copyright law . . . . The 1976 Copyright Act and adherence to the Berne Convention marked a sea change in U.S. copyright law—a profound shift in philosophy.

Id.; Malla Pollack, *Towards a Feminist Theory of the Public Domain, or Rejecting the Gendered Scope of United States Copyrightable and Patentable Subject Matter*, 12 WM. & MARY J. WOMEN & L. 603, 603–04 (2006) ("The Copyright Revision Act of 1976 moved the general line of protection from the point of publication to the point of fixation. In combination with the Berne Implementation Act, it eliminated most of the prior need for copyright formalities."); *see also* Pamela Brannon, *Reforming Copyright To Foster Innovation: Providing Access To Orphaned Works*, 14 J. INTELL. PROP. L. 145, 145, 158 (2006) ("Copyright protection prior to the 1976 Act was attended by a bevy of formalities. . . . The 1976 Copyright Act discarded most of these formalities, shifting to an 'opt-out' system that granted copyright protection upon the initial creation and fixation of a work."); Wendy J. Gordon, *Toward a Jurisprudence of Benefits: The Norms of Copyright and the Problem of Private Censorship*, 57 U. CHI. L. REV. 1009, 1010 (1990) ("One major reason for the increasing breadth of copyright scholarship is the 1976 Copyright Act, which simplified and rationalized the complexities and formalisms of prior law . . . ."); Matt Jackson, *The Digital Millennium Copyright Act of 1998: A Proposed Amendment to Accommodate Free Speech*, 5 COMM. L. & POL'Y 61, 71 (2000) ("Prior to the 1976 Copyright Act, authors had to comply with a laundry list of formalities in order to enjoy federal copyright protection.").

27. 17 U.S.C. § 102(a) (2006) ("Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device.").

28. 17 U.S.C. § 411(a) (2006) (requiring registration of a copyrighted work prior to the initiation of an infringement suit based thereon).

Under 1909 Act, a prevailing plaintiff could recover statutory damages and attorneys' fees without timely registration.[29] All of that changed with the 1976 Act. What the 1976 Act gave to creators of copyrighted works through its purported reduction of vesting formalities, it more than took away through the imposition of timely registration as a precondition for recovery of statutory damages and attorneys' fees—two of the most powerful weapons in a copyright holder's arsenal.[30] Thus, although creative works are now "protected" under federal law from the moment of creation, litigation rarely makes sense without proper and timely registration. In an ordinary case of copyright infringement—where an infringed work is not registered before the infringement begins—a plaintiff can only recover actual damages that directly result from the defendant's action. As we shall see, such a remedy is rarely adequate to enable a copyright holder to vindicate his or her interests in even the most clear-cut and brazen case of infringement.

Moreover, rather than further harmonizing our copyright regime with those of other countries, the 1976 Act, through its timely registration requirement, has ironically enhanced the exceptionalism of the American copyright system. Specifically, the United States is the only major country in the world with a timely registration prerequisite for the recovery of certain forms of damages and attorneys' fees.[31] In other countries, full legal vindication of one's exclusive rights does not require the added procedure of registration, let alone timely registration.[32]

The registration requirement plays an instrumental role in the enforcement of copyright in the United States. As Nimmer reminds us,

---

29. 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 7.16[C][2] (2008). Nimmer states,

> Under the 1909 Act . . . registration was only required . . . prior to the filing of an infringement action, and, in such an action, there might be a recovery (including, under the 1909 Act, statutory damages and attorney's fees) with respect to infringing acts that occurred prior to, as well as after, registration.

Id. (internal citations omitted).

30. Under the traditional American rule, parties bear the costs of their own representation, regardless of the outcome. By statute, however, prevailing plaintiffs in copyright infringement suits are eligible to receive their attorneys' fees—but only if the work was timely registered. Although the grant of fees lies squarely within the discretion of courts, prevailing plaintiffs often recover their fees. *See, e.g.*, Columbia Pictures TV v. Krypton Broad. of Birmingham, Inc., 106 F.3d 284, 296 (9th Cir. 1997) ("[A] plaintiff in a copyright action is generally awarded fees by virtue of prevailing in the action . . . .").

31. 4 NIMMER & NIMMER, *supra* note 29, § 17.01.

32. *See, e.g.*, *id.* § 17.03 ("[U]nlike the United States copyright law, under virtually all foreign copyright laws . . . there are no administrative formalities that must be satisfied in order to create or to perfect a copyright.").

"statutory damages may often constitute the only meaningful remedy available to a copyright owner for infringement of his work."[33] Yet not all copyright holders can qualify for statutory damages in an infringement suit—far from it, in fact. Under the reigning 1976 Copyright Act, statutory damages are only available to a certain class of copyright holders: those who register their works with the United States Copyright Office in a timely manner in relation to the infringement. The absence of timely registration also precludes an award of attorneys' fees to a prevailing plaintiff. As 17 U.S.C. § 412 provides,

> [N]o award of statutory damages or of attorney's fees . . . shall be made for . . . any infringement of copyright commenced after first publication of the work and before the effective date of its registration, unless such registration is made within three months after the first publication of the work.[34]

Thus, to qualify for statutory damages and the potential recovery of attorneys' fees, a copyright plaintiff must register before a defendant's act of infringement or within ninety days of publication. Registration, especially timely registration, therefore represents a pivotal feature on the copyright landscape. Without it, a plaintiff's remedies are dramatically constrained.

Assume, for example, that a pharmaceutical company usurps five of an artist's illustrations, without permission, for use on the packaging of their new male enhancement drug. With proper and timely registration, the artist can immediately force the pharmaceutical company to pay attention to her infringement claims and to cease the infringing conduct. Even without a demonstration of actual damages or profit from the infringement, a suit would expose the defendant to potential liability for statutory damages in the amount of $750,000 ($150,000 for each of five acts of infringement), reimbursement of the artist's reasonable attorneys' fees, and expenditure of its own attorney's fees.

Without proper and timely registration, however, the situation is radically different. The artist can generally recover only actual economic damages from the company—lost sales, or disgorgement of profits. Not surprisingly, the amount of these damages is often riddled with ambiguity. Moreover, unless the artist is world-renowned, her damages claim will rarely amount to more than a few thousand dollars. But pursuing an infringement suit will cost her several hundred thousand dollars in attorneys' fees. And although the artist might receive an injunction to prevent further infringement, it will be

33. 2 *id.* § 7.16[C][1].
34. 17 U.S.C. § 412(2) (2006).

costly to obtain since the significant fees she would have to incur are not recoupable. Thus, even under the most optimistic scenario, legal action will not be worth pursuing, unless she has a desire to fight for principal and end up bankrupt. In short, the artist may ultimately recover $5000 from the defendant, but such a victory would be pyrrhic at best, especially after accounting for the $250,000 invoice from her attorney.[35]

To make matters worse, a quarter of a million dollars is a conservative estimate for the cost of copyright litigation. According to the 2007 Report of the Economic Survey conducted by the American Intellectual Property Law Association, the mean cost of taking a relatively small instance of copyright infringement (one involving less than $1 million in potential liability) to trial in United States is $310,000.[36] For a middle-of-the-road infringement case (one involving $1-$25 million in potential liability), that figure rises to a mean of $749,000.[37]

As such, it frequently makes no economic sense to pursue litigation. The cost of filing a complaint in federal court will likely exceed the total amount of damages recoverable under even the most sanguine scenario. And this is true even though a defendant has undoubtedly infringed the work and done so with gusto. In short, the law fails to provide an effective remedy from the wrongdoing the artist has suffered at the hands of the pharmaceutical company, even though we think of our laws as protecting the sanctity and inviolability of intellectual property rights, especially when they are indisputably infringed.

This all-too-common situation allows large, sophisticated corporations to enforce their copyright with a vast array of tools, including statutory damages and attorneys' fees, while simultaneously enabling them to laugh in the face of less sophisticated players who lodge infringement claims against them. If you infringe the copyrights of the major motion picture studios or the major record labels, the specter of statutory damages and fees will squarely put you on the defensive. Witness the onslaught of suits against ordinary Americans filed by the Recording Industry Association of American (RIAA) for unauthorized peer-to-peer file sharing.[38] By contrast, if a large corporation

35. *See infra* text accompanying notes 150–55155.

36. AM. INTELL. PROP. LAW ASS'N, 2007 REPORT OF THE ECONOMIC SURVEY I-100 (2007).

37. *Id.* at I-101. When the amount at controversy exceeds $25 million, the mean cost of taking a case to trial is $1.292 million. *Id.*

38. Leslie Walker, *New Movement Hits Universities: Get Legal Music*, WASH. POST, Mar. 17, 2005, at E1 (noting that the RIAA has filed "thousands of suits against people for sharing copyrighted material").

violates your copyright, it can often thumb its nose at claims of infringement. In many cases, there is little you can do since most copyright holders are unlikely to register their copyright on a timely basis. As such, you are left with an appeal not to law but to morality. Thus, the dynamics of the existing registration regime—put into place on January 1, 1978 and largely unchanged by the implementation of the Berne Convention and other subsequent amendments to the Copyright Act—elevate procedural steps into outcome-determinative hurdles. The impact is both dramatic and underappreciated.

C. HIERARCHY AND THE UNSOPHISTICATED CONTENT CREATOR

A closer examination of the language and extant interpretation of the Copyright Act reveals the particular difficulties facing unsophisticated creators in seeking to vindicate their rights in the United States, especially in comparison to their peers in foreign countries. First, courts have found that § 412 precludes recovery of statutory damages and attorneys' fees when an infringement continues after registration. As a result, an infringer has veritable carte blanche to continue its wrongful activity with impunity if a work is not timely registered at the moment of first infringement. Second, unsophisticated creators face a dangerous one-way risk of attorneys' fees. Plaintiffs who fail to register their copyright on a timely basis are *never* eligible to recover their fees if they prevail in a suit, no matter how wanton the infringement at issue. By contrast, defendants are *always* eligible to recover their fees if they prevail. This unbalanced fees matrix dampens any enthusiasm that unsophisticated creators might have about seeking redress through the judicial system. Third, unsophisticated creators have no ability to seek punitive or reputational damages; the actual damages which they can receive are insufficient to make them whole and provide no deterrent effect against infringers. All told, these factors combine to create a rather bleak enforcement regime for the rights of creators who do not register in a timely manner.

1. Interpreting § 412's Timeliness Requirement: The Unavailability of Partial Eligibility for Statutory Damages or Attorneys' Fees

First, the challenges facing unsophisticated creators seeking to be made whole for unauthorized exploitation of their copyrighted works have grown more pronounced with the courts' reading of § 412's timely registration requirement. Specifically, even in cases of egregious and continuing infringement after registration, courts have denied plaintiffs access to statutory damages and attorneys' fees. On one hand, courts may have had no other alternative but to do so: there is little doubt that § 412 is perfectly clear in proscribing the imposition of statutory damages and attorneys' fees when

all acts of infringement by a defendant occur before registration. On the other hand, the issue is more ambiguous when an infringement occurs, the copyright holder then registers the work, and the infringement continues. Nevertheless, courts have almost uniformly resolved this issue in favor of defendants, holding that in such an instance statutory damages and attorneys' fees may not be awarded for any act of infringement—either before or after registration.

Courts have adopted this narrow view of the scope of plaintiff eligibility for statutory damages and attorneys' fees through their interpretation of the term "any infringement." Indeed, the reported cases considering the issue have determined that "infringement 'commences' for the purposes of § 412 when the first act in a series of acts constituting continuing infringement occurs."[39] Thus, "*the first act of infringement* in a series of ongoing infringements of the same kind marks the commencement of one continuing infringement under [§] 412."[40] As a result, if an artist discovers a company is infringing her work and then registers that work and sues for infringement, the artist cannot seek any recompense in the way of statutory damages or attorneys' fees, even for those acts of infringement that occur *after* the registration. Moreover, the narrow construction of "any infringement" under § 412 has led courts to insulate from statutory damages and attorneys' fees defendants who infringed prior to registration of the operative work, even if they conduct new acts of infringement after registration occurs.[41] The results of this rule are dramatic and perverse. As one disgruntled copyright claimant put it, this judicial interpretation grants infringers a "license to steal."[42]

---

39. Johnson v. Jones, 149 F.3d 494, 506 (6th Cir. 1998); *accord* Derek Andrew, Inc. v. Poof Apparel Corp., 528 F.3d 696, 700–01 (9th Cir. 2008); Bouchat v. Bon-Ton Dep't Stores, Inc., 506 F.3d 315, 330 (4th Cir. 2007), *cert. denied*, 128 S.Ct. 2054 (2008); Troll Co. v. Uneeda Doll Co., 483 F.3d 150, 158 (2d Cir. 2007); Mason v. Montgomery Data, Inc., 967 F.2d 135, 142–44 (5th Cir. 1992).

40. *See* Derek Andrew, Inc. v. Poof Apparel Corp., 528 F.3d 696, 701 (9th Cir. 2008).

41. *See, e.g.*, Qualey v. Caring Ctr. of Slidell, 942 F. Supp. 1074, 1076 (E.D. La. 1996). The court noted,

> [B]ecause the defendants commenced the first alleged infringement (preparing derivative works) prior to registration and publication, plaintiff is barred from recovering statutory damages or attorneys fees not only for that specific act of infringement, but also for any subsequent infringements of the drawings commenced after registration (or within the three-month period between first publication and registration).

Id.; *see also* Mason v. Montgomery Data, Inc., 967 F.2d 135, 144 (5th Cir. 1992) (holding that "a plaintiff may not recover an award of statutory damages and attorney's fees for infringements that commenced after registration if the same defendant commenced an infringement of the same work prior to registration.").

42. Teevee Toons, Inc. v. Overture Records, 501 F. Supp. 2d 964, 966 (E.D. Mich.

Consider a scenario where officials at a major clothing manufacturer decide to use the work of an artist for their new autumn line. Assume that they contemplate approaching the artist for a license but ultimately decide to use the work without authorization. Maybe they cannot be bothered to track down the author, perhaps they are concerned about the extra costs that a license would add to product development, maybe they think the chance of getting caught is remote, or perhaps they attempt to obtain a license but talks with the artist break down when the parties cannot agree on a rate of compensation. Regardless of the particular context and motivations at play, the officials are aware of the need to license under federal law but they decide to bring the product to market without a license. Several months later, the artist discovers the wholesale infringement. He immediately registers his work with the Copyright Office and then files suit against the company.

Because of the courts' narrow interpretation of § 412, the best the artist can hope for is to receive actual damages for the company's willful infringement. As a result, there is little incentive for the company to stop infringing. After all, whether the company ceases and desists now that it has been caught red-handed will play little role in any damages that the artist can receive. There are no punitive damages available in copyright actions. And statutory damages, which alone provide discretionary enhancements for willful infringement, are not available. The company can continue to infringe with impunity and, at the most, pay only actual damages that leave the unrepentant infringer without an incentive to respect copyright ex ante and effectively granting the infringer a compulsory license. Even if the artist attempts to obtain preliminary injunctive relief, the prospects for such relief are dubious. By the time he obtains an injunction, its value has diminished significantly: the clothing manufacturer may have moved on to its next line of seasonal clothing. Moreover, obtaining injunctive relief is expensive and the fees for doing so are non-recoupable.

The underlying rationale for this result is particularly problematic and rests on several unfounded assumptions about our registration system. In 2008, the Ninth Circuit addressed the registration timing issue on damages for the first time with its ruling in *Derek Andrew, Inc. v. Poof Apparel Corp*.[43] In its opinion, the court adopted the reasoning and conclusion of the other circuits that have considered the issue by finding that such an interpretation

---

2007).

43. Derek Andrew, Inc. v. Poof Apparel Corp., 528 F.3d 696 (9th Cir. 2008). The Ninth Circuit joined with the Second, Fourth, Fifth, and Sixth Circuits in barring recovery of statutory damages and fees when an infringement begins pre-registration and continues post-registration. *Id.* at 701.

of § 412 advanced Congress's intent to "promote early registration of copyright" by (1) "provid[ing] copyright owners with an incentive to register their copyright promptly" and (2) "encourag[ing] potential infringers to check the Copyright Officer's database." [44]

While the statute may not leave sufficient room for alternate interpretations, these rationalizations themselves do not hold up under scrutiny. First, if a court held that each of a series of on-going infringements constituted a new act of infringement for the purposes of § 412, copyright holders would still have a strong incentive to register. After all, there are numerous other advantages to registration besides qualification for statutory damages and attorneys' fees. Specifically, timely registration serves an important evidentiary function by enabling rights holders to make out infringement cases more easily. For example, registration provides a plaintiff with a *prima facie* presumption of copyright validity—but only if a work is registered within five years of publication.[45] Thus, plaintiffs have a strong incentive to register on relatively timely basis in order to enjoy the presumption of validity. Registration also provides proof of the date of creation. This benefit encourages timely registration because, *ipso facto*, a work's date of registration sets its latest possible date of creation. Thus, a registration certificate dated March 6, 2003 irrefutably establishes that the registered work must have been created before March 6, 2003. Such proof can be instrumental in many cases, especially where a defense of independent, or earlier, creation is proffered.

Second, as we explore later, the idea that potential infringers can confidently check the Copyright Office's database for complete registration information is deeply flawed. In short, the entire notion that registration will serve a notice function to potential infringers is vastly exaggerated. Moreover, the consequences of the current reading of § 412 are perverse. In many instances, it immunizes defendants to infringe with impunity.

Of course, it is fair to ask whether a different reading of § 412 by the courts, in which each violation of an exclusive right secured under § 106 constitutes a new infringement for the purposes of § 412, would fare any better. As it turns out, such an alternate reading would lead to a whole new set of policy problems. First, courts would have to address the thorny issue of when one violation ends and the next one begins. In the pre-Internet age,

---

44. *Id.* at 700–01.

45. 17 U.S.C. § 410(c) (2006) ("In any judicial proceedings the certificate of a registration made before or within five years after first publication of the work shall constitute prima facie evidence of the validity of the copyright and of the facts stated in the certificate.").

it was easier to make such a determination. After all, infringements came in more discrete bits—each day's edition of the newspaper featuring the infringement or each broadcast of an infringement could constitute a new act. But in the non-discrete world of the Internet, online newspapers and blogs are updated continuously and websites stream infringing content at all hours of the day. Second, a different reading of § 412 might nullify the entire purpose of the timely-registration incentive for statutory damages and attorneys' fees. After all, a plaintiff could wait for an infringement to occur, register and then sue with a colorable demand for statutory damages and attorneys' fees so long as defendant violates just a single exclusive right of the plaintiff after the effective date of registration. In the end, however, the shortcomings of an alternative interpretation of § 412 provides further reason to rethink the registration requirement as it pertains to statutory damages and attorneys' fees.

2. The One-Way Risk of a Fees Award

The 1976 Copyright Act disincentivizes unsophisticated artists from vindicating their legal rights through litigation in another way: the one-way risk of an attorneys' fee award. Section 505 gives courts the general discretion to grant fees to a prevailing party in an infringement suit, and frequently courts do so. For example, the Ninth Circuit, billed by Judge Alex Kozinski as "the Court of Appeals for the Hollywood Circuit,"[46] has, in the past, generally awarded fees to prevailing plaintiffs in copyright suits.[47] However, § 412 of the Act prevents courts from awarding fees to a plaintiff who has not timely registered.[48] This creates a one-way risk for any individual artist attempting to be made whole. Most individual artists do not timely register and are therefore ineligible for attorneys' fees. Yet a prevailing defendant *always* enjoys the potential to recover fees. Thus, without timely registration, you can never obtain attorneys' fees if you prevail in your infringement suit. But, should you lose, the defendant can recover fees against you.

---

46. White v. Samsung Elecs. Am., Inc., 989 F.2d 1512, 1521 (9th Cir. 1993) (Kozinski, J., dissenting).

47. Columbia Pictures Television v. Krypton Broad. of Birmingham, Inc., 106 F.3d 284, 296 (9th Cir. 1997) ("[A] plaintiff in a copyright action is generally awarded fees by virtue of prevailing in the action.").

48. 17 U.S.C. § 412 (2006). The statute states,

> [N]o award of statutory damages or of attorney's fees . . . shall be made for . . . any infringement of copyright commenced after first publication of the work and before the effective date of its registration, unless such registration is made within three months after the first publication of the work.

Id.

This state of affairs significantly dissuades the individual artist from pursuing even a clear-cut case of infringement, lest something go wrong at trial. Even in the most obvious case of infringement, there is always a chance that the case may go awry due to an error in registration, a difference of opinion over "substantial similarity," or a generous reading of the fair use doctrine by the trier of fact.

For example, errors in registration[49] are commonplace and almost always play a role in a defense to an infringement suit. As Charles Ossola reminds us, parties will often spend hundreds of thousands of dollars in a suit dealing with the inevitable claim of fraud on the Copyright Office based on mistakes in an application for registration. As he explains, "[t]here are almost always mistakes, or at least arguable mistakes, [in a registration application] which are invariably discovered during litigation."[50] On occasion, courts have looked askance at such errors, throwing out suits in their entirety, no matter how strong the merits and how blatant the infringement.[51]

---

49. Registration, even if untimely for purposes of § 412, is required to have standing to bring an infringement suit. 17 U.S.C. § 411(b) (2006) ("[N]o action for infringement of the copyright in any United States work shall be instituted until preregistration or registration of the copyright claim has been made in accordance with this title.").

50. Charles Ossola, *Registration and Remedies: Recovery of Attorney's Fees and Statutory Damages under the Copyright Reform Act*, 13 CARDOZO ARTS & ENT. L.J. 559, 561 (1993).

51. *See, e.g.*, Torres-Negron v. J & N Records, LLC, 504 F.3d 151, 162 (1st Cir. 2007) (affirming judgment as a matter of law to defendants on a copyright infringement claim on the grounds of improper registration since the songwriter's deposit of a reconstruction with his registration paperwork resulted in an incomplete application); Morris v. Bus. Concepts, Inc., 259 F.3d 65, 71 (2d Cir. 2001) (affirming grant of summary judgment to defendant for lack of subject matter jurisdiction on the grounds that plaintiff had not properly registered articles which appeared in a magazine since only the magazine itself had been registered); Raquel v. Educ. Mgmt. Corp., 196 F.3d 171, 180 (3d Cir. 1999) (dismissing a suit for lack of standing due to improper registration since the registrant had mischaracterized the work in question as audiovisual rather than musical in nature). *See also* Kodadek v. MTV Networks, Inc., 152 F.3d 1209, 1212 (9th Cir. 1998) (invalidating a copyright registration on the grounds that the works deposited did not constitute bona fide copies of the original works). As Ossola points out,

> If anyone is of the opinion that there are no such registration errors, he should sit through a deposition with a client when he is asked to justify his position on work made for hire in light of Reid factors, joint work in light of the recent case law, or what constitutes preexisting material for the purpose of derivative works. These are all questions that must be filled in on the application registration form, and they each provide fertile territory for attack in litigation.

Ossola, *supra* note 50, at 561.

3. The Inadequacy of Remedies Absent Timely Registration: The Limits of Actual Damages and Injunctive Relief

Absent access to statutory damages and attorneys' fees, a creator has little meaningful ability to punish an infringer or deter future acts of infringement. As Judge Richard Posner has stated,

> [T]here is no basis in the law for requiring the infringer to give up more than his gain when it exceeds the copyright owners' loss. Such a requirement would add a punitive as distinct from a restitutionary element to copyright damages, and . . . the statute contains no provision for punitive damages.[52]

Posner's conclusion is widely shared. As Nimmer observes, "[t]he cases are clear that exemplary or punitive damages should not be awarded in a statutory copyright infringement action."[53] On one hand, it would appear that the Copyright Act provides plenty of room for punitive style damages under the guise of the statutory damages regime and pursuant to the courts' authority to enhance those statutory damages five-fold on the grounds of willful infringement. However, in the vast majority of real world infringements, untimely registration of a copyright precludes a prevailing plaintiff from recovering statutory damages (or any willfulness enhancement), no matter how egregious the conduct of the defendant.

Thus, with punitive damages unavailable, a prevailing unsophisticated creator is left with only actual damages or disgorgement of profits. However, for copyright infringement, actual damages are often speculative and expensive to prove. Moreover, case law interpreting the Copyright Act has specifically excluded psychological injury from the equation for actual damages. Thus, if your artwork is used in an unauthorized manner that is repulsive to you, the modern damages analysis provides you with no relief on these grounds.[54]

---

52. Bucklew v. Hawkins, Ash, Baptie & Co., 329 F.3d 923, 931–32 (7th Cir. 2003) (dictum).

53. 4 NIMMER & NIMMER, *supra* note 29, § 14.02.

54. *Id.* Nimmer observes,

> The Act provides that the "copyright owner is entitled to recover the actual damages suffered by him or her as a result of the infringement. . . ." Yet neither its text nor the Committee Reports attempt to define the nature of those actual damages. Reference must therefore be made to both statutory and common law copyright case law.

Id.; *see, e.g.*, Mackie v. Rieser, 296 F.3d 909, 917 (9th Cir. 2002) (noting that "hurt feelings" cannot form the basis of damages awards under copyright law); Baker v. Urban Outfitters, Inc., 254 F. Supp. 2d 346, 356–57 (S.D.N.Y. 2003) (providing that an award to plaintiff based upon his personal feelings of moral debt is without basis).

Disgorgement of profit can certainly be a useful remedy, but it is often difficult to quantify. While case law does establish a presumption that all profits stem from the infringing act, defendants can rebut this presumption, leading courts to engage in the problematic task of accurate apportionment.[55] Moreover, a defendant will often show no profit. For example, the entertainment industry is notorious for its ability to show a loss on virtually every project, including some of its biggest hits.[56] To scrutinize such troublesome accounting, a plaintiff needs to spend extensive time in discovery without any hope of recovering fees for the effort.

When it becomes too difficult for a plaintiff to demonstrate profit by the defendant from an act of infringement, a few courts have analogized to patent law's reasonable royalty analysis for damages and have allowed the assessment of a hypothetical license value for the unauthorized use of the copyrighted work.[57] But as Nimmer notes, this line of authority is of relatively recent vintage, and only a "smattering" of decisions have explicitly followed the lead of the Seventh Circuit, which first allowed such a recovery in its 1985 ruling.[58] Under this analysis, a court determines ex ante the price

---

55. *See* 17 U.S.C. § 504(b) (2006); Harper & Row Pub., Inc. v. Nation Enter., 471 U.S. 539, 567 (1985). The court noted,

> With respect to apportionment of profits flowing from a copyright infringement, this Court has held that an infringer who commingles infringing and noninfringing elements "must abide the consequences, unless he can make a separation of the profits so as to assure to the injured party all that justly belongs to him."

Id. (quoting Sheldon v. Metro-Goldwyn Pictures Corp., 309 U.S. 390, 406 (1940))

56. *See, e.g.*, Buchwald v. Paramount Pictures, Corp., No. 706083, 1992 WL 1462910, *1–2 (Cal. Super. Ct. Mar. 16, 1992) (declaring Paramount's accounting methodology unconscionable when it showed that the hit movie *Coming to America* had earned no net profits).

57. *See, e.g.*, Deltak, Inc. v. Advanced Sys., Inc., 767 F.2d 357 (7th Cir. 1985). Plaintiffs had not timely registered and therefore could not recover statutory damages. *Id.* at 359. Actual damages proved difficult to ascertain, as there was no provable out-of-pocket harm, such as lost sales, to the plaintiff. *Id.* at 360. Moreover, the court was "unable to determine what portion of the gross revenues were due to the infringement and what portion were due to their factors such as lawful marketing methods," thereby preventing any disgorgement recovery. *Id.* at 359. With what Nimmer dubs "triple circumstances (no out-of-pocket losses to plaintiff, no profits to defendant, no ability to recover statutory damages)" in play, the Seventh Circuit reversed an order denying plaintiffs any damages and allowed recovery of an implied license fee. *Id.* at 364. *See also* 4 NIMMER & NIMMER, *supra* note 29, § 14.02[B][1].

58. *Id.*; *see, e.g.*, Kleier Adver., Inc. v. Premier Pontiac, Inc., 921 F.2d 1036, 1040 (10th Cir. 1990); Roeslin v. Dist. of Columbia, 921 F. Supp. 793, 799–800 (D.D.C. 1995); Kleier Adver. Co. v. James Miller Chevrolet, Inc., 722 F. Supp. 1544, 1546 (N.D. Ill. 1989). *But see* Widenski v. Shapiro, Bernstein & Co., 147 F.2d 909, 911–12 (1st Cir. 1945). The court found that the Copyright Act's provision for statutory damages serves as an absolute bar to the recovery of a reasonable license:

upon which a willing licensor and licensee would have agreed for the use. But unlike patents, copyrighted works are not merely commodities. As such, there are many more unwilling copyright licensors than unwilling patent licensors. One's willingness to license one's invention might differ markedly from one's willingness to license one's song. In the realm of patents, for example, there are no hold-outs from licensing on the grounds that licensing will diminish the inherent beauty of the invention. The same is not true of copyrighted content, largely because of the personal and artistic content that form its subject matter. Consider the reticence of some major bands to license their music for use in advertising. R.E.M. famously rejected an offer of more than $10 million from Microsoft for use of their tune *It's the End of the World as We Know It (And I Feel Fine)* for the launch of Windows 95—though The Rolling Stones were more than happy to step in and ultimately ended up licensing *Start Me Up* for use).[59] However, a medical device maker is

---

> [I]t seems to us highly significant that we have been referred to and have found no case applying the patent rule contended for by the defendant in a copyright case, and that the Supreme Court in the Sheldon case, supra, refused to sanction the closely analogous contention that damages in a copyright case ought to be the price at which the copyright proprietor had indicated his willingness to sell to the infringer.

Id. Cf. Childress v. Taylor, 798 F. Supp. 981, 990 (S.D.N.Y. 1992); Lundberg v. Welles, 93 F. Supp. 359 (S.D.N.Y. 1950). Arguably, older cases finding an absolute bar to the award of a reasonable licensing fee are distinguishable since they were decided under the 1909 Copyright Act. However, even modern cases under the 1976 Copyright Act have concluded that the law bars recovery of a reasonable license fee when there are no lost sales or disgorgable profits. In *Business Trends Analysts, Inc. v. Freedonia Group, Inc.*, 887 F.2d 399 (2d Cir. 1989), for example, the Second Circuit explicitly

> decline[d] to adopt *Deltak's* approach . . . We see no room for such a speculative and artificial measure of damages under Section 504(b) . . . It is surely true that where an infringer such as TFG sells the offending publication at a nominal price, and there is no evidence of lost sales of the infringed publication, a conventional profits test may seem inadequate. Nevertheless, we believe we must follow the statutory scheme.

Id. at 405.

59. *See* Barnet D. Wolf, *Selling Out*, COLUMBUS DISPATCH, Sept. 29, 2002, at 01E. On a related note, when U2 allowed *Vertigo* to be used in an iTunes commercial, the band took pains to explain to fans that the use of the song was like a 30 second music video and that Apple did not pay the band anything directly for use of the song in the advertisement. *See* Chris Ayres, *U2 Online Deal Hastens Last Spin for the CD*, THE TIMES (U.K), Oct. 30, 2004 ("Some fans feel cheated that the band is getting so corporate. Apple's latest iPod advertising campaign . . . features U2 performing their new single, Vertigo—in what could be construed as the band's first commercial endorsement."). However, the band's comments ignored the profit sharing arrangement U2 enjoyed from sales of the special U2 iPod. *Id.* (noting the "unprecedented joint marketing and licensing deal" between U2 and Apple that was "by far the most lucrative [deal] signed by any rock band in history").

unlikely to refuse to license their patented technology for a rate far above market.

The one area of potentially meaningful leverage that an unsophisticated creator does possess is the threat of injunctive relief. But this leverage is limited in several critical ways. To begin with, the creator must first discover the infringement at the ideal time in order to effectively utilize the threat of injunctive relief. This occurs on the eve of a product release, when enjoining the distribution of an infringing product is both at its most feasible and economically painful. Often, however, infringement is not discovered until much later, after the infringer has already enjoyed significant unauthorized use of the work.

Second, the burdens facing a plaintiff seeking injunctive relief in an intellectual property dispute have grown markedly in recent years. Specifically, there is no automatic entitlement to injunctive relief in a copyright dispute, no matter what the merits of a claim. Rather, the choice to grant an injunction—either preliminary or permanent—resides within the sound discretion of the trial court.[60] Courts used to routinely grant permanent injunctions to prevailing plaintiffs in intellectual property cases, absent exceptional circumstances. However, starting with dicta in *New York Times Co. v. Tasini*,[61] and culminating in an express statement in *eBay v. Mercantile Exchange*,[62] which ruled on the issue in the patent context, the Supreme Court has abandoned this general rule.[63] By allowing judges the discretion to transform patent, copyright, and trademark protection from property rights to a liability regime, the Court reasserted the importance of a critical element sometimes overlooked in the adversarial setting: the public interest. Courts

60. 17 U.S.C. § 502 (2006) (providing that a court "may" grant injunctive relief "on such terms as it may deem reasonable to prevent or restrain infringement of a copyright"). Appellate courts have cautioned that preliminary injunctions, even in the intellectual property context, are considered to be "an extraordinary remedy involving the exercise of a very far-reaching power, which is to be applied 'only in [the] limited circumstances which clearly demand it.' " Direx Israel, Ltd. v. Breakthrough Med. Corp., 952 F.2d 802, 811 (4th Cir. 1991) (reversing a grant of preliminary injunction in a trade secret case).

61. 533 U.S. 483, 505 (2001) ("[I]t hardly follows from today's decision [finding infringement] that an injunction against [the infringing use] must issue."); *see also* 17 U.S.C. §502(a) (2006) (stating that a court "may" enjoin infringement); Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 578 n. 10 (1994) (holding that goals of copyright law are "not always best served by automatically granting injunctive relief").

62. eBay Inc. v. MercExchange L.L.C., 547 U.S. 388 (2006).

63. In rejecting this general rule, *eBay* claimed true fealty to the traditional four-part balancing test historically used by courts of equity when contemplating injunctive relief. *Id.* at 392. As such, the Court actually characterizes an automatic-injunction rule as wayward and inconsistent with precedent. Nevertheless, the automatic-injunction rule dominated intellectual property jurisprudence in the twentieth century.

therefore possess the option to order damages but allow an act of infringement to continue unabated. As the Supreme Court held in *eBay*, "this Court has consistently rejected invitations to replace traditional equitable considerations with a rule that an injunction automatically follows a determination that a copyright has been infringed."[64]

Courts have begun to apply the principals of *eBay* in the context of preliminary injunctions as well. A recent wave of cases has questioned the presumption of irreparable harm that all intellectual property plaintiffs used to enjoy when applying for injunctive relief.[65] Of course, there may be plenty of good public policy reasons to make obtaining injunctive relief in intellectual property disputes more difficult, both at the preliminary and permanent levels. As Justice Kennedy noted in his *eBay* concurrence,

> When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest. In addition injunctive relief may have different consequences for the burgeoning number of patents over business methods, which were not of much economic and legal significance in earlier times.[66]

Along the same lines, during a widespread outbreak of Anthrax, it may not make sense to enjoin an infringing company from distributing life-saving drugs to the infected. That said, however, the increased difficulty in obtaining injunctive relief in intellectual property disputes dramatically and disproportionately impacts unsophisticated creators seeking to vindicate their rights.

---

64. *eBay*, 547 U.S. at 392–93.

65. *See, e.g.*, MGM Studios, Inc. v. Grokster, Ltd., 518 F. Supp. 2d 1197, 1212 (C.D. Cal. 2007) ("[T]he longstanding rule that irreparable harm can be a presumed after a showing of likelihood of success for purposes of a copyright preliminary injunction motion may itself have to be reevaluated in light of *eBay*."); Allora, LLC v. Brownstone, Inc., No. 07-87, 2007 WL 1246448, at *5 (W.D.N.C. Apr. 27, 2007) (applying *eBay* to increase the burden on plaintiffs in requests for preliminary injunctions in copyright claims, thereby trumping older circuit court precedent); Canon, Inc. v. GCC Int'l Ltd., 450 F. Supp. 2d 243, 254 (S.D.N.Y. 2006) (drawing upon *eBay* to hold that "the movant must demonstrate the likelihood of irreparable injury in the absence of a grant of the requested [preliminary] injunction"). *Cf.* Lorillard Tobacco Co. v. Engida, 213 Fed. Appx. 654 (10th Cir. 2007) (declining to address whether *eBay* changed the standards for preliminary injunctions in intellectual property cases, but affirming a district court's decision to deny preliminary injunctive relief based on a balance of hardships).

66. *eBay*, 547 U.S. at 396–97 (Kennedy, J., concurring).

Additionally, with the burden on the plaintiff to demonstrate a likelihood of success on the merits, a defendant need only poke sufficient doubt into a single issue, such as registration or ownership, in order to defeat an application for preliminary injunctive relief. Moreover, even if a court grants an injunction, it can be stayed pending an appeal and, in the end, can only issue after the plaintiff posts a bond.[67] Under the Federal Rules, a court must therefore set the security to an amount sufficient "for the payment of such costs and damages as may be incurred or suffered by any party who is found to have been wrongfully enjoined."[68] Thus, a bond can be especially expensive when seeking to halt the distribution of a valuable product and the high price can be cost-prohibitive for many plaintiffs.[69] Finally, the attorneys' fees expended to obtain injunctive relief are never recoupable for a plaintiff who has not timely registered an infringed work. In all, therefore, there are numerous shortcomings to the remedies available for plaintiffs whose works are not timely registered but are willfully infringed.

4. Pitfalls for the Unsophisticated Even with Timely Registration: The Inevitable and Wasteful Scrutinization of Registration Applications.

At the same time, the registration regime elevates form over substance, leading to a disproportional emphasis on compliance with formalities when one attempts to vindicate one's intellectual property rights. Indeed, the most profoundly time-consuming and taxing aspect of many copyright infringement suits is the inevitable attack lodged by defendants against the propriety of the registration. Though this arguably should be one of the least important aspects of litigation—after all, questions of copyright ownership, validity and substantial similarity would appear to trump in significance—registration often becomes a central question because the entire value of the case rides on the issue. If registration is declared invalid, a plaintiff loses standing to bring the suit and must begin the litigation anew after filing a new registration form. More problematically, if the registration is deemed invalid, the plaintiff loses the right to recover statutory damages and attorneys' fees against the defendant, even if the suit is re-filed. Quite simply, valid registration will not have occurred prior to the commencement of infringement.

---

67. FED. R. CIV. P. 65(c).

68. *Id.*

69. *See* Lawrence v. St. Louis-San Francisco Ry., 278 U.S. 228, 233 (1929) (holding that the recipient of a preliminary injunction assumes the risk of "being required to restore [the status quo ex ante] if it should be held that the . . . injunction was improvidently granted . . . and also the risk of having to compensate the [enjoined] . . . for any damages suffered by reason of the [injunction].")

Because of the value of timely registration, procedural formalities are crucial in copyright infringement litigation, often overshadowing the merits of a case. For a system purportedly seeking to vindicate the legitimate rights of creators, this can be a devastating turn. Since statutory damages, attorneys' fees, and even the ability to bring a suit in the first place rely on proper and timely registration, the tendency of some courts to lapse into hyper-formalism when scrutinizing registration applications has dramatic consequences. Although there is ample basis to view errors in the registration process forgivingly,[70] many courts have deviated from this scheme—certainly enough to give unsophisticated plaintiffs pause when pursuing an infringement action. Several cases from the federal circuit courts illustrate this cautionary note.

Consider *Raquel v. Education Management Corp.*,[71] an infringement suit involving the rock band Nirvana. In the case, the copyright holder to the song *Pop Goes the Music* had sued, inter alia, Nirvana and its record label, Geffen, for the unauthorized use of the song in Nirvana's music video *About a Girl*. The defendants responded by arguing, inter alia, that the plaintiff lacked a proper registration for the work. As it turned out, the plaintiff had filed a copyright registration application for a musical composition and described the "nature of the work" as an "audiovisual work." The reason for this designation seemed plausible enough: the claimant had submitted a videotape of a television commercial in which the claimant's song had been performed. Moreover, the claimant had correctly noted on the registration application that the nature of the authorship claim was "[a]ll music and lyrics and arrangement." Nevertheless, the court invalidated the registration on the grounds that the claimant had made a material misrepresentation to the Copyright Office that obfuscated information that would have led to the application's rejection. Moreover, the Court held that the misrepresentation was not inadvertent or innocent, a fact that would ordinarily prevent invalidation. The case drew a vigorous dissent from future Supreme Court Justice Samuel Alito, who charged the majority with irrationally and unfairly elevating form over substance and mandating "a forfeiture of a valid

70. As Nimmer argues, absent fraud, "a misstatement or clerical error in the registration application . . . shall neither invalidate the copyright nor render the registration certificate incapable of supporting an infringement action." 2 NIMMER & NIMMER, *supra* note 29, § 7.20; *see, e.g.*, Harris v. Emus Records Corp., 734 F.2d 1329, 1335 (9th Cir. 1984) ("Absent intent to defraud and prejudice, inaccuracies in copyright registrations do not bar actions for infringement"); Advisers, Inc. v. Wiesen-Hart, Inc., 238 F.2d 706, 708 (6th Cir. 1956) ("[I]nnocent misstatement . . . in the affidavit and certificate of registration, unaccompanied by fraud . . . does not invalidate copyright.").

71. 196 F.3d 171 (3d Cir. 1999).

copyright because of a misstatement that the trial court had already labeled inadvertent."[72]

Indeed, the court's concern about intentional misrepresentation appears particularly misplaced when one considers the real facts. The premise underlying the entire opinion—fraud on the Copyright Office—was simply untrue. As it turns out, the Copyright Office was not misled in any way. The Office took pains to announce this when it formally addressed *Raquel* by issuing a statement of policy on Registration of Claims to Copyright.[73] The Copyright Office unequivocally and resoundingly rejected the reasoning of *Raquel*:

> The Copyright Office is issuing this policy statement to clarify that it was not misled in registering the copyright claim in the *Raquel* case, and that the Copyright Office knew that the copyright claim was in a musical work, and not an audiovisual work. The Office is also issuing this statement to clarify that in the 'nature of this work' space on Form PA, it has been and continues to be acceptable to describe the physical nature of the deposit submitted with the application.[74]

Despite the Copyright Office's firm rebuke of the holding in *Raquel*, it was too late to help the plaintiff. Moreover, despite the Copyright Office's statement of policy, the ultimate question of registration validity remains in the hands of the courts. As the *Raquel* case and others demonstrate, there is always the risk of invalidation of a registration on relatively minor grounds.[75] Indeed, other circuit courts have also invalidated registrations (although not on such a flimsy basis) because of errors contained in the application forms.[76]

At the turn of the last century, the Register of Copyrights at the time, Thorvald Solberg, expressed his profound distaste for the registration regime and its ability to punish seemingly innocuous errors and omissions with the

---

72. *Id.* at 182 (Alito, J., dissenting).

73. Registration of Claims to Copyright, 65 Fed. Reg. 41508-09 (Jul. 5, 2000).

74. *Id.*

75. *See, e.g.*, Tavory v. NTP, Inc., 297 Fed. App'x 986 (Fed. Cir. 2008) (affirming dismissal of copyright infringement claim on the grounds of improper registration because the programmer's deposit copy was not an original or bona fide copy); Torres-Negron v. J & N Records, LLC, 504 F.3d 151 (1st Cir. 2007) (affirming dismissal of copyright infringement claim on the grounds of improper registration because the songwriter's submission of a reconstruction with his registration application resulted in an incomplete application).

76. Kodadek v. MTV Networks, Inc., 152 F.3d 1209 (9th Cir. 1998) (invalidating a registration because of a mistake in the application in an infringement suit involving the Beavis and Butthead characters from MTV); Whimsicality, Inc. v. Rubie's Costume Co., 891 F.2d 452 (2d Cir. 1989) (invalidating registrations because of a mistake in the application in an infringement suit involving popular Halloween costumes).

dramatic loss of substantive rights. In a report to the Librarian of Congress dated December 1, 1903, he wrote, "[A] system has gradually grown up under which valuable literary rights have come to depend upon exact compliance with these statutory formalities which have no relation to the equitable rights involved, and the question may very well be raised whether this condition should be continued."[77] Over a century later, the same concern continues to resonate.

5. Comparing American and Foreign Infringement Remedies

When one considers the remedies available to any creator under foreign laws, the tremendous disadvantages facing unsophisticated creators in the United States become all the more remarkable. Compare our infringement remedies to those of the United Kingdom and Canada—two countries whose legal regimes are most closely aligned to our own. The United Kingdom has no government registration system at all. And while statutory damages are not available, punitive damages—called " 'additional' damages"—are. The United Kingdom's Copyright, Designs and Patent Act 1988 provides that a plaintiff can recover actual damages plus 'additional' damages to both deter future infringers and punish defendants who willfully violate a plaintiff's intellectual property rights.[78] Furthermore, prevailing plaintiffs (whether in copyright cases or otherwise) recover attorneys' fees.[79]

Canada similarly lacks a registration requirement. And although the country's remedies are more limited in some ways, they are also more expansive in other ways. First, any prevailing plaintiff in a copyright infringement suit is entitled to recovery of statutory damages and attorneys' fees, regardless of the existence or date of any copyright registration.[80] But, to

77. THORVALD SOLBERG, LIBRARY OF CONGRESS, REPORT ON COPYRIGHT LEGISLATION 25 (1904).

78. Copyright, Designs and Patent Act, 1988, ch. 48, § 97(2) (U.K.) (allowing award of "additional" damages based on "flagrancy of the infringement" and "benefit accruing to the defendant by reason of the infringement").

79. Copyright, Designs and Patent Act, 1988, ch. 48, §§ 96(2), 103 (U.K.).

80. For statutory damages *see, e.g.*, Copyright Act, R.S.C., ch. 42, § 34 (1985) (Can.), amended by 1997 S.C., ch. 24 (Can.) ("Where copyright has been infringed, the owner of the copyright is, subject to this Act, entitled to all remedies by way of injunction, damages, accounts, delivery up and otherwise that are or may be conferred by law for the infringement of a right."); Copyright Act, R.S.C., ch. 42, § 38.1(1) (1985) (Can.), amended by 1997 S.C., ch. 24 (Can.) ("Subject to this section, a copyright owner may elect, at any time before final judgment is rendered, to recover, instead of damages and profits referred to in subsection 35(1), an award of statutory damages for all infringements involved in the proceedings"). For attorneys' fees, *see* Copyright Act, R.S.C., ch. 42, § 34(3) (1985) (Can.), amended by 1997 S.C., ch. 24 (Can.) (granting courts the discretion to grant full costs, including attorneys' fees, to a prevailing party).

counter this expansion in remedy eligibility, Canadian law limits the amount of statutory damages to a maximum award of CN $20,000 per act of infringement[81]—a small fraction of the maximum statutory damages award allowed under American law.[82] In addition, courts have discretion to award exemplary damages to punish infringers and effectively deter future infringements.[83] Thus, in both Canada and the United Kingdom, sophisticated or unsophisticated creators share an equal footing, as far as available remedies for the vindication of their copyright interests go. Additionally, unlike the United States, both the United Kingdom and Canada recognize non-economic injuries, such as moral prejudice or harm to one's reputation, as cognizable damages in an infringement suit. For example, Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights instructed member states to grant judges the authority to fashion infringement awards, when appropriate, based on "moral prejudice"[84] to rights holders. Pursuant to this directive, the United Kingdom passed the Intellectual Property (Enforcement, etc.) Regulations in 2006.[85] The Regulations instruct judges to take into account "all appropriate aspects" including negative economic consequences and non-economic factors such as "moral prejudice caused by the infringement."[86] In Germany, for example, authors, photographers, and

81. Copyright Act, § 38.1(1).

82. 17 U.S.C. § 504(c)(2) (2006) ("In a case where the copyright owner sustains the burden of proving, and the court finds, that infringement was committed willfully, the court in its discretion may increase the award of statutory damages to a sum of not more than $150,000.").

83. Manitoba Inc. v. Parks, [2007] N.S.J. No. 128, 2007 NSCA 36 (Can.) (citing GEORGE S. TAKACH, COMPUTER LAW 122-23 (2nd ed. 2003)). The court stated,

> A court may award damages for copyright infringement even where the infringer made no profits. The Copyright Act also contains a statutory damages provision that permits a court to award monetary damages between $500 and $20,000. Punitive or exemplary damages for copyright infringement or trade secret misappropriation can also be awarded where the defendant's conduct is egregious and shows virtual contempt for the intellectual property rights of the plaintiff.

Id.; *see also* The Queen v. James Lorimer & Co., [1984] 1 F.C. 1065 (Can.) ("[I]t is well established that [exemplary damages] are, in appropriate circumstances, available . . . , [and there is] no reason why appropriate circumstances should be different in the case of copyright infringement than in the case of any other civil invasion of another's rights."); Osmont v. Petit Journal Inc., [1934] 73 Que. S.C. 465, 473 (Can.) (providing for availability of exemplary damages to punish copyright infringement as a species of theft).

84. Council Directive 2004/48, art. 13, § 1, 2004 O.J. (L 157) 78 (EC).

85. Intellectual Property (Enforcement, etc.) Regulations, 2006, S.I. 1028, art. 3, ¶¶ 1, 2 (U.K.).

86. *Id.* ¶ 2.

performers may "recover, as justice may require, a monetary indemnity for the injury caused to them even if no pecuniary loss has occurred."[87] Thus, even in absence of actual economic damages, German courts can fashion awards based on subjective, non-economic harms stemming from infringement. Similarly, in Canada, a computation of actual damages, if elected in lieu of statutory damages, can include a claim of non-economic injury, i.e., harm to a copyright holder's reputation.[88]

All told, the United States' registration regime fails to protect unsophisticated creators adequately due to the unavailability of statutory damages even for post-registration infringement, the one-way risk of attorneys' fees, and the absence of any punitive or reputational damages. This situation is all the more concerning when compared to the state of protection under foreign regimes, including those close to our own.

D. REGISTRATION AND THE FAILURE OF THE NOTICE FUNCTION

Besides putting unsophisticated creators at a profound disadvantage when seeking to vindicate their legal interests, the registration system fails to achieve even its basic function of notice. Despite our rhetorical distaste for copyright formalities, courts and commentators have explained the endurance of the registration regime based on the important notice function it serves. As numerous jurists have asserted, the registration requirement "encourages potential infringers to check the Copyright Office's database" to ascertain protection status.[89] However, this claim is vastly overstated.

At the outset, it is a dubious proposition that potential infringers would even check the database at all, let alone prior to their infringement. Such an assumption may have made marginally more sense in previous centuries, when the law provided greater incentives to conduct a registration check. Prior to the enactment of the 1909 Copyright Act, for example, a work did not receive any copyright protection at all unless it was registered prior to, or

---

87. Bürgerliches Gesetzbuch [BGB] [Civil Code] Sept. 9, 1965, BCB II 27 at 1273, § 97(2), as amended by the Law of July 16, 1998 (F.R.G.).

88. Groller v. Wolofsky, [1934], 72 Que. S.C. 419 (Can.). For a more extensive discussion on damages available, see Chaplin v. Hicks, [1911] 2 K.B. 786.

89. Derek Andrew, Inc. v. Poof Apparel Corp., 528 F.3d 696, 700 (9th Cir. 2008); *see also* Johnson v. Jones, 149 F.3d 494, 505 (6th Cir. 1998). The court noted that

> [i]n addition to giving copyright owners incentive to register, [section] 412 also provides potential infringers with an incentive to check the federal register. If [section] 412 succeeds in encouraging copyright owners to register and in encouraging potential infringers to check registration, then it will have reduced both the search costs imposed on potential infringers and the enforcement costs borne by copyright owners.

Id.

simultaneous with, first publication.[90] The 1909 Act, which governed until December 31, 1977, reduced this formality somewhat by allowing a statutory copyright so long as a registration took place, and a deposit was made, "promptly" after publication.[91] Although the promptness requirement was substantially eviscerated with the Supreme Court's 1939 ruling in *Washingtonian Publishing Co. v. Pearson*,[92] delayed registration could give rise to a laches defense preventing enforcement of a copyright.[93] Timely registration remained a requirement until 1964 in order to have the option to renew a copyright after the first twenty-eight year term. Thus, in the past, registration determined the copyright status of many works. Failure to conform to certain formalities, such as the provision of proper copyright notice on a work, was fatal to a work's protection.[94] Since many published works were not registered, renewed, or published with certain notice formalities, there was a decent chance that any given work had no copyright protection. However, copyright now subsists from the moment of creation for all works created after January 1, 1978, meaning that just about any creative work authored in the past few decades enjoys copyright protection.[95] As a result, all creative works are copyrighted and are the potential subject of a lawsuit, regardless of registration status.[96]

At the same time, for works registered prior to 1978, the Catalog is not available in any official online format. It is only accessible by visiting the Copyright Office Public Records Room, by paying the Copyright Office to conduct a check, or by accessing a copy of the Office's Catalog of Copyright

---

90. 2 NIMMER & NIMMER, *supra* note 29, § 7.16[A][2][b].

91. Copyright Act of 1909 §§ 13–14 (codified as amended at 17 U.S.C. 11, 61 Stat. 652 (1947)) (repealed by Copyright Act of 1976, Pub. L. No. 94-553, 90 Stat. 2541).

92. 306 U.S. 30 (1939), *reh'g denied*, 306 U.S. 668 (1939).

93. *See, e.g.*, Samet & Wells, Inc. v. Shalom Toy Co., No. 74-C-695, 1975 U.S. Dist. LEXIS 13683, at *20 (E.D.N.Y. Feb. 24, 1975) ("[T]he delay in filing the copyright notice may prevent plaintiff from complaining of any [infringing] titles which were sold [prior to filing of registration]."); Kontes Glass Co. v. Lab Glass, Inc., 250 F. Supp. 193 (D.N.J. 1966), *aff'd*, 373 F.2d 319 (3d Cir. 1967) (stating that delay in registration may create a defense of laches).

94. 2 NIMMER & NIMMER, *supra* note 29, § 7.14[A][1] (describing how failure to observe proper notice requirements on a work used to lead to its dedication to the public domain); *see* Neimark v. Ronai & Ronai, LLP, 500 F. Supp. 2d 338, 341 (S.D.N.Y. 2007) (noting that "works published without a copyright notice prior to the enactment of the Berne Convention on March 1, 1989 are injected into the public domain and thus lose any copyright protection to which they might otherwise have been entitled").

95. One notable exception is works authored by citizens of countries that have not signed the Berne convention or that do not have copyright laws.

96. However, available remedies are profoundly affected by registration status. This affects the viability of many potential lawsuits.

Entries. Even after conducting a thorough search, however, one cannot be sure of a work's copyright status. The Office's own searches produce only a "factual, noninterpretive report"[97] and, as the Office takes pains to caution, no investigation, no matter how comprehensive, can determine copyright status with certainty: "Copyright investigations often involve more than one of these methods [examining the work for proper copyright notice, searching the Copyright Office catalog, having the Copyright Office make a search for you]. Even if you follow all three approaches, the results may not be conclusive."[98]

Thus, on the off-chance that potential infringers do check the Copyright Office's records, they will frequently have difficulty gleaning accurate information about the registration status of a particular work. As regular practitioners know, the Office's database is not up-to-date. Although registration becomes effective upon the Copyright Office's receipt of a complete application,[99] a copyright registration certificate usually does not issue for several months—until after the Copyright Office has had the opportunity to evaluate and process an application. Even after a certificate is issued, there is an additional delay before the information is entered into the database. Although the Copyright Office's recent move to elective electronic registration may reduce some of these delays, there is still a significant gap between effective registration and the availability of such information on the Copyright Office's database.[100]

Additionally, the Copyright Office database—at least in its most accessible, online form—does not contain a single image or copy of a registered work. Instead, one must search for a work via text alone.

---

97. U.S. COPYRIGHT OFFICE, CIRCULAR 23, THE COPYRIGHT CARD CATALOG AND THE ONLINE FILES OF THE COPYRIGHT OFFICE 1 (2009), *available at* http://www.copyright.gov/circs/circ23.pdf.

98. U.S. COPYRIGHT OFFICE, CIRCULAR 22, HOW TO INVESTIGATE THE COPYRIGHT STATUS OF A WORK 1 (2009), *available at* http://www.copyright.gov/circs/circ22.pdf [hereinafter CIRCULAR 22].

99. 2 NIMMER & NIMMER, *supra* note 29, § 7.16[B][1][a][i]. Nimmer observes that

> [17 U.S.C. § 411(a)] provides that "[t]he effective date of a copyright registration is the day on which an application, deposit, and fee, which are later determined by the Register of Copyrights or by a court of competent jurisdiction to be acceptable for registration, have all been received in the Copyright Office." The legislative history explains that "[w]here the three necessary elements [of application, deposit and fee] are received at different times, the date of receipt of the last of them is controlling . . . ."

Id.

100. CIRCULAR 22, *supra* note 98, at 4 ("Since searches are ordinarily limited to registrations that have already been cataloged, a search report may not cover recent registrations for which catalog records are not yet available.").

Moreover, textual searches yield only the information that an applicant has actually provided. Matching a work to a registration can, therefore, represent a task rife with uncertainty. For example, imagine that you are a potential infringer who wants to use a photograph documenting the infamous night in 2006 when Hollywood starlets Britney Spears, Lindsay Lohan, and Paris Hilton hit Los Angeles's infamous Sunset Strip to celebrate Britney's divorce from Kevin Federline. That November evening, the paparazzi were on the prowl and they caught the moment on film. The photographs hit the Internet and caused an immediate sensation, as several candid shots caught Britney *in flagrante commando*. Spears's unfortunate decision to disregard undergarments that evening led to the exposure of her nether regions to the world. It also raised two central legal issues. First, it traumatized millions, resulting in a potentially viable action for intentional infliction of emotional distress. Second, and less facetiously, it spurred a wave of copyright infringement. Within hours, thousands of blogs reproduced the images without authorization so that they could feature them on their front pages.

If you were seeking to use the photographs legally, you could have approached one of the sources of the photographs: X17.[101] However, as easy as it would have been to contact X17 and obtain a license (for the right price, of course), it would have been impossible to determine whether the photographs were registered with the Copyright Office. First, since the photographs had just been published, there was no way to know yet about their registration status. Second, even if you were seeking to use the photographs a year later, definitively determining registration status from the Copyright Office would be next to impossible. After all, you can only search by text and not image on the database. And, to make things worse, the work lacks a determinate title. While movies and music usually have offical titles, photographs often do not. Consider some of the most famous images of the 20th century. With a simple description, most readers will immediately recall the photographs to which I am referring: Mohammed Ali (then Cassius Clay) lording over a defeated Sonny Liston,[102] an unnamed couple kissing jubilantly at celebrations marking the end of World War II in New York,[103] troops raising the American flag on Iwo Jima,[104] the first panoramic color view of planet Earth from space,[105] or the iconic shot of Che Guevara qua

101. X17 is one of Hollywood's leading celebrity and news photography agencies.

102. Muhammad Ali Knocks Out Sonny Liston (Photograph) (1965); Muhammad Ali Taunting Sonny Liston (Photograph) (1965).

103. Alfred Eisenstaedt, *V-J Day in Times Square* (Photograph) (1945).

104. Joe Rosenthal, *Raising the Flag on Iwo Jima* (Photograph) (1945).

105. Apollo 17 Crew, *The Blue Marble* (Photograph) (1972). Because the image is likely

revolutionary.[106] But while recognizing these photographs may be easy, ascertaining their registration statuses or their titles is something else altogether.

For example, to check on the registration status of the Che Guevara photo, it would help to know that the work is actually titled *Guerrillero Heroico*. Even then, the information you learn might lead you to the wrong conclusion. On the Copyright Officer's website, one would find a registration to the work that was based upon the Uruguay Round Agreement Act (URAA). The URAA restored copyright protection to certain foreign works that as of January 1, 1996 had fallen into the public domain in the United States because of a failure to comply with certain American formalities. The first problem with the registration is that it is unclear which photograph the registration relates to—the original version of the Che Guevara photograph or the more stylized, cropped version of the photograph that is more famous. Regardless of which version the registration refers to, even though the copyright records suggest the photograph is protected, the work is likely in the public domain. According to the photographer,[107] the original version was first published in Cuba around 1960, where copyright protection for a photograph extended only ten years from the date of first use.[108] Thus, the photograph has fallen into the public domain in Cuba. But assuming the photograph originally received protection in the United States at the time of

---

considered a government work, it is in the public domain. 17 U.S.C. § 105 (2006) ("Copyright protection under this title is not available for any work of the United States Government.").

106. Alberto Korda, *El Guerrillero Heroico* (Photograph) (1960).

107. Sarah Levy, *A Copyright Revolution: Protecting the Famous Photograph of Che Guevara*, 13 L. & BUS. REV. AM. 687, 693 (2007). Levy notes that

> [t]he newspaper likely did not request that Korda photograph Guevara specifically, as evidenced by the fact that the newspaper did not even use the photo in its article about the funeral the following day. Korda recalled that the newspaper did keep the photo on file, however, and used it in a subsequent publication alongside an announcement that Guevara would be speaking at a public event.

Id. It should be noted that others claim that publication did not occur until later. *See id.* ("[O]ther sources claim the photo remained unpublished in Korda's studio for the next seven years, leaving Korda's possession for the first time in 1967 when an Italian publisher named Giangiacomo Feltrinelli requested a copy of the image.").

108. Copyright Law, Gaceta Oficial de la República de Cuba, No. 49, art. 47, 30 de diciembre de 1977 (Cuba), *translated in* 1 COPYRIGHT LAWS & TREATIES OF THE WORLD (U.N. Educ., Scientific & Cultural Org. et al. eds., 2000). Decree Law no. 156, September 28, 1994 extended the copyright term for photographs to twenty-five years from date of first publication. However, Korda's work would have already fallen into the public domain by 1971. Moreover, even if Decree Law No. 156 resurrected protection for works already in the public domain, the copyright would have expired in 1986.

its Cuban publication, its present status in the United States is determined by the fact that no renewal application was filed for the photograph in the twenty-eighth year after its publication, leading it to fall into the American public domain no later than 1988. And even though the URAA restored American copyrights for certain foreign works, it only applies to works that are still in copyright in the country of first publication. Since the photograph has likely fallen into the public domain in Cuba, it is ineligible for restoration. A further wrinkle is that the UGAA may not even be constitutional in the first place.[109] Thus, the work is likely in the public domain in the United States, even though its registration status claims otherwise. [110] The cropped version—to the extent that it does not add substantial creativity to the original—is therefore also likely to be in the public domain.[111] Thus, even if one effectively combs through the Copyright Office's registration documents, one might conclude that the work is both still in copyright and registered such that an infringer would face the enhanced penalties of statutory damages and attorneys' fees if caught engaging in unauthorized use. In fact, the work is likely free for anyone to use as it is in the public domain.

Moreover, even if one definitively concludes that no registration under the name of a copyright owner exists, that does not end the inquiry on registration. Copyrights may also be registered by an exclusive licensee of a work.[112] As a result, determining that a copyright owner has not registered its own work does not conclusively settle the issue of registration and protection.[113] In fact, the work may have been registered by any number of

109. *See, e.g.*, Golan v. Gonzales, 501 F.3d 1179, 1197 (10th Cir. 2007) (finding that the URAA does not violate the Copyright Clause, but remanding to the district court for First Amendment review).

110. *But see* Levy, *supra* note 107 (concluding that the work may still be protected).

111. *Cf.* Hearn v. Meyer, 664 F. Supp. 832 (S.D.N.Y. 1987) (suggesting that more than minimal creativity is required to copyright modifications to a work in the public domain). *But cf.* Feist Publ'ns v. Rural Tel. Serv. Co., Inc., 499 U.S. 340 (1991) (holding that, generally, only minimal creativity is required for copyrightability).

112. *See* 17 U.S.C. § 408(a) (2006) ("[T]he owner of copyright or *of any exclusive right in the work may obtain registration* . . . .") (emphasis added); 37 C.F.R. § 202.3(c)(1) (2009) ("An application for copyright registration may be submitted by any author or other copyright claimant of a work, or the owner of any exclusive right in a work, or the duly authorized agent of any such author, other claimant, or owner.").

113. *See* Huthwaite, Inc. v. Sunrise Assisted Living, Inc., 261 F. Supp. 2d 502, 510 (E.D. Va. 2003) (noting that it is not necessary that "the party bringing the infringement must have itself registered the claim"); Tang v. Hwang, 799 F. Supp. 499, 503–05 (E.D. Pa. 1992) (noting that "there is no requirement under the statute that the only person who may bring an action is the person who applies for the copyright registration" and that the law merely provides that "there must be registration of the copyright"); 3 NIMMER & NIMMER, *supra* note 29, § 12.02[B] (stating that "the plaintiff in court obviously need not be the same party

the owners' exclusive licensees, a potentially large pool of entities given the difficulty courts have faced in drawing a clear distinction between exclusive and non-exclusive licensees.[114]

Additionally, many works (including photographs) may be constructively registered if they are published as part of a collection, such as a periodical, that is itself registered. In such cases, the Copyright Office records do not separately list the contents of such collections. As the Copyright Office concedes in its Circular 22, "Individual works such as stories, poems, articles, or musical compositions that were published as contributions to a copyrighted periodical or collection are usually not listed separately by title in our records."[115] Nevertheless, courts have conferred the benefits of registration to such individual works in a variety of circumstances. For example, in *Abend v. MCA, Inc.*, the Ninth Circuit found that a magazine publisher's registration of a blanket copyright for a particular issue effectively registered a story within the issue, although the author of the story had conveyed to the publisher only magazine publication rights and retained all other rights and the author did not separately register any copyright for the story.[116] Thus, the benefits of registration were found to extend to a smaller work contained in a larger work, even when the authors of the two works (the magazine as a whole versus the story itself) were different. Although *Abend* does not deal with the availability of statutory damages and attorneys' fees—it was decided under the 1909 Copyright Act, when timely registration was not a prerequisite for such relief—it does suggest that the benefits of a collection's registration could extend to all of the works within the collection. Other courts have made similar suggestions.[117]

---

who initially registered the subject work").

114. 3 NIMMER & NIMMER, *supra* note 29, § 10.02. Nimmer explains that

> [a]n exclusive license, even if it is "limited in time or place of effect," is equated with an assignment, and each is considered to be a "transfer" of copyright ownership. Nonexclusive licenses, however, do not constitute "transfers," and some residue of the impact of indivisibility with respect to licenses under the 1909 Act remains under the current Act *vis-a-vis* nonexclusive licenses.

Id. (internal citations omitted); David C. Tolley, Note, *Regulatory Priorities Governing Stem Cell Research in California: Relaxing Revenue Sharing & Safeguarding Access Plans*, 23 BERKELEY TECH. L.J. 219, 240 (2008) ("The distinction between an exclusive and a non-exclusive license is not easy to draw in practice. As one treatise author points out, 'commercial practice yields a wide variety of differing transactional frameworks . . . making drawing a simple distinction between exclusive and nonexclusive licenses difficult.' ") (quoting RAYMOND NIMMER & JEFF DODD, MODERN LICENSING LAW § 5:1 (2007)).

115. CIRCULAR 22, *supra* note 98, at 3.

116. Abend v. MCA, 863 F.2d 1465, 1469 (9th Cir. 1988).

117. *See, e.g.*, Kay Berry, Inc. v. Taylor Gifts, Inc., 421 F.3d 199, 204–05 (3d Cir. 2005)

In the end, this exhaustive examination of the vagaries of the search process and the law surrounding it leads to an inescapable conclusion: registration status can be difficult to determine with certainty. Returning to our example involving Britney Spears and her night on the town with Paris Hilton and Lindsay Lohan, this point becomes clear. In one of my more dubious professional accomplishments, I actually registered the copyright to a set of these photographs.[118] Late one night, I was left to decide the title of each photograph, including the infamous "upskirt" shots. In an homage to Spears's hit, *Oops! . . . I Did It Again*, I briefly contemplated going with *Oops! . . . She Did It Again*. But I ultimately went with the distinctly less titillating, but eminently more descriptive, appellations *Britney Spears exposes her derriere*[119] and *Britney Spears exposes herself (again)*,[120] so as to increase the possibility (no matter how unlikely) that would-be infringers could in fact identify the registration status of the work. However, as this example illustrates, there is no assurance that the title will match the work at all.

All told, the registration requirement—at least as presently implemented—fails its basic notice function. It is doubtful whether would-be infringers would really engage in ex ante consultation of the registration rolls. Even if they did, they would have trouble finding a definitive answer on registration status. For works registered before 1978, registration records are not easily available. For more recent works, the available database is not up to date. Information provided in the database is text-only, making it difficult to identify the registration status of certain works. This problem is especially acute for visual works. Moreover, divining the title of some works can prove to be challenging. Furthermore, registration can be achieved not only by a copyright holder but also by any of its exclusive licensees, of which there may be several. And works can be constructively registered as a part of a larger

---

(finding that a garden sculpture included in a copyrighted catalog of sculptures was an individually recognizable element of that single work, and thus was entitled to the benefits of the catalog's copyright registration, regardless of whether it was "related" to other sculptures in catalog); Warren v. Fox Family Worldwide, Inc., 171 F. Supp. 2d 1057, 1065 (C.D. Cal. 2001) (finding that a copyright registration of a motion picture or television show serves to register the musical compositions contained on the soundtrack of the film or show); Greenwich Film Prods., S.A. v. DRG Records, Inc., 833 F. Supp. 248, 250 (S.D.N.Y. 1993) (musical compositions in motion picture were registered with copyright office by virtue of registration of motion picture in which they were contained, and musical compositions did not have to be separately registered).

118. The registrations were part of a suit by X17, Inc. against Mario Lavandeira, aka Perez Hilton, a celebrity gossip blogger accused of infringing X17's copyrighted photographs en masse. *See* X17 v. Lavandeira, 563 F. Supp. 2d 1102 (C.D. Cal. 2007).

119. X17, Inc., VA0001390186, registered December 1, 2006.

120. X17, Inc., VA0001390193, registered December 1, 2006.

work. In short, the registration system remains relatively opaque, thereby undermining its purported utility.

E. REGISTRATION, REMEDIES, AND INTERNATIONAL TREATY OBLIGATIONS

Besides failing its basic notice function, our registration system also runs into a legal concern that has remained largely unscrutinized: it potentially shirks our international treaty obligations. The timely registration requirement for statutory damages and attorneys' fees arguably flouts the tenets of the Berne Convention. Moreover, the potential incompatibility of the United States' registration regime with international law becomes even clearer when one considers the language of the WIPO Copyright Treaty. These issues are particularly salient at a time when we are taking other countries to task for their failure to honor international copyright obligations. Moreover, it is difficult to call out foreign countries for their alleged disrespect for the value of intellectual property when we do not adequately protect many domestic creators.

The world's oldest international copyright agreement, the Berne Convention for the Protection of Literary and Artistic Works, was first drafted in 1886. But the United States did not accede to the Convention until 1988, more than a century later.[121] According to the principal congressional architect of its implementation, the prohibition on formalities is "the central feature of Berne."[122] According to conventional wisdom, therefore, Berne's implementation in 1988 eliminated most of the remaining formalities of the American copyright regime. Yet, despite this, the requirement for timely registration in order to recover statutory damages or attorneys' fees continues to survive. All the while, many observers have blithely rejected any possible conflict between § 412 and the dictates of the Convention. For example, as Ralph Oman, who drafted the Practicing Law Institute's publication *The Impact of Berne on U.S. Copyright Law*, dismissively argues, "we could scotch the requirement of timely registration as a precondition to statutory damages and attorneys' fees, but, whatever the policy arguments pro and con, it is difficult to argue with a straight face that this fringe-benefit is a formality barred by the Berne Convention."[123]

---

121. Berne Convention Implementation Act of 1988, Pub. L. No. 100-568, 102 Stat. 2853 (codified as amended in scattered sections of 17 U.S.C.).

122. 134 CONG. REC. H3079-02 (daily ed. May 10, 1988) (statement of Rep. Kastenmeier).

123. Ralph Oman, *The Impact of the Berne Convention on U.S. Copyright*, PATENT, COPYRIGHTS, TRADEMARKS, AND LITERARY PROPERTY COURSE HANDBOOK SERIES,

It is certainly possible that the registration prerequisite for statutory damages and fees does not violate Berne. After all, it was on that understanding that the Berne Convention Implementation Act of 1988 was passed. Before the United States formally enacted the Convention, the State Department put together the Ad Hoc Working Group on U.S. Adherence to the Berne Convention to evaluate areas of American copyright law falling outside of the parameters of Berne. The Group's report formed the basis for the Implementation Act, which, in the words of one scholar, took a "minimalist approach to adherence."[124] Upon consideration of whether § 412's registration prerequisite for statutory damages and attorneys' fees constituted an impermissible formality, the Ad Hoc Group glibly concluded that "[s]ection 412 is compatible with Berne since it deals with certain specific remedies rather than the ability to obtain redress at all."[125] Nimmer appears to agree, concluding that "Berne imposes a condition that copyright subsistence for works emanating from other member states may not be premised on formal requirements. It does not, however, prohibit formalities as a condition to certain types of remedies, licenses, exemptions, etc."[126]

However, the contrary position—that the statutory damages prerequisite violates Berne's prohibition on formalities—is hardly as implausible as Oman and others may suggest. Here, perhaps, Oman's impressive background comes into play: he spent almost a decade as the Registrar of the Copyright Office, where he led the federal government's efforts to enter the Berne Convention and served as chief counsel of the Senate Subcommittee on Patents, Copyrights, and Trademarks.[127] As a result, he may be disinclined to second-guess his own work by legitimizing questions as to whether our statutory damages and fees scheme complies with Berne. Additionally,

---

Practising Law Institute, PLI Order No. G4-3981, 455 PLI/Pat 233, 255 (October, 1996).

124. William Belanger, *U.S. Compliance with the Berne Convention*, 3 GEO. MASON INDEP. L. REV. 373, 393 (1995).

125. *Final Report of the Ad Hoc Working Group on U.S. Adherence to the Berne Convention, reprinted in* 10 COLUM.-VLA J.L. & ARTS 513 (1986).

126. 4 NIMMER & NIMMER, *supra* note 29, § 17.01(B).

127. U.S. Copyright Office, http://www.copyright.gov/history/bios/oman.pdf. Oman's bio states,

> In 1982, Mr. Oman became Chief Counsel of the newly revived Subcommittee on Patents, Copyrights, and Trademarks, and in 1985 he scheduled the first Senate hearing in 50 years on U.S. adherence to the Berne Convention for the Protection of Literary and Artistic Works. From the Chief Counsel position, he was appointed Register of Copyrights on September 23, 1985. As Register, Mr. Oman helped move the United States into the Berne Convention in 1989.

Id.

Oman's characterization of statutory damages as mere *fringe* benefits is intellectually disingenuous. Rather than a peripheral or secondary feature of copyright law, statutory damages are in practice the only effective means under U.S. law by which plaintiffs can enforce their copyrights in a matter that deters future infringements.

Indeed, there is reason to believe that the registration prerequisite for statutory damages and attorneys' fees may actually violate treaty obligations such as the Berne Convention. In relevant part, the Berne Convention dictates that "[t]he enjoyment and the exercise of these rights shall not be subject to any formality."[128] These rights include the exclusive right of reproduction of a copyrighted work.[129] Because copyright registration is undoubtedly a formality, the threshold question is whether it affects a copyright holder's "enjoyment and exercise" of rights purportedly secured by the Berne Convention.

Though Oman, the Ad Hoc Group, and Nimmer appear to draw a distinction between copyright subsistence and copyright remedies—arguing that Berne prohibits formalities attaching to the former but not the latter—such a conclusion is not inescapable. The language of Berne draws no clear-cut subsistence/remedies dichotomy. Instead, it speaks of the "enjoyment and exercise" of rights, something to which remedies are inextricably related. Consider the most foundational case in American constitutional jurisprudence, *Marbury v. Madison*, which advanced the critical link between the creation of a right and the affordance of a meaningful remedy for violation of that right.[130] Admittedly, there is no language in the Berne Convention that explicitly requires a member state to provide prevailing plaintiffs with statutory damages or even attorneys' fees. Indeed, the Convention makes no mention of remedies whatsoever. However, one can argue—with a straight face, to boot—that enjoyment and exercise of one's

---

128. Berne Convention for the Protection of Literary and Artistic Works art. 5(2), Sept. 9, 1886, as last revised at Paris, July 24, 1971, 828 U.N.T.S. 221, S. TREATY DOC. NO. 99-27.

129. *Id.* at art. 9(1).

130. Marbury v. Madison, 5 U.S. 137, 163 (1803). The court stated,

> The very essence of civil liberty certainly consists in the right of every individual to claim the protection of the laws, whenever he receives an injury. One of the first duties of government is to afford that protection. . . . The government of the United States has been emphatically termed a government of laws, and not of men. It will certainly cease to deserve this high appellation, if the laws furnish no remedy for the violation of a vested legal right.

Id.

exclusive right of reproduction necessarily requires the ability to *deter* infringement.[131]

Thus, while Berne may not ban a copyright registration system that serves a procedural end, its language appears to render any copyright registration system that affects substantive rights, including significant remedies, suspect. In the absence of an ability to pursue statutory damages and attorneys' fees, there is little deterrent effect in copyright enforcement under American law. Indeed, the worst-case scenario for a defendant is that they might have to pay damages ex post in an amount similar to what they might have paid for a license ex ante. They may also face an injunction, but not one that issues automatically issue upon a finding of infringement. Moreover, making attorneys' fees unavailable to a prevailing plaintiff in an infringement case stifles his ability to be made whole for the injury to his rights. Over the years, I have been approached by hundreds of artists who have no viable infringement suit against monied defendants who have undoubtedly and willfully infringed their copyrights. The first question that any experienced copyright litigator asks a potential plaintiff-client is: "Were the works registered before the infringement occurred?" If the answer is negative, the artist is often left with only extralegal means, such as moral force or business sanctions (where available), to rectify the wrongdoing.

Furthermore, even if one accepts the view that Berne does not proscribe the existence of formal prerequisites for certain remedies, that does not end the discussion on international obligations. More recent treaties also need to be considered. For instance, the WIPO Copyright Treaty (WCT) calls into serious question the continued viability of the registration prerequisite for statutory damages and attorneys' fees.

The United States played a key role in drafting the WCT, which went into force domestically on March 6, 2002. The WCT serves as an extension to the rights established by the Berne Convention and was passed pursuant to Article 20 of the Convention.[132] Unlike Berne, the WCT actually makes specific reference to the remedy obligations of "Contracting Parties," mandating that the parties "shall ensure that enforcement procedures are

131. Although she does not necessarily argue that the registration requirement explicitly violates the requirements of Berne, Shira Perlmutter has flagged its philosophical incompatibility with Berne. *See* Shira Perlmutter, *Freeing Copyright from Formalities*, 13 CARDOZO ARTS & ENT. L.J. 565, 565–66, 575–76 (1995).

132. WIPO Copyright Treaty art. 1(1), Dec. 20, 1996, 36 I.L.M. 65, S. TREATY DOC. NO. 105-17 ("This Treaty is a special agreement within the meaning of Article 20 of the Berne Convention for the Protection of Literary and Artistic Works, as regards Contracting Parties that are countries of the Union established by that Convention.").

available under their law so as to permit *effective action* against any act of infringement of rights covered by this Treaty, including expeditious remedies to prevent infringements and remedies which constitute a *deterrent to further infringements*."[133] For creators who do not timely register, American law provides no punitive damages, statutory damages, or attorneys' fees. With only actual damages or disgorgement of profits left, there is no deterrent effect and plaintiffs are frequently unable to take effective legal action against infringers.

III. HIERARCHY AND REFORM

The American copyright registration system not only frustrates the ability of many creators to be made whole for even the most egregious infringements of their copyrights, but it also fails to fulfill its basic notice function and possibly violates our international treaty obligations. Nevertheless, efforts to eliminate § 412 have been met with steep resistance and therefore have not been enacted. While it is impossible to ascertain precisely why such proposed amendments were never enacted, the hearings surrounding this issue provide clues as to who opposed the elimination of § 412 and why they opposed this change.

As the following analysis reveals, by creating a two-tiered system of protection, the registration requirement constructs a hierarchy of works defined by their violability. Works by sophisticated creators have the opportunity to become part of the commercial canon. Their aura is secured through artificial scarcity perpetuated by copyright law and the dramatic penalties facing infringers for unauthorized exploitations of such works. Thus, sophisticated creators can dangle copyright's Sword of Damocles over the heads of would-be infringers. Almost any book, periodical, recording, movie, television show, or computer program distributed by a large press, magazine publisher, music label, film studio, broadcast network, or software developer enjoys similar protection, even though many such works may lack continued economic value.[134] The recent wave of high profile infringement suits involving peer-to-peer file sharing clarifies this point: the expansive remedies provided by the Copyright Act allow organizations such as the RIAA to hand individual defendants their heads on a platter with more fervor than Salomé's dance (to licensed music, of course).

133. *Id.* at art. 14(2).

134. Of course, this is only true so long as these works remain under copyright protection.

For an illustration of this concept, consider the case of Jammie Thomas-Rasset, a single mother of four who earned her living working as a natural-resources coordinator for a Native American tribe in Minnesota. In 2005, Thomas was sued by the RIAA for sharing 24 songs on the peer-to-peer file-sharing site Kazaa. Initially, the court found Thomas liable for willful copyright infringement in the amount of $222,000. While the $9250 per track judgment may seem high, it was far less than the $150,000-per-track statutory damages that courts are permitted to award plaintiffs, even in the absence of any proof of actual harm. Ultimately, however, Thomas earned a retrial when the verdict was thrown out based on an error in jury instructions.

Unfortunately for Thomas, things went even worse the second time around. In 2009, a jury returned another judgment against her, this time for the whopping amount of $1.92 million. At $80,000 per song, this is almost a full order of magnitude larger than the earlier verdict. The infringed songs were sold on iTunes at a price of ninety-nine cents each, arguably making the ratio between the verdict and the actual damages 80,000 to 1.[135] Ironically, in the context of punitive damages awards, the Supreme Court has ruled that any ratio in excess of ten to one violates the due process clause of the Constitution.[136]

On the other hand, non-registered works—generally those produced by unsophisticated creators such as individual artists—serve as fodder for remix, reinterpretation, transformation, and unauthorized use. These works lack any aura, their violability is not patrolled, and they may be infringed, sometimes even with impunity. As a result, the current system does not uniformly protect the interests of all authors so much as it privileges a certain class of works. The primary beneficiaries of this system are the major players in the copyright industry—the large corporations that are both generators and users of content. A system with more uniformly harsh consequences for infringement would be unfavorable to these players when they are on the receiving end of suits. This is especially so because the law imposes liability

135. *Cf.* Chris Williams, *Big Fine Could Be Big Trouble in Music Downloading Case*, Associated Press Newswire, June 19, 2009. The article noted that Tom Sydnor, director of the Progress & Freedom Foundation's Center for the Study of Digital Property, defended the verdict, arguing that "[l]egally acquiring a license to give copies of a song to potentially millions of Kazaa users might well have cost $80,000 per song." *Id.*

136. *See* BMW of N. Am. v. Gore, 517 U.S. 559, 581 (1996) (concluding that the relevant ratio for determining punitive damages as compared to compensatory damages is "not more than 10 to 1"); *see also* State Farm Mut. Auto. Ins. Co. v. Campbell, 538 U.S. 408, 425 (2003) ("Our jurisprudence and the principles it has now established demonstrate . . . few awards exceeding a single-digit ratio between punitive and compensatory damages, to a significant degree, will satisfy due process.").

on both direct and vicarious infringers, regardless of mens rea (though mens rea can affect the amount of damages in a case of timely registration).[137] A two-tiered system is ideal for these corporations. When they infringe the materials of others—even if they do so willfully—the consequences are relatively benign. Meanwhile, when an outsider infringes their work, the penalties are draconian. The strategic preference of certain sophisticated creators for this dichotomous structure becomes clear when one examines the failed efforts to eliminate § 412 in the early 1990s.

A. SECTION 412 REFORM AND ITS DISCONTENTS

In 1993, Representative William J. Hughes of New Jersey introduced legislation—dubbed the Copyright Reform Act—that, inter alia, repealed the registration requirement for standing, statutory damages, and attorneys' fees.[138] The House ultimately passed the bill on November 20, 1993 and again on September 20, 1994.[139] However, the bill died when the Senate failed to act.[140]

The available paper trail provides insight into the various interests that worked to prevent the bill from passing. A number of groups that represent libraries or smaller creators including the Graphic Artists Guild, the American Society of Media Photographers, the Software Publishers Association, the Committee for Library Property Studies, and the American Association of Law Libraries all spoke in favor of the bill.[141] Just three groups came to speak in opposition.[142] Yet the bill never passed. These three groups were, oddly enough, the Association of American Publishers (AAP), the

---

137. *See, e.g.*, NIMMER & NIMMER, *supra* note 29, § 13.01 (reflecting that state of mind is not a relevant element in making a prima facie case of copyright infringement).

138. Copyright Reform Act of 1993, H.R. 897, 103d Cong. (1993). Senator Dennis DeConcini introduced a related bill in the Senate. Copyright Reform Act of 1993, S. 373, 103d Cong. (1993). On November 17, 1993, Hughes's bill received a favorable report from the Subcommittee on Intellectual Property and Judicial Administration of the House Committee. *See* John B. Koegel, *Bamboozlement: The Repeal of Copyright Registration Incentives*, 13 CARDOZO ARTS & ENT. L.J. 529, 529 n.1 (1995); Copyright Reform Act of 1993, H.R Rep. 103-388 (1993).

139. H.R. 4307, 103d Cong. (1994).

140. Shira Perlmutter, *Freeing Copyright from Formalities*, 13 CARDOZO ARTS & ENT. L.J. 565, 572 (1995).

141. *Copyright Reform Act of 1993: Hearing on S. 373 Before the Subcomm. on Patents, Copyrights and Trademarks of the S. Comm. On the Judiciary*, 103d Cong. (1993) [hereinafter *Hearnings*] (witness list).

142. *Id.* Poet and novelist Erica Jong, a member of the Authors Guild, testified in support of the legislation, but she appeared only in her individual capacity because her opinion was not shared by the organization. *Id.* (statement of Erica Jong) ("I emphasize that I am here as an individual author and not in any official capacity.").

American Association of University Presses (AAUP) and the Authors Guild (AG). The AAP is, in its own words, "the principal trade association of the U.S. book publishing industry."[143] The AAUP is a trade association with more than 130 members worldwide, consisting of both non-profit academic and scholarly publishers.[144] Finally, despite its name, the AG neither represents most authors, nor speaks for them. Indeed, by their own description at the time, the Guild was made up of "6500 published writers—authors of fiction, history, biography, textbooks, periodical articles, short stories and other literary works—and includes winners of the Nobel Prize in Literature, the Pulitzer Prize and countless other literary awards."[145] Their membership numbers have grown since the date of the 1993 hearing,[146] but membership decisions continue to be made on a case-by-case basis upon application. At a minimum, book authors must have published their work with an established American publisher and received a "significant advance" in order to receive consideration.[147] The AG therefore constitutes an elite group of only the most successful commercial authors—individuals who have a strong interest in the inviolability of their works. Not surprisingly, their spokesperson at the hearing on the Copyright Reform Act of 1993 was novelist Scott Turow, an author who has profited handsomely from adaptations of his works in a number of contexts.[148]

At first blush, one would think that groups representing the interests of publishers and authors would appreciate the ability to vindicate their rights with fewer formalities. However, such groups already register their works and therefore already have the opportunity to vindicate their rights to the fullest extent possible under the law. Yet when they are on the receiving end of a lawsuit, they would also prefer that their opponents not have the full panoply of remedies available. After all, many creators make use of other copyrighted

143. Association of American Publishers, http://www.publishers.org/ (last visted Feb. 13, 2010). According to their "Membership" page, approximately 260 publishers belong to the AAP.

144. Association of American University Publishers, http://aaupnet.org/membership/directory.html (last visited Feb. 21, 2010).

145. *Hearings*, *supra* note 141, (statement of Scott Turow) ("appearing . . . in [sic] behalf of The Author's Guild, Inc.").

146. There are currently over 8000 authors in the Authors Guild. The Authors Guild, History, http://www.authorsguild.org/about/history.html (last visited Feb. 13, 2010).

147. The Authors Guild, *Membership Eligibility*, https://www.authorsguild.org/join/eligibility.html (last visted Feb. 13, 2010).

148. *See, e.g.*, PRESUMED INNOCENT (Warner Bros. Pictures 1990) (theatrical film version of Turow's novel of the same name); THE BURDEN OF PROOF (1992) (television miniseries version of Turow's novel of the same name); REVERSIBLE ERRORS (2004) (television mini-series version of Turow's novel of the same name).

works. With the current registration regime, sophisticated interests get the best of both worlds: the full range of remedies when they are plaintiffs seeking vindication of their intellectual property interests, but incredibly narrow remedies when they are defendants, thereby helping them to fend off suits for infringement.

Strangely, the forces opposing § 412's repeal dismissed the possibility that the existing system might frustrate the ability of copyright holders to vindicate their intellectual property rights. After what it claimed to be "an elaborate process of consultation with its own members and representatives of other writers groups," the Authors Guild concluded that

> [o]ur efforts to find an example of a meritorious claim by a writer that was lost or seriously frustrated under the present system was unsuccessful. Undoubtedly, there must be such cases; but our diligent efforts to study the issue empirically suggest that instances where the lack of statutory damages have prevented writers from bringing infringement claims are far less widespread than imagined and that the currently available remedies appear to be accomplishing their intended effect.[149]

Unfortunately, it seems that the Guild's researchers were not looking particularly hard, as such a claim appears thoroughly disingenuous in light of our prior discussion. Indeed, a simple examination of case law reveals numerous cases demonstrating just this point.

Take *Deltak, Inc. v. Advanced Systems, Inc.*,[150] a case decided a decade before the debate over the Copyright Reform Act of 1993 took place. The litigation involved Deltak's claims of literary infringement, but not of the highbrow variety. Rather, copy from a corporate pamphlet describing programs that taught data processing skills were lifted wholesale by a rival company, Advanced Systems, for use in its brochures. In the suit, Judge Richard Posner, who was sitting as by designation, found that the defendant had not only infringed plaintiff's copyright, but had done so willfully. On this basis, he noted that

> [i]f Deltak had registered its copyright within the time provided by the Copyright Act, I would have no hesitation in awarding not only the maximum statutory damages [at the time] under section 504(c)(2) of $50,000, but also attorney's fees, which are authorized

149. *Hearings*, *supra* note 141, (statement of Scott Turow).

150. Deltak, Inc. v. Advanced Systems, Inc., 574 F. Supp. 400 (N.D. Ill. 1983).

> by section 505 and are frequently awarded in cases of willful infringement even if no actual damages are proved.[151]

Not surprisingly, however, the plaintiff had not timely registered its copyright, and was unable to establish damages with sufficient certainty.[152] The defendant's plan to use Deltak's brochures for its own marketing purposes ultimately failed, and therefore there were no proven profits to disgorge or actual sales lost by the plaintiff.[153] As a result, despite the willful infringement of its works, Posner felt it was left with no alternative to but to award the plaintiff *nothing*, despite the verdict in its favor.[154] As the Second Circuit lamented when dealing with the same issue in a different case, the existing § 412 structure can lead to "the anomaly of affording plaintiffs a right without a remedy."[155] These cases represent only the tip of the iceberg because the dictates of § 412 render such infringements—no matter how unabashed—impractical to litigate.

Conversely, opponents of the reform efforts claimed that the elimination of § 412 would undermine the legitimate fair use of copyrighted works by spurring frivolous and vexatious litigation by rapacious rights holders. As the AAP and AAUP stated,

> We oppose repeal because it would upset the careful and critical balance struck by [§]412 among the interests of authors and publishers of pre-existing works and those who would transform, build upon and make reasonable use of those works. Repeal would discourage legitimate and important activities of historians, biographers, journalists, and other authors and publishers.[156]

Taken at face value, and viewed narrowly, the sentiments reflected in this testimony seem to make eminent sense. However, when one considers the positions taken by the AAP and AAUP in litigation and its public representations and one assesses the situation in the broader context of rights management, things appear quite different.

---

151. *Id.* at 402.

152. *Id.* at 411.

153. *Id.* at 403–04, 411–12.

154. *Id.* at 412. Ultimately, on appeal, the Seventh Circuit carved out a relatively novel exception by allowing recovery of a reasonable license fee in such situations. The court therefore remanded the case for recalculation of damages. However, in other circuits, such as the Second, the possibility of such recovery has been clearly disavowed, regardless of the harsh and seemingly inequitable consequences. *See* Business Trends Analysts, Inc. v. Freedonia Group, Inc., 887 F.2d 399, 406 (2d Cir. 1989).

155. *Id.* at 406 (citing 3 NIMMER & NIMMER, *supra* note 29, § 14.02[A]).

156. *Hearings*, *supra* note 141, (statements of the American Association of Publishers and the American Association of University Presses).

Unless we make the dubious assumption that the AAP and AAUP are somehow more socially responsible and altruistic than other copyright holders, the arguments of the AAP and AAUP would militate against the availability of statutory damages and fees in *all* infringement suits, lest they encourage frivolous litigation. After all, many of the works to which the AAP and AAUP claim copyright are works of which historians, biographers, journalists, and other authors and publishers would like to make transformative, accretive, or reasonable use. Nevertheless, the AAP and AAUP were clearly not willing to take such a stand. After all, by timely registering their works, they enjoy dramatic benefits when, as rights holders, they seek to assert their copyrights and pursue alleged infringers with the threat of statutory damages and attorneys' fees. Indeed, their palaver regarding legitimate educational activities is readily betrayed by the fact that they have no qualms about opposing the unauthorized use of *their* copyrighted works as primary materials for other historians, biographers, journalists, and researchers in many contexts. As it turns out, the AAP and AAUP have repeatedly asserted that *any* unauthorized use of their works constitutes an act of infringement and they have demonstrated a willingness to sue on the basis of this principle.

For example, in just the past few years, the major academic publishers have filed dozens of lawsuits across the country against reproduction shops that produce course "readers" used on college campuses on the theory that such packets violate the publishers' copyrights.[157] Although the makers of the course readers undoubtedly profit in their provision of these services, the packets directly serve an educational purpose. Section 107 of the Copyright Act explicitly states, "the fair use of a copyrighted work, . . . for purposes such as . . . teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright."[158] Section 107's language may be merely preambulary, or it may create a bright-line rule protecting the use of copyrighted materials for teaching and research

157. *See, e.g.*, Princeton Univ. Press v. Michigan Document Servs., Inc., 99 F.3d 1381, 1400 (6th Cir. 1996); Basic Books, Inc. v. Kinko's Graphics Corp., 758 F. Supp. 1522, 1532 (S.D.N.Y. 1991) (holding that while the particular course packets at issue were copied for educational purposes, they did not qualify as fair use because, inter alia, they were made for profit).

158. 17 U.S.C. § 107 (2006). It should be noted that, in spite of this language, courts have still managed to find a plethora of instances where use of a copyrighted work for teaching, research, or scholarship constitutes infringement. *See, e.g.*, *Princeton Univ. Press*, 99 F.3d at 1391; Am. Geophysical Union v. Texaco, 37 F.3d 881, 899 (2d Cir. 1994); Television Digest, Inc. v. U.S. Tel. Ass'n, 841 F. Supp. 5, 11 (D.D.C. 1993); *Basic Books, Inc.*, 758 F. Supp. at 1547.

purposes. But either way, for the concept of fair use to have any meaning, there must be some threshold at which the use of a copyrighted work for such purposes is excused—whether it is the quotation of a single sentence or the unauthorized reproduction of many pages. Nevertheless, the major publishers serving the academic community have vigilantly maintained that any use, no matter how small, by the copy shops requires their authorization and payment through the Copyright Clearance Center—a centralized clearing house for published content.[159] And, in their public statements, it appears that this position is not simply limited to copy-shops but extends to any unauthorized use of their copyright works. In language that has grown almost de rigueur in the industry, one leading academic publisher warns: "No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, on in any information storage or retrieval system, without permission in writing from publishers."[160]

More broadly, an examination of copyright stances taken by some sophisticated creators quickly undermines any assumption that they are less likely to engage in extreme and aggressive positions vis-à-vis their purported intellectual property rights. As Jason Mazzone points out, major publishers often claim copyright protection over works that are indisputably in the public domain. Many of these works constitute important primary source materials for historians and others. In a quintessential example of overbearing copyright claims, Mazzone observes that a pocket version of the United States Constitution sold on the market sternly warns anyone against reproducing the work without written permission from the publisher. As Mazzone quips, "Whatever the Constitution's framers and ratifiers had in mind when they authorized Congress to create copyright law, they surely did not expect that somebody would one day claim a copyright in the Constitution itself."[161]

All told, there is no reason to think that sophisticated creators, such as the AAP, AAUP, or the AG are any more altruistic or socially responsible with the enforcement of their copyrights than any other rights holder. As such, drawing a line on the availability of statutory damages and attorneys'

159. Founded in 1978 by a group of publishers and authors, the CCC is a clearinghouse for the licensing of "millions of books, journals, newspapers, websites, ebooks, images, blogs and more." Copyright Clearance Center, About Us, http://copyright.com/viewPage.do?pageCode=au1-n (last visited Nov. 25, 2009).

160. See, for example, the copyright insert for STEVE NEALE, GENRE AND HOLLYWOOD (2000).

161. Jason Mazzone, *Copyfraud*, 81 N.Y.U. L. Rev. 1026, 1028 (2006).

fees with the timely registration requirement in order to prevent a tide of frivolous litigation makes little sense. Moreover, by comparing the purported reasoning of such groups as the AAUP and APA in opposing § 412 reform with their litigation agendas and fair use policies, it becomes reasonable to at least suspect that § 412 serves sophisticated creators quite well by granting them expansive rights to use the works of unsophisticated creators without authorization while simultaneously enabling them to enforce their own copyrights with tenacity and severity.

B. HIERARCHY IN HOLLYWOOD

In light of our foregoing discussions, it is natural to ask how artists could fail to register their works. After all, artists dedicate countless years and make many sacrifices to bring their work to fruition. It may seem surprising that they do not take that extra step to obtain full legal protection for their works. Part of the problem is that many artists are not aware of the importance of registration. Others do not want to be bothered with paperwork and its apparent complexities. However, it is not simply a matter of ignorance or myopia. For many artists, effective and timely registration is cost-prohibitive. For example, for artists such as photographers who create a large volume of works, only a small number of which may attain high value, registration can be a costly affair.[162] In recent years, the Copyright Office has adopted regulations that allow group registration of some works, but these regulations are highly restrictive.[163] Moreover, many artists are effectively dissuaded from registration by relying on private registration regimes—regimes that grant them some of the benefits of going through the Copyright Office, but not the recovery of statutory damages and attorneys' fees. As we shall see, screenwriters are such a group. At the same time, the field of screenwriting also provides a salient illustration of the modern two-tiered protection system, its perpetuation of hierarchy and the powers that it serves.

---

162. *See* Ossola, *supra* note 50, at 560 (1995). Ossola states,

> Photographers present special problems, given the tremendous volume of works created, but these problems are not unique to them. With thousands of images created each year, it is literally impossible for anybody, even the most successful photographers, to register those images in the Copyright Office. As a result, even the most successful photographers consistently fail to register their works.

Id. In their analysis of copyright registrations and renewals from 1910 through 2000, William Landes and Richard Posner determined that even small fee increases can result in precipitous declines in registrations and renewals. WILLIAM M. LANDES & RICHARD POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 245 (2003).

163. 17 U.S.C. § 408 (2006) (providing for group registration of certain copyrighted works).

In Hollywood, it is no secret that screenwriters often feel unappreciated and disenfranchised. Directors are, after all, viewed as the CEOs of movies—their names drive critical discussions about oeuvres and masterpieces. Through the "a film by" credit, directors are designated by fiat as a film's auteur. Similarly, actors serve as the industry's public face and have always represented a key engine of its financial success—at least until recently.[164] And, while it is true that a movie cannot get made without a script, the overriding sentiment towards screenwriters is perhaps best captured by one studio mogul's famous musing: "If we could only figure out a way to make movies without writers."[165] In the early days of the industry, some of our finest novelists sought to pay their bills by trying their luck in Tinsel Town. The products of these ill-advised ventures by William Faulkner, Nathanael West, James Agee, Ernest Hemingway, and others are notorious.[166] As critic Edmund Wilson would later observe, the failures of Fitzgerald and West "may certainly be laid partly to Hollywood, with its already appalling record of talent depraved and wasted."[167] The writers often wound up defeated and desolate. To Wilson, however, the result was not surprising: Hollywood was "an intractable magnetic mountain, which twists American fiction askew."[168]

There are, of course, many factors that allow such "twisting" to take place: commercial realities, the multidimensional nature of the movie-making process, bargaining power disparities, the history of the industry, and the roles of the various guilds representing above-the-line talent such as directors, actors, and writers. However, the "twisting" is also aided by a de facto norm in the industry that is not usually analyzed: the absence of strong copyright protection for scripts.

In Hollywood, paranoia over the unauthorized usurpation of the heart of one's screenplay or treatment runs rampant. And the reason is simple: it happens. Screenwriters have responded, but not with widespread registration of their works with the Copyright Office. Instead, since 1927, the Writers Guild of America, West (WGAW) has administered a registration system for works that is convenient, easy to use, and relatively inexpensive.

---

164. *See, e.g.*, Willa Paskin, *WHO KILLED THE MOVIE STAR?: Hollywood's A-list Idols Are Losing Their Movie-Selling Mojo*, RADAR MAGAZINE, (July/August 2008).

165. *See* Andrew McWhirter, *Film: Fameless Faces – Hidden Art of Screenwriting Revealed*, TRIBUNE MAGAZINE, Jan. 19, 2009, *available at* http://www.tribunemagazine.co.uk/2009/01/19/film-fameless-faces-%E2%80%93%C2%A0hidden-art-of-screenwriting-revealed/.

166. Edmund Wilson, *The Boys in the Back Room*, *in* CLASSICS AND COMMERCIALS: A LITERARY CHRONICLE OF THE FORTIES 19, 56 (1950).

167. *Id.*

168. HUBERT BUTLER, INDEPENDENT SPIRIT: ESSAYS 271, 272 (2000).

Unfortunately, it is also largely useless when utilized in lieu of a copyright registration, as it frequently is.

Billed as "the world's number one screenplay and intellectual property registration service,"[169] and the "the industry standard in the creation of legal evidence for the protection of writers and their work,"[170] the WGAW Registry is not entirely without utility. Individuals—both the general public and WGAW members alike—can deposit copies of their works, including screenplays and treatments prepared for radio, film, television, video, interactive media, and other works such as theatrical plays, novels, short stories, poems, commercials, lyrics, drawings, and music, with the WGAW. This helps to establish date of creation by producing a record of a screenplay or treatment being held in deposit by the Guild. This record can be useful should charges of plagiarism, misappropriation, or copyright infringement later emerge.

However, there is no good reason to opt for the WGAW's registration system over that of the Copyright Office. Unfortunately, however, many individuals, both inside and outside of the industry, do. The reason is not surprising. In a blurb buried within its Frequently Asked Questions section, the WGAW website does disclaim that "[r]egistering your work with the WGAW Registry does not take the place of registering with the Library of Congress, U.S. Copyright Office." But the Guild obfuscates the registration issue by eschewing explanation of the dramatic consequences of failing to register a work with the Copyright Office.[171] Moreover, its website states that WGAW registration and Copyright Office registration "both create valid legal evidence that can be used in court," thereby promoting a deceptive sense of interchangeability between the two regimes.[172]

But the two forms of registration are far from equal. Any work that is capable of being registered with the WGAW is, by its very nature, capable of being registered or pre-registered with the U.S. Copyright Office, since it is an original work of authorship fixed in a tangible medium.[173] Yet the WGAW registration system fails to provide several of the key advantages of

169. Writers Guild of America, West, Registry, http://www.wga.org/subpage_register.aspx?id=1183 (last visited Sep. 29, 2009).

170. WGA West: Registry, http://www.wgawregistry.org/ (last visited Sep. 29, 2009).

171. WGA West: Registry, Frequently Asked Questions, Does Registration Take the Place of Copyright?, http://www.wgawregistry.org/webrss/regfaqs.html#quest14 (last visited Sept. 29, 2009).

172. *Id.*

173. The copyright protection that a treatment or story outline receives may be thin, depending on the nature of the content and considerations such as the idea/expression dichotomy and the scènes à faire doctrine.

registration with the Copyright Office—specifically, the presumption of copyright validity and, most importantly, qualification for statutory damages and attorneys' fees in the event of an infringement suit. Although WGAW registration is less expensive than Copyright Office registration, it is only marginally so.[174]

The result of a world dominated by WGAW registrations is dramatic. The movie studios register their films with the Copyright Office. So infringement of those works is subject to harsh penalties. By contrast, the underlying screenplays, treatments, and outlines are usually registered only with the WGAW. As a result, the screenplay becomes a low-tiered work in the copyright schema, subject to manipulation, reinterpretation, transformation, and even unauthorized exploitation to a degree. It is not a sacred text; its inviolability is not ensured by law. Normatively, some will conclude that this is exactly as it should be. Others will be appalled. Either way, however, descriptively, the system constructs a hierarchy of works. In the end, the screenplay is malleable, submissive, and yielding; the movie is untouchable, consecrated, and unassailable. Formality serves function in determining Hollywood's chain of command.

C. CONSECRATION, CRITICAL THEORY, AND MUSIC

The history of the modern music industry also reveals the stratification process emerging from formalities in action. As K.J. Greene has observed, while ostensibly neutral, the technicalities of our copyright regime undoubtedly exist in a "concrete social milieu"[175] where "not all creators of intellectual property are similarly situated."[176] Often times, the privileging of certain works takes on the qualities of other social stratifications that divide along lines of class, gender, and race. As Greene argues, inequalities in bargaining power, a fundamental tension between structural components of copyright law and "the oral predicate of Black culture," and discrimination that resulted in the devaluation of Black creative contributions have resulted in the historical disadvantaging of African-American creators, especially African-American musicians.[177] A prominent example is the growth of the modern music industry. Driven by rock 'n roll, the industry saw much of its early success from the unauthorized exploitation of old blues riffs, many

174. Online WGAW Registration costs $10 for WGA members in good standing and $20 for the general public. Online registration of a copyright with the Copyright Office currently costs $35.

175. K.J. Greene, *Copyright, Culture & Black Music: A Legacy of Unequal Protection*, 21 HASTINGS COMM. & ENT. L.J. 339, 358–59 (1999).

176. *Id.* at 343.

177. *Id.* at 356–57.

stolen directly from unacknowledged and uncompensated African-American folk artists.[178] Although the resulting musical compositions and sound recordings represent the product of unsanctioned pastiche, the industry continues to vigilantly protect them—not just from wholesale reproduction, as the flood of file sharing suits demonstrates, but from transformative remixing, as the case law on sampling illustrates.

Although sound recordings generally did not receive federal copyright protection until 1972,[179] musical compositions have qualified for protection since 1831.[180] However, in order for a musical composition to receive a copyright, it had to be fixed in a tangible medium—in other words, it had to be written. This situation resulted in what Keith Aoki terms a "dual economy" of music. Under this system, certain kinds of music (and, therefore, composers) received legal protection, and certain kinds of music (and composers) did not. In general, copyrighted (or copyrightable), notated, written scores were composed by upper middle class educated whites, while un-notated musical compositions, including those created by or within folk collectives, did not receive copyright protection. Many unprotected works were intertemporal, intergenerational, anonymous, communal, or improvisational in their composition. Thus, in general, those works that arose within collective experiences of slavery, the struggle for freedom, and post-Reconstruction subordination did not receive protection.[181] Indeed, the strictures of our modern copyright regime, with its mystification and fetishization of the Romantic author trope, have often privileged Western forms of (ostensibly) individualistic creation over other modalities.[182]

---

178. *See* SIVA VAIDHYANATHAN, COPYRIGHTS AND COPYWRONGS: THE RISE OF INTELLECTUAL PROPERTY AND HOW IT THREATENS CREATIVITY 117–48 (2001) (tracing the appropriation of blues by rock 'n roll artists over time); K.J. Greene, *Intellectual Property at the Intersection of Race and Gender: Lady Sings the Blues*, 16 AM. U. J. GENDER, SOC. POL'Y & L. 365, 371–74 (2008); K.J. Greene, *"Copynorms," Black Cultural Production, and the Debate Over African-American Reparations*, 25 CARDOZO ARTS & ENT. L.J. 1179, 1193 (2008) ("The fleecing of Black artists was the basis of the success of the American music industry."). One example where rights were reasserted several decades later occurred when the Estate of Willie Dixon successfully went after Led Zeppelin for their uncredited, unauthorized, and uncompensated lifting of *You Need Love* for their song *Whole Lotta Love*. See Keith Aoki, *Distributive Justice and Intellectual Property: Distributive and Syncretic Motives in Intellectual Property Law*, 40 U.C. DAVIS L. REV. 717, 763 (2007).

179. 17 U.S.C. § 102(a)(7) (2006) (placing sound recordings within the subject matter of copyright protection); 1 NIMMER & NIMMER, *supra* note 29, § 2.10 (noting that copyright protection was extended to sound recordings fixed on or after Feb. 15, 1972).

180. General Revision of Copyright Act of Feb. 3, 1831, ch. 16, § 1, 4 Stat. 436.

181. Aoki, *supra* note 178, at 760.

182. *See, e.g.*, Peter Jaszi, *Contemporary Copyright and Collective Creativity*, *in* THE CONSTRUCTION OF AUTHORSHIP: TEXTUAL APPROPRIATION IN LAW AND LITERATURE 29

In recent decades, the widespread dissemination of recording and publishing technology has ensured that virtually all creative works are fixed in a tangible medium, thereby avoiding one iteration of the "dual economy" problem identified by Aoki. However, the timely registration requirement has stepped in to enforce a hierarchy of works by distinguishing between sophisticated and unsophisticated creators. Creative works by those at the legal and social margins remain unregistered and therefore unprotected. These intellectual properties become low-tier works, relegated to the status of raw materials subject to remixing, reinterpretation, and transformation. But from these low-tier works come the inviolable commercial products whose iterations are carefully controlled, and whose scarcity is assiduously patrolled. Registered and enforced by the RIAA, the commercial product represents a sacred work that cannot be manipulated without authorization of its rights holders. Thus, the potential socioeconomic and racial dimensions to the cultural hierarchy of copyrighted works live on through the timely registration requirement.

IV. CAVEATS AND CONCLUSIONS

As it turns out, American copyright militancy is vastly overstated. Indeed, by comparison to most developed countries, we continue to provide inadequate remedies to a large class of content creators: authors who do not timely register their copyrights. Instead, we practice a two-tiered protection system that privileges sophisticated creators. Their works become sacred, inviolable matter protected from unauthorized exploitation or transformation by a series of remedies that often rise to draconian levels. Meanwhile, the works of unsophisticated creators remain fodder for remix and reinterpretation. Thus, the registration system plays a critical role in perpetuating a sacralization process. While the emergence of mass reproduction and digital dissemination has threatened the consecration of privileged works, our registration regime has rekindled the aura. What technology has undermined, our two-tiered copyright hierarchy has reinstated, at least in part.

Indeed, a close examination of the language and current interpretation of the Copyright Act reveals several difficulties facing unsophisticated creators

---

(Martha Woodmansee & Peter Jaszi eds., 1994); Olufunmilayo B. Arewa, *From J.C. Bach to Hip Hop: Musical Borrowing, Copyright and Cultural Context*, 84 N.C. L. REV. 547, 550–51 (2006) ("Copyright legal structures and the classical music canon have thus relied on a common vision of musical authorship that embeds Romantic author assumptions. Such assumptions are based on a vision of musical production as autonomous, independent and in some cases even reflecting genius.").

who seek to vindicate their rights in the United States. First, courts have found that § 412 precludes recovery of statutory damages and attorneys' fees, when an infringement continues after registration. As a result, an infringer can continue its wrongful activity with impunity if a work is not timely registered at the time of first infringement. Second, unsophisticated creators face a one-way risk of attorneys' fees. Plaintiffs who fail to register their copyright on a timely basis are never eligible to recover their fees if they prevail in a suit, even if the infringement was willful.[183] By contrast, defendants are always eligible to recover their fees if they prevail. Third, creators cannot seek punitive or reputational damages, and actual damages are often insufficient to make them whole or to deter future infringement. All told, these factors combine to create a rather bleak enforcement regime for the rights of creators who do not register in a timely manner.

Of course, the idea that access to legal counsel and adherence to certain legal formalities can improve the effective scope of one's rights is certainly not novel or surprising. But, the consequences in copyright law are particularly dramatic, virtually determining the rights to and in cultural production. Sophisticated, economically powerful interests receive full protection for their creative works, making their cultural production sacred and inviolable. The Bourdieuian act of cultural reproduction[184] is therefore controlled and patrolled by copyright law—with the hallowed works of elites subject to use and re-use only with proper authorization and payment.

---

183. 17 U.S.C. § 412 (2006). The section states,

> In any action under this title, other than an action brought for a violation of the rights of the author under section 106A(a), an action for infringement of the copyright of a work that has been preregistered under section 408(f) before the commencement of the infringement and that has an effective date of registration not later than the earlier of 3 months after the first publication of the work or 1 month after the copyright owner has learned of the infringement, or an action instituted under section 411(b), *no award of statutory damages or of attorney's fees*, as provided by sections 504 and 505, shall be made for—
>
> (1) any infringement of copyright in an unpublished work commenced before the effective date of its registration; or
>
> (2) any infringement of copyright commenced after first publication of the work and before the effective date of its registration, unless such registration is made within three months after the first publication of the work.

Id. (emphasis added).

184. *See* PIERRE BOURDIEU & JEAN-CLAUDE PASSERON, REPRODUCTION IN EDUCATION, SOCIETY, AND CULTURE (Richard Nice trans. 1977) (using the concept of cultural reproduction to explain the hegemonic process through which the dominant class retains its power).

Meanwhile, the output of the rest of society does not receive such beatification. For unsophisticated players, their production is subject to remix, reinterpretation, and re-commercialization, all without authorization or payment. Thus, while the law purports to grant copyright protection to any work of authorship with minimal creativity fixed in a tangible medium irrespective of whether it was made by Manet or "the Man on the Street," in practice, such is not the case. All works and creators are not treated alike, and the formalities of the registration requirement establish a hierarchy of protected and less protected works, the untouchable and the readily manipulable. The resulting system enables dominant social forces to freely usurp the creative content of the masses for their own use while simultaneously enjoying the ability to prevent any unauthorized use of their own privileged creative content. Within the confines of this regime, it is the underclass that typically ends up with minimal protection.

It is the African-American blues musician, who just two generations ago saw his riffs and melodies appropriated by the burgeoning modern music industry to "develop" rock 'n' roll. As his creative efforts developed a multi-billion dollar marketplace, he watched from the economic sidelines, unable to vindicate the legal rights to his intellectual property because he had failed to conform to the procedural strictures of copyright law.

It is the unheralded rural landscape painter whose evocative depictions of nature are used to decorate and set the mood for the outdoorsy, Western-themed catalog of a major retailer. Although her work is used without permission or payment, her failure to register on a timely basis leaves her without meaningful remedies if she threatens to pursue legal action.

It is the screenwriter—the least valued of Hollywood's traditional above-the-line creative triumvirate[185]—who followed industry protocol by registering his screenplay with the Writers Guild of America and then found original dialogue and a unique action sequence from his work in the summer's leading blockbuster. Absent a well-timed and successful application for injunctive relief and posting of the necessary bond, he will find himself without meaningful remedies to pursue legal action simply because registration with the WGA fails to provide the legal benefits of registration with the United States Copyright Office.[186]

185. Actors, directors, and writers. One could also add producers to this list.

186. The Writer's Guild Registry does not "bestow any statutory protections." *See* WGA West Registry, Registration Details, http://www.wgawregistry.org/webrss/reg details.html (last visitied Nov. 8, 2009).

It is the dance choreographer whose uniquely sequenced moves make their way into the new music video for a leading pop star. Having failed to register her choreography on a timely basis with the United States Copyright Office, she is without an ability to recover statutory damages and attorneys' fees, making legal action of dubious worth.

It is the small businesswoman who drafts effective copy for a marketing pamphlet, only to see it copied wholesale by a competitor. In the absence of timely registration, legal action makes no practical sense. She finds herself holding "a right without a remedy."[187]

It is the solo architect whose structural design plans for an industrial building integrating green photovoltaic technology are taken and utilized by a multinational corporation to save itself the cost of generating its own plans. In the unlikely event that the architect catches the infringement and sues, she will, at best, likely recover only the value of the plans in the first place. Facing no punitive consequences for its illicit conduct, the multinational corporation has every incentive to infringe. Absent timely registration, the architect is without meaningful remedies, especially when one considers the (non-recoupable) cost of litigation.

It is the graffiti artists in urban corridors whose renderings eventually make their way into the newly sacrililized work of the modern art world's latest sensation, whose multi-million dollar originals and 'affordable' limited edition prints are carefully controlled and regulated, in order to obtain the Benjaminian aura that generates value in the art marketplace.

All the while, back-breaking penalties await those who would touch the copyrighted works of the modern music industry, the major Hollywood studios, the elite art world, or the fashion industry without permission, even when these works are built on the unprotected works of others. One might even posit that by controlling the manipulation and transformation of cultural content through its hierarchical system of protection, copyright law's registration requirement plays a significant role in personhood development and identity formation as it determines the ways in which we can and cannot interact with the cultural content.

Although this Article scrutinizes the peculiar epistemological role of copyright law's registration requirement in controlling the manipulation and transformation of cultural content by creating and patrolling a hierarchical system of protection, it does not necessarily advocate immediate repeal of § 412. First, the registration requirement advances the ability of individuals

187. Bus. Trends Analysts, Inc. v. Freedonia Group, Inc., 887 F.2d 399, 406 (2d Cir. 1989) (citing 3 NIMMER & NIMMER, *supra* note 29, § 14.02[A]).

and corporations to engage in activity akin to efficient breaches of contracts—at least in theory. However, to engage in an efficient infringement, one must be aware of one's rights and potential liabilities ex ante. Unfortunately, as we have seen, the copyright system does not enable this since there is no good way to know if a work is registered and therefore entitled to enhanced protection. One can know when a work is registered, but it is difficult to conclude with assurance that a work is *not* registered. Thus, the value of making efficient breaches available through the registration scheme is heavily dissipated.

Additionally, the elimination of the timely registration requirement for statutory damages and fees eligibility could certainly lead to some problems, especially if accomplished without other significant changes to our infringement and remedies regime. As Jon Baumgarten and Peter Jaszi have argued, repealing of § 412 might indeed increase potentially frivolous copyright litigation.[188] Such a claim is not to be taken lightly, especially in an era characterized by overreaching copyright claims and a disparity between copyright norms and laws that has left us all vulnerable to infringement litigation for dozens of our quotidian acts.[189] Indeed, as a critic of over-expansive copyright claims and the threat that copyright enforcement run amuck can inflict on our daily lives, I am keenly aware of the benefits that accrue to society from many works not being registered and, therefore, remaining ineligible for attorneys' fees and statutory damages. By making infringement litigation more profitable for more rights holders, we certainly risk an uptick in litigation by giving even plaintiffs with petty claims more leverage.

However, there are at least some reasons to think that a tide of frivolous litigation may not be unleashed by reform. First, such a position presumes that the existing holders of copyrights that are timely registered are less likely to pursue frivolous litigation than the masses. As we have already discussed, this assumption is relatively untenable. Second, we have direct experience suggesting otherwise. Specifically, under the pre-1976 regime, fees and statutory damages were available to all, regardless of registration. Furthermore, a registration requirement for standing to bring a suit was largely eviscerated by the Supreme Court's decision in *Washingtonian Publishing*

188. *See* Jon Baumgarten & Peter Jaszi, *Why Section 412 Should Be Retained*, *reprinted in* ROBERT WEDGEWORTH & BARBARA RINGER, THE LIBRARY OF CONGRESS, ADVISORY COMMITTEE ON COPYRIGHT REGISTRATION AND DEPOSIT, REPORT OF THE CO-CHAIRS A85-A91 (1993) [hereinafter ACCORD REPORT]; Koegel, *supra* note 138; Peter Jaszi, *Section 412*, reprinted in ACCORD REPORT A92-A93 (1993).

189. *See* John Tehranian, *Infringement Nation*, 2007 UTAH L. REV. 537 (2007).

Co. v. Pearson,[190] in 1939. Despite this lack of a registration requirement, no boom in litigation resulted. Moreover, no other country has a registration requirement. Yet we have not witnessed a flood of frivolous copyright litigation either pre-1976 or in other countries, especially those sharing common legal traditions.

Nevertheless, given a risk of increased litigation, it is important that no change to § 412 occur in a vacuum. Concomitant reform in several areas is needed.[191] For example, broader general protection of transformative rights (for both sacred and low-tier works), including the implementation of some bright-line rules, would reduce the likelihood of some socially undesirable litigation.[192] Moreover, a limitation on the recovery of statutory damages would conform the remedies regime to constitutional due process dictates that require punitive assessments to bear some reasonable relationship to actual damages.[193] Indeed, the disproportional size of statutory damages, decoupled from any proof of actual damages, may do far more to encourage frivolous litigation than putting all creators—sophisticated and unsophisticated—on a level playing field for remedies. As a result, it may make sense to think about simultaneously reducing the upper range of statutory damage awards and providing better protections for innocent infringers, while expanding the availability of statutory damages to all copyrighted holders, regardless of formalities. We should simultaneously consider improving the powers that defendants have in fighting meritless infringement claims by considering greater penalties for overreaching copyright claims[194] and making attorneys' fees recovery for prevailing defendants easier to attain.[195]

---

190. 306 U.S. 30 (1939), *reh'g denied,* 306 U.S. 668 (1939).

191. *See, e.g.*, Pamela Samuelson, *Preliminary Thoughts on a Copyright Reform Project*, 2007 UTAH L. REV. 551 (making the case for preliminary consideration of a holistic reform of copyright law).

192. *See* John Tehranian, *Whither Copyright? Transformative Use, Free Speech, and an Intermediate Liability Proposal*, 2005 B.Y.U. L. REV. 1201 (2005).

193. *See* Pamela Samuelson & Tara Wheatland, *Statutory Damages in Copyright Law: A Remedy in Need of Reform*, 51 WM. & MARY L. REV. 439 (2009) (arguing that the present statutory damages scheme is both inconsistent with congressional intent and in violation of the Supreme Court's due process jurisprudence regarding punitive damages awards).

194. *See, e.g.*, Jason Mazzone, *Copyfraud*, 81 N.Y.U. L. REV. 1026 (2006) (arguing that Congress should amend the Copyright Act to create a cause of action against overreaching claims of copyrightability by purported rights holders).

195. Prevailing parties can, at the court's discretion, receive attorneys' fees. 17 U.S.C. § 505 (2006). However, despite the absence of any statutory language distinguishing between prevailing plaintiffs and defendants, courts have historically applied a bifurcated analysis on the fees question, depending on which side prevailed.

Prior to 1994, many courts—including the copyright-rich Second and Ninth

Thus, any reform of the copyright system that seeks to treat content creators of all stripes on a more equal basis must remain cognizant of the critical need to balance the rights of creators with those of users. In the end, we should not necessarily seek more copyright or less copyright, but, rather, *better* copyright.

Circuits—explicitly adopted an unbalanced approach to the award of fees under § 505, and courts would routinely grant them to prevailing plaintiffs, see, e.g., Frank Music Corp. v. Metro-Goldwyn-Mayer, Inc., 886 F.2d 1545, 1556 (9th Cir. 1989) ("Plaintiffs in copyright actions may be awarded attorney's fees simply by virtue of prevailing in the action: no other precondition need be met, although the fee awarded must be reasonable."), but deny fees to prevailing defendants absent a finding of frivolousness or bad faith—an exacting standard. *See, e.g.*, Warner Bros. Inc. v. Dae Rim Trading, Inc., 877 F.2d 1120 (2d Cir. 1989); Olson v. N.B.C., Inc., 855 F.2d 1446, 1454 (9th Cir. 1988) (declining to adopt an evenhanded approach to awarding fees). *But see* Lieb v. Topstone Indus., Inc., 788 F.2d 151 (3d Cir. 1986) (adopting an evenhanded approach to the grant of attorneys' fees under the Copyright Act in the Third Circuit).

The Supreme Court rejected this dual standard for fees in *Fogerty v. Fantasy, Inc.*, 510 U.S. 517, 534 (1994), ordering that "[p]revailing plaintiffs and prevailing defendants are to be treated alike" under § 505. *Id.* However, upon an examination of infringement decisions since 1994, it appears that a dual standard continues to prevail. For example, the Ninth Circuit—home of Hollywood and the entertainment industry—has continued to award fees as a general matter of right to prevailing plaintiffs. *See* Columbia Pictures Television v. Krypton Broad. of Birmingham, Inc., 106 F.3d 284 (9th Cir. 1997), *rev'd on other grounds sub nom* Feltner v. Columbia Pictures Television, 523 U.S. 340 (1998).

Meanwhile, defendants continue to be denied fees "absent bad faith motivation (such as to dominate the market in question), hard-ball tactics [such as discovery abuse] . . . or objective unreasonableness (such as pursuing a claim against a defendant after dismissal of the identical claim against a co-defendant)." 4 NIMMER & NIMMER, *supra* note 29, § 14.10[D][3][b]. Thus, to receive fees, the case must demonstrate almost extraordinary gall and bad judgment by the plaintiff. *See, e.g.*, Bond v. Blum, 317 F.3d 385 (4th Cir. 2003) (awarding fees to defendant on grounds that suit was frivolous and motivated by a desire to suppress the underlying facts of the plaintiff's work—not to protect the creative expression embodied in the manuscript).

THE *PAS DE DEUX* OF PHARMACEUTICAL REGULATION AND INNOVATION: WHO'S LEADING WHOM?

Ron A. Bouchard, Jamil Sawani, Chris McLelland,
Monika Sawicka, & Richard W. Hawkins[†]

ABSTRACT

Global drug development and regulation is undergoing a substantial transition, including redefinition of the roles of public and private actors responsible for developing, regulating, and paying for therapeutic products. This shift has been accompanied by growing debate over the validity of the claim that an efficiently functioning global public health system requires acceptance of models of drug development that promote early access to therapeutic products in exchange for strong intellectual property rights. Without these rights, advocates claim pioneering drug development will not occur. Here, we challenge this view, arguing that recent regulatory efforts designed to encourage the development of new and innovative drugs through the provision of strong patent and "linkage" rights, which legally tie drug patenting and drug approval, have in fact had the opposite effect. We provide data to suggest that the pharmaceutical industry is leaning away from the development of new drugs and towards incremental changes in existing drugs as a result of firms locking in to discrete rights targets provided for by law.

TABLE OF CONTENTS

---

† Dr. Ron A. Bouchard is an Associate Professor in the Faculties of Medicine & Dentistry and Law at the University of Alberta. Dr. Richard W. Hawkins is a Professor and Canada Research Chair in the Faculty of Social Sciences at the University of Calgary. Jamil Sawani, Chris McLelland, and Monika Sawicka are Research Assistants and students in the Faculty of Law at the University of Alberta. The authors thank Professors David Levine, Joel Lexchin and Stephen Newman, as well as Jacob Foster, Kirsten Burrows and Cameron Sklar for comments and criticism as the work evolved. The work was funded by a career establishment award (RAB) from the Alberta Heritage Foundation for Medical Research (AHFMR). Correspondence to: ron.a.bouchard@gmail.com.

INTRODUCTION

Global drug development is currently undergoing a substantial transition, including major redefinition of the responsibilities of those who develop, regulate and consume therapeutic products. This shift has been accompanied by growing debate over the validity of the claim that an efficiently functioning public health system requires acceptance of emerging lifecycle, or "real world," models of drug regulation that promote early access to innovative therapeutic products, enhanced post-market surveillance, and strong intellectual property and regulatory (IPR) rights. Indeed, IPR rights are assumed necessary for all stages of the therapeutic product lifecycle, including publicly funded medical research, university technology transfer, private research and development activities, regulatory submission, and now even the post-market stage. Advocates claim that without IPR rights pioneering drug development would not occur and that the public would be left without breakthrough remedies. The goal of the research discussed in the present Article is to investi-

gate this claim empirically and to assess how IPR rights might be used more effectively to encourage innovation in the medical sciences. In particular, we investigate whether regulatory incentives specifically intended to stimulate innovation in the pharmaceutical sector via IPR rights are producing such innovations.

The study is split into three sections. The first is an empirical investigation into the type of drugs approved by domestic Canadian regulators as regulatory incentives intended to stimulate innovation came into force. The primary goal of this study is to quantitatively analyze various types of "new" and "follow-on" drugs. A related, though smaller, component is to investigate trends for these drug types in the context of Canada's emerging lifecycle regulatory regime for drug approval, referred to as the "Progressive Licensing Framework."[1] Progressive licensing is currently enshrined in Bill C-51. Given its emphasis on promoting early access, enhanced post-market scrutiny, and strong IPR rights, progressive licensing offers an excellent opportunity to probe the relationship between drug approval, drug patenting, and innovation in an emerging drug regulation model.

The second is an empirical study of patents and patent litigation associated with the various types of drug approvals identified in the first section. The primary goal of this project is to show how government regulation shapes the domestic market for brand name and generic products. Particular attention is given to changes in patenting and litigation patterns before and after the establishment of the Canadian "linkage regulations" regime in 1993, referred to as the Patented Medicines (Notice of Compliance) Regulations (NOC Regulations).[2] Linkage regulations are critical to drug development, as they legally tie drug approval to drug patenting and litigation and thus represent a primary mechanism by which regulators promote drug development in exchange for IPR rights.

The third section is an analytical model of regulated pharmaceutical innovation, which focuses on the effectiveness of regulatory incentives intended to encourage innovation. Of particular interest is the synchronization of drug approval, patenting, and litigation data to the establishment of NOC

1. *See generally* HEALTH CANADA, BLUEPRINT FOR RENEWAL: TRANSFORMING CANADA'S APPROACH TO REGULATING HEALTH PRODUCTS AND FOOD (2006), *available at* http://www.hc-sc.gc.ca/ahc-asc/alt_formats/hpfb-dgpsa/pdf/hpfb-dgpsa/blueprint-plan-eng.pdf [hereinafter HEALTH CANADA, BLUEPRINT]; HEALTH CANADA, THE PROGRESSIVE LICENSING FRAMEWORK CONCEPT PAPER FOR DISCUSSION (2006), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfbdgpsa/pdf/prodpharma/proglic_hom prog_concept-eng.pdf [hereinafter HEALTH CANADA, PLF CONCEPT PAPER]; Neil Yeates et al., *Health Canada's Progressive Licensing Framework*, 176 CAN. MED. ASS'N J. 1845 (2007).

2. Patented Medicines (Notice of Compliance) Regulations SOR/1993-133 (Can.).

Regulations and progressive licensing. Given that progressive licensing is still being formally incorporated into the nation's regulatory regime, the majority of the analysis focuses on the relationship between drug approval, patenting, and litigation under the NOC Regulations.

I. BACKGROUND

A. PHARMACEUTICAL MARKET

Pharmaceutical products occupy an established and rapidly growing niche in modern health care. Estimated global pharmaceutical sales were U.S. $773 billion in 2008, up from $605 billion and only $298 billion in 2005 and 1998, respectively.[3] Sales growth has been strong in North America (12.6% per year from 1998 to 2005) compared to Europe (9.3%),[4] with the former accounting for the largest share of global sales (46%) compared to the latter (29.97%).[5] Even in a relatively small and growing market such as Canada,[6] more than 22,000 pharmaceutical products are currently available[7] and this number is growing rapidly.[8] Indeed, prescription drugs comprise the fastest rising component of domestic health care spending.[9] By 2006, drug expenditures in Canada rose to 17.4% of total health expenditures, up from 9.6% in 1985.[10]

---

3. INTERCONTINENTAL MARKETING SERVICES HEALTH INC., GLOBAL PHARMACEUTICAL SALES, 2001–2008 (2008); INTERCONTINENTAL MARKETING SERVICES HEALTH INC., GLOBAL PHARMACEUTICAL SALES, 1998–2005 (2005). As noted by IMS, the "value of the global pharmaceutical market in 2010 is expected to grow 4–6 percent on a constant-dollar basis, exceeding $825 billion, driven by stronger near-term growth in the U.S. market" and "is expected to expand to $975+ billion by 2013." Gary Gatyas & Clive Savage, *IMS Forecasts Global Pharmaceutical Market Growth of 4 - 6% in 2010; Predicts 4 - 7% Expansion Through 2013*, INTERCONTINENTAL MARKETING SERVICES HEALTH CAN., OCT. 7, 2009.

4. OFFICE OF FAIR TRADING, ANNEX D: GLOBAL OVERVIEW OF THE PHARMACEUTICAL INDUSTRY 8 (2007).

5. MEDICINES AUSTRALIA, GLOBAL PHARMACEUTICAL INDUSTRY FACTS AND FIGURES 1 (2007).

6. *See* PATENTED MEDICINE PRICES REVIEW BOARD, ANNUAL REPORT 2008, at 37 (2009), *available at* http://www.pmprb-cepmb.gc.ca/cmfiles/PMPRB-AR08-E.pdf. Canada's share of drug sales in major markets increased from 2.4% in 2001 to 3.8% in 2008. More significantly, domestic growth in pharmaceutical sales was 7% from 2007 to 2008 compared with 2.7% in all major markets and 1% in the United States over the same time frame. *Id.*

7. HEALTH CAN., ACCESS TO THERAPEUTIC PRODUCTS: THE REGULATORY PROCESS IN CANADA 3 (2006), *available at* http://www.hc-sc.gc.ca/ahc-asc/alt_formats/hpfb-dgpsa/pdf/pubs/access-therapeutic_acces-therapeutique-eng.pdf.

8. CAN. INST. FOR HEALTH INFO., DRUG EXPENDITURE IN CANADA 1985 to 2008, at 6 (2009).

9. Trudo Lemmens & Ron A. Bouchard, *Regulation of Pharmaceuticals in Canada*, *in* CANADIAN HEALTH LAW AND POLICY 311, 312 (Jocelyn Downie et al. eds., 3rd ed. 2007).

10. CAN. INST. FOR HEALTH INFO., *supra* note 8, at 3. Total drug expenditures were CN $4 billion, $10 billion, and $18 billion in 1985, 1995, and 2002, increasing to $25.5 billion

Indeed, drug expenditures grew faster than all other expenses within the Canadian health care system, with an average growth rate of 9.4% between 1985 and 2006 compared with 6.6% for total health spending.[11] Similarly, per capita expenditures increased on average 8.2% per annum between 1985 and 2006, faster than France, Germany, Japan, Sweden, Finland, Norway, and other European nations.[12] Between June 2004 and June 2005 alone, a total of 378 million prescriptions were filled in Canada.[13] According to Organisation for Economic Co-operation and Development (OECD) data, Canada ranked third in the world in per capita drug expenditures by 2002, behind only the United States and France.[14] Drug sales with patent protection lead the way in pharmaceutical expenditures. Between 1990 and 2008, patented drug product sales in Canada increased 764%, from CN $1.7 billion to $13 billion per annum.[15] Global and domestic pharmaceutical markets therefore are entrenched and growing more rapidly than other health care segments.

B. DRUG APPROVAL

While drug products have become an essential element of domestic and global public health systems, concerns have nevertheless been raised about the willingness of the public to underwrite the cost of drugs that are extensions of already marketed products. Indeed, there has been considerable debate over the last 25 years relating to the social benefits of "new" drug products versus those referred to variously as "follow-on," "incremental," "line extension," "me too," and "supplemental" products. To this list one can add generic drugs that are bioequivalent to already marketed products. This is because all drug products that are not considered breakthrough or pioneering in nature represent by definition some form of technology appropriation, i.e., they come into being as a result of a party's ability to capture profits generated from their own or related inventions.[16] Many commentators have derided the social value of follow-on innovations.[17] Others, however, have claimed

---

in 2006. Similarly, per capita expenditures were CN $150, $350 and $600 for the same fiscal years, increasing to $776 in 2006. *Id.* at 6–8.

11. *Id.* at 60–63.

12. *Id.* at 31.

13. INTERCONTINENTAL MARKETING SERVICES HEALTH, COMPUSCRIPT REPORT 2004, at 1 (2004).

14. ORG. FOR ECON. CO-OPERATION AND DEV., OECD HEALTH DATA 2004 (2004).

15. PATENTED MEDICINE PRICES REVIEW BOARD, *supra* note 6, at 23.

16. *See generally* David J. Teece, *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy*, 15 RES. POL'Y 285 (1986).

17. *See, e.g.*, JAMES LOVE, CONSUMER PROJECT ON TECHNOLOGY, EVIDENCE REGARDING RESEARCH AND DEVELOPMENT INVESTMENTS IN INNOVATIVE AND NON-INNOVATIVE MEDICINES 20 (2003); Joel Lexchin, *Intellectual Property Rights and the Canadian*

that follow-on drugs represent a critical component of pharmaceutical industry innovation and that dire consequences will follow should policy-makers alter the current basket of legal and regulatory incentives for innovation.[18] An example of the tension between the utility of new and existing therapies is provided by the intensity of debate over Health Technology Assessment (HTA)[19] and Cost Effectiveness Research (CER),[20] particularly as it relates to the American Recovery and Reinvestment Act of 2009.[21]

1. Drug Approval Process and Terms

Given decades of effort towards global regulatory harmony, it is not surprising that the regulatory framework for drug approval in Canada parallels that of the U.S. Food & Drug Administration (FDA).[22] In both countries, drugs submitted through "New" or "Supplementary" pathways, can be classified as "First in Class," "Me Too," or "Line Extensions," under appropriate circumstances undergo some form of "expedited review," and can contain a "New Chemical Entity" (NCE) or "New Active Substance" (NAS). Typically, a sponsor files a New Drug Submission (NDS)[23] containing sufficient data on drug safety, efficacy, and quality to warrant approval (referred to as No-

Pharmaceutical Marketplace: Where do We Go from Here?, 35 INT'L. J. HEALTH SERV. 237, 243 (2005); *Drugs in 2001: A Number of Ruses Unveiled*, 11 PRESCRIRE INT'L 58, 58 (2002).

18. *See, e.g.*, Joshua Cohen & Kenneth Kaitin, *Follow-On Drugs and Indications: The Importance of Incremental Innovation to Medical Practice*, 15 AM. J. THERAPEUTICS 89, 91 (2008).

19. *See generally* Egon Jonsson, *Development of Health Technology Assessment in Europe*, 18 INT'L J. TECH. ASSESSMENT HEALTH CARE 171 (2002).

20. COMM. ON COMPARATIVE EFFECTIVENESS RES. PRIORITIZATION, INST. OF MED. OF THE NAT'L ACADS., INITIAL NATIONAL PRIORITIES FOR COMPARATIVE EFFECTIVENESS RESEARCH (2009); FED. COORDINATING COUNCIL FOR COMPARATIVE EFFECTIVENESS RES., U.S. DEPT. OF HEALTH & HUMAN SERVS., REPORT TO THE PRESIDENT AND THE CONGRESS (2009), http://www.hhs.gov/recovery/programs/cer/cerannualrpt.pdf; G. Caleb Alexander & Randall S. Stafford, *Does Comparative Effectiveness Have a Comparative Edge?*, 301 J. AM. MED. ASS'N 2488 (2009); Jerry Avorn, *Debate about Funding Comparative-Effectiveness Research*, 360 NEW ENG. J. MED. 1927 (2009); John K. Iglehart, *Prioritizing Comparative-Effectiveness Research—IOM Recommendations*, 361 NEW ENG. J. MED. 325 (2009); Peter Singer, *Why We Must Ration Health Care*, N.Y. TIMES MAG., July 15, 2009, at MM38; Hans-Georg Eichler et al., *Use of Cost-Effectiveness Analysis in Health-Care Resource Allocation Decision-Making: How Are Cost-Effectiveness Thresholds Expected to Emerge?*, 7 VALUE IN HEALTH 518 (2004).

21. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009).

22. *See* Lemmens & Bouchard, *supra* note 9, at 321; *see generally* Patricia I. Carter, *Federal Regulation of Pharmaceuticals in the United States and Canada*, 21 LOY. L.A. INT'L & COMP. L. REV. 215 (1999).

23. Lemmens & Bouchard, *supra* note 9, at 325; *see also* Food and Drug Regulations, C.R.C., ch. 870, at § C.08.002(1)(a) (2009). The Food and Drug Regulations are propagated under the general authority of the Food and Drugs Act, R.S.C., ch. F-27 (1985).

tices of Compliance or NOCs).[24] A Supplemental New Drug Submission (SNDS) may be filed for changes to a drug already marketed by that sponsor.[25] These include amendments to dosage, strength, formulation, manufacture, labeling, route of administration, or indication.[26] Products associated with an SNDS are typically referred to as line extensions, referring to the fact that they are extensions of already marketed products.[27] Generic manufacturers submit an Abbreviated New Drug Submission (ANDS) to obtain an NOC requiring that generic drugs be pharmaceutically equivalent to the reference brand name product.[28] Generic sponsors may also submit Supplemental Abbreviated New Drug Submissions (SANDS) when changes are made to a drug already on market. Consequently, both brand name and generic firms can make "new" and "supplemental," or "follow on," submissions.

NOCs can be granted in an expedited fashion under domestic food and drug law in two ways.[29] One is through Priority Review, which refers to the fast-tracking of eligible drug candidates "intended for the treatment, prevention or diagnosis of serious, life-threatening or severely debilitating diseases or conditions" with an "unmet medical need or for which a substantial improvement in the benefit/risk profile is demonstrated."[30] Evidentiary requirements for safety, efficacy, and quality parallel those for non-priority submissions; the main difference is an accelerated review time.[31] In the second path, sponsors may be granted an "NOC with conditions"

24. Food and Drug Regulations, C.R.C., ch. 870 § C.08.002(2) (2009); Lemmens & Bouchard, *supra* note 9, at 325; *see also* HEALTH CAN., THERAPEUTIC PRODUCTS PROGRAMME GUIDELINE: PREPARATION OF HUMAN NEW DRUG SUBMISSIONS (1991), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/prephum-eng.pdf.

25. Food and Drug Regulations, C.R.C., ch. 870 § C.08.003 (2009).

26. *Id.* at § C.08.003(2); *see also* Lemmens & Bouchard, *supra* note 9, at 326.

27. Lexchin, *supra* note 17, at 243; *see generally* Song Hee Hong et al., *Product-Line Extensions and Pricing Strategies of Brand Name Drugs Facing Patent Expiration*, 11 J. OF MANAGED CARE PHARMACY 746 (2005).

28. The term "bioequivalence" refers to the requirement that the generic product must be equivalent to the already marketed "reference product" with regard to chemistry, manufacturing, route of administration, use, and therapeutic and adverse systemic effects. *See also* Food and Drug Regulations, C.R.C., ch. 870, at §§ C.08.001.1, C.08.002.1(1) (defining and discussing "Canadian reference product" and "pharmaceutical equivalent").

29. *See generally* Ron A. Bouchard & Monika Sawicka, *The Mud and the Blood and the Beer: Canada's New Progressive Licensing Framework for Drug Approval*, 3 MCGILL J.L. & HEALTH 49, 58–59 (2009).

30. HEALTH CAN., GUIDANCE FOR INDUSTRY: PRIORITY REVIEW OF DRUG SUBMISSIONS 1–2, 4 (2009), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/priordr-eng.pdf.

31. Lemmens & Bouchard, *supra* note 9, at 328.

(NOC/c)[32] for eligible NDS or SNDS submissions directed to serious, life-threatening or severely debilitating diseases, or conditions for which there is promising evidence of clinical effectiveness based on available data.[33] In addition to less onerous evidentiary requirements, the targeted review time for NOC/c approval is significantly accelerated compared to that for standard NDS review.[34] The main difference with Priority Review is that NOC/c licensure is granted on the condition that the sponsor will perform additional post-market studies to confirm alleged benefits and risks.

While the definitions of new and supplementary (NDS and SNDS) brand name submissions, standard and supplementary generic submissions (ANDS and SANDS), and pathways for expedited review (Priority Review and NOC/c) are relatively simple and straightforward, the definitions of First in Class and Me Too drugs are much less so.[35] In Canada, First in Class drugs are those that consist of either (a) a new family of active ingredient(s), also known as New Active Substance (NAS),[36] or (b) old active ingredient(s) used for the treatment of a new indication. A drug is First in Class if there is no other drug on the market that belongs to the same compound family that is used for the same indication.[37] Conversely, Me Too drugs are those that offer "important therapeutic options," but that may have little or no change to the benefit-risk profile.[38] Essentially, Me Too drugs are comparable to other drugs in terms of compound and indication.[39]

Previously referred to as a "New Chemical Entity,"[40] the definition of

32. NOC/c approvals are granted pursuant to § C.08.004(1), in compliance with the conditions of use stipulated in §§ C.08.002(1)(g), C.08.002(1)(h), C.08.006(2)(b), and C.05.006(2)(a) of the Food and Drug Regulations, C.R.C., ch. 870 (2009).

33. HEALTH PRODS. & FOOD BRANCH, HEALTH CAN., GUIDANCE DOCUMENT: NOTICE OF COMPLIANCE WITH CONDITIONS (NOC/C) (2007), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/noccg_accd-eng.pdf.

34. HEALTH CANADA, *supra* note 7, at 10–11.

35. *See* Monika Sawicka & Ron A. Bouchard, *Empirical Analysis of Canadian Drug Approval Data 2001-2008: Are pharmaceutical players "Doing More With Less"?*, 3 MCGILL J.L. & HEALTH 85, 97–114 (2009).

36. DRUGS DIRECTORATE, HEALTH CAN., POLICY ISSUES—NEW ACTIVE SUBSTANCE (1991), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/nas_nsa_pol-eng.pdf.

37. Letters between author, David K. Lee, Dir., Office of Legislative and Regulatory Modernization, Health Can., Dr. Maurica Maher, Senior Scientific Advisor, Progressive Licensing Project, Health Can., and Lesley Brumell, Supervisor, Submission and Info. Policy Div., Health Can. (April–July 2008) (on file with author) [hereinafter Health Canada Personal Communication].

38. *Id.*

39. *Id.*

40. *Id.*

NAS encompasses a wide range of chemically active substances, including (a) a chemical or biological substance that has not been previously approved for sale as a drug, (b) an isomer, derivative, or salt of a chemical substance that is already approved for sale as a drug but differs in safety and efficacy properties, or (c) a biological substance previously approved for sale as a drug that differs in molecular structure, nature of the source material, or manufacturing process.[41] The scope of regulatory approval based on an NAS is thus wide and forms the basis for NDS, SNDS, First in Class, and Me Too categories, depending on the chemical nature and use of the compound.

Drugs approved through NDS and SNDS routes can be classified as either First in Class or Me Too. For the NDS route, First in Class drugs are those that contain either an NAS or are directed to a new use (or indication), whereas NDS Me Too drugs neither contain a new ingredient nor are directed to a new use, but instead have an improved benefit-risk profile. For the traditional "line extension" SNDS route, relatively small changes to existing chemical structures such as salts or isomers may still yield First in Class or Me Too designations. The difference is that while both SNDS First in Class and Me Too drugs can cover new chemical forms,[42] only drugs directed to a new use may be deemed First in Class SNDSs.[43] Those that do not are deemed Me Too.[44] Because even a follow-on First in Class drug must be directed to a new use as opposed to just a new chemical form with altered benefit-risk, a higher level of innovation is typically ascribed to SNDS and SANDS First in Class drugs as opposed to Me Too drugs.[45] It is not surprising that drugs containing an NAS can be approved via the SNDS route given the broad overlap between SNDS (change in dosage, strength, formulation, manufacture, labeling, route of administration, or indication) and NAS (isomers, derivatives, or salts of existing drugs with differing safety and efficacy profiles and/or source material and manufacturing process) requirements.[46]

2. Lifecycle Model and IPR Rights

Emerging global drug policy places increasing importance on the need to

41. DRUGS DIRECTORATE, *supra* note 36; Health Can., Drugs and Health Products—NOC Database Terminology, *available at* http://www.hc-sc.gc.ca/dhp-mps/prodpharma/notices-avis/noc-acc/term_noc_acc-eng.php.

42. Health Canada Personal Communication, *supra* note 37.

43. *Id.*

44. *Id.*

45. For a comparison of Canadian and WHO Family of International Classifications (WHO-FIC) and Me Too classifications schemes, see Sawicka & Bouchard, *supra* note 35, at 108 (comparing Tables 2 and 5).

46. See *infra* Section III.A for discussion of the difference between Me Too and First in Class drugs particularly in regards to Figure 3b and Table 4.

adopt the principles of "lifecycle" regulation.[47] Lifecycle regulation of pharmaceuticals involves all relevant research and development, clinical trial studies, regulatory approval, market authorization, and normative post-market prescribing and use by physicians and the general population.[48] As Canadian regulators recognize, the unique aspect of lifecycle regulation is the recognition that valuable knowledge about a product is continuously accumulated over its lifecycle, especially with respect to data regarding benefit-risk analysis.[49] This progression has obvious ramifications for safety problems that arise after market penetration. The assumption is that as a drug's benefit-risk profile changes with time, so too should its approval status,[50] thus allowing for an opportunity for regulators to adapt to changing conditions over time.[51]

Canada is currently in the process of integrating the lifecycle approach into its regulatory regime.[52] Under the terms of the progressive licensing framework, plans regarding post-market studies, monitoring, safety surveillance, and risk management will be required when a sponsor files its submission.[53] The standard for initial market authorization is a positive or favorable benefit-risk profile, with maintenance of market authorization requiring a continuing favorable benefit-risk profile throughout the product's life span.[54] Canada is not alone in its efforts to legislate lifecycle approaches. Indeed, the FDA,[55] U.S. Institute of Medicine (IOM),[56] and European Medicines Agency

47. *See* Hans-Georg Eichler et al., *Balancing Early Market Access to New Drugs with the Need for Benefit/Risk Data: A Mounting Dilemma*, 7 NATURE REVS. DRUG DISCOVERY 818, 823–24 (2008).

48. HEALTH CANADA, BLUEPRINT, *supra* note 1, at 3.

49. *Id.* at 16.

50. *Id.* at 17.

51. *See id.* at 12.

52. *See* Bouchard & Sawicka, *supra* note 29, at 72–77.

53. HEALTH CANADA, PLF CONCEPT PAPER, *supra* note 1, at 5.

54. *Id.* at 17, 20.

55. CENTER FOR DRUG EVALUATION AND RESEARCH, U.S. FOOD AND DRUG ADMIN., CONCEPT PAPER: PREMARKETING RISK ASSESSMENT (Mar. 3, 2003) (draft, on file with the author); CENTER FOR DRUG EVALUATION AND RESEARCH, U.S. FOOD AND DRUG ADMIN., CONCEPT PAPER: RISK MANAGEMENT PROGRAMS (Mar. 3, 2003) (draft, on file with the author); CENTER FOR DRUG EVALUATION AND RESEARCH, U.S. FOOD AND DRUG ADMIN., CONCEPT PAPER: RISK ASSESSMENT OF OBSERVATIONAL DATA: GOOD PHARMACOVIGILANCE PRACTICES AND PHARMACOEPIDEMIOLOGIC ASSESSMENT (Mar. 3, 2003) (draft, on file with the author); FOOD AND DRUG ADMIN., U.S. DEP'T. HEALTH AND HUMAN SERVS., INNOVATION STAGNATION: CHALLENGE AND OPPORTUNITY ON THE CRITICAL PATH TO NEW MEDICAL PRODUCTS (2004); Jeffery L. Fox, *FDA embraces risk-management approach*, 21 NATURE BIOTECH. 1120 (2003); *see also* Guidance on Drug Safety Information, 72 Fed. Reg. 10224 (Mar. 7, 2007).

56. BD. ON HEATH CARE SERVS., INST. OF MED. OF THE NAT'L ACADS., PATIENT SAFETY: ACHIEVING A NEW STANDARD OF CARE (Philip Aspden et al. eds., 2004). For ex-

(EMEA)[57] recognized early that drug safety was well served by lifecycle models, including articulating the need for regulating therapeutic products in light of "real world" drug use.

IPR rights remain a pivotal element of lifecycle models of drug regulation. In accordance with the terms of its National Pharmaceutical Strategy[58]

---

ample,

> Reviewers in the Center for Drug Evaluation and Research (CDER) at the Food and Drug Administration (FDA) must weigh the information available about a drug's risk and benefit, make decisions in the context of scientific uncertainty, and integrate emerging information bearing on a drug's risk-benefit profile throughout the lifecycle of a drug, from drug discovery to the end of its useful life.

Id. at S-2. For a discussion of a comprehensive, rather than silo-based, response to errors in patient care, see also COMM. ON QUALITY OF HEALTH CARE IN AM., INST. OF MED. OF THE NAT'L ACADS., TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM (Linda T. Kohn et al., 2000).

57. COMM. FOR MEDICINAL PRODS. FOR HUMAN USE, EUROPEAN MEDS. AGENCY, REPORT OF THE CHMP WORKING GROUP ON BENEFIT-RISK ASSESSMENT MODELS AND METHODS, EMEA/CHMP/15404/2007 (2007), *available at* http://www.emea.europa.eu/pdfs/human/brmethods/1540407en.pdf. The EMEA states,

> The current report describes the technical and scientific highlights of all these consultations, incorporates reflections and draws recommendations from the think-tank group. Areas for improvement in the operations of the EMEA and its scientific Committees include strengthening of both the informal and formal dialogue already in place, in order to ensure a continual exchange throughout the life-cycle of the products.

Id. at 6. For general discussion of "continuing and contextual" pre-market and post-market analysis of benefit-risk approach, see generally: COMM. FOR MEDICINAL PRODS. FOR HUMAN USE, EUROPEAN MEDS. AGENCY, GUIDELINE ON THE SCIENTIFIC APPLICATION AND THE PRACTICAL ARRANGEMENTS NECESSARY TO IMPLEMENT COMMISSION REGULATION (EC) NO 507/2006 ON THE CONDITIONAL MARKETING AUTHORISATION FOR MEDICINAL PRODUCTS FOR HUMAN USE FALLING WITHIN THE SCOPE OF REGULATION (EC) NO 726/2004, EMEA/509951/2006 (2006); COMM. FOR MEDICINAL PRODS. FOR HUMAN USE, EUROPEAN MEDS. AGENCY, REFLECTION PAPER ON BENEFIT-RISK ASSESSMENT METHODS IN THE CONTEXT OF THE EVALUATION OF MARKETING AUTHORISATION APPLICATIONS OF MEDICINAL PRODUCTS FOR HUMAN USE, EMEA/CHMP/15404/2007 (2008), *available at* http://www.emea.europa.eu/pdfs/human/brmethods/1540407enfin.pdf.

58. FED./PROVINCIAL/TERRITORIAL MINISTERIAL TASK FORCE ON THE NAT'L PHARMS. STRATEGY, NATIONAL PHARMACEUTICALS STRATEGY: PROGRESS REPORT (2006), *available at* http://www.hc-sc.gc.ca/hcs-sss/alt_formats/hpb-dgps/pdf/pubs/2006-nps-snpp/2006-nps-snpp-eng.pdf. Intellectual property rights and pharmaceutical innovation comprise three of the five "pillars" of the nation's pharmaceutical policy. According to the Government of Canada, the five "pillars" of federal pharmaceutical policy are the following: (1) intellectual property, (2) pharmaceutical research and development, (3) international trade policy, (4) health care, and (5) consumer protection. Barbara Oullet, Pharmaceutical Management and Price Control in Canada 7 (Mar. 31, 2006) (presentation to the North American Pharmaceutical Summit, on file with the Berkeley Technology Law Journal). The National Pharmaceutical Strategy states that "Governments recognize the crucial role the innovative

and Smart Regulations initiative,[59] the government of Canada sees itself as a leader in developing an innovative drug regulation platform and in providing unique regulatory incentives to the pharmaceutical industry.[60] In this capacity, Canadian regulators are acting in tandem with their American and European counterparts, all of which claim that therapeutic product development is crucial for national prosperity and productivity in the global marketplace.[61] The specific goals of the latest round of reform are to: (1) facilitate biomedical innovation; (2) create incentives for drug development when the market itself does not; (3) allow for earlier access to new drugs; (4) create an informed consumer; and (5) increase the threshold for post-market drug safety. The emphasis on providing IPR rights incentives to the industry in order to support innovation follows numerous reports from the government and its consultants over the last number of years on the growing productivity gap in Canada and the commercialization of novel therapeutic products emanating from publicly funded medical research.[62]

A cornerstone of Canadian domestic lifecycle regulation is NOC/c-type approval.[63] This refers to a recalibrated balance between faster access to nov-

---

pharmaceutical industry plays in the development of breakthrough drugs and that intellectual property protection is key to encouraging and supporting innovation." NATIONAL PHARMACEUTICALS STRATEGY, at 39.

59. EXTERNAL ADVISORY COMM. ON SMART REGULATION, SMART REGULATION: A REGULATORY STRATEGY FOR CANADA (2004), *available at* http://dsp-psd.pwgsc.gc.ca/Collection/CP22-78-2004E.pdf.

60. *See* Robert Peterson, Dir. General, Therapeutic Products Directorate, Lecture to the Ottowa Regional Conference, Innovation in Drug Regulation: Canada as a Leader (Feb. 11, 2005).

61. *See* Ron A. Bouchard, *Balancing Public and Private Interests in the Commercialization of Publicly Funded Medical Research: Is There a Role for Compulsory Government Royalty Fees?*, 13 B.U. J. SCI. & TECH. L. 120, 158–64 (2007).

62. *See, e.g.*, EXPERT PANEL ON COMMERCIALIZATION, PEOPLE AND EXCELLENCE: THE HEART OF SUCCESSFUL COMMERCIALIZATION 6 (2006); BRIAN GUTHRIE & TREFOR MUNN-VENN, CONFERENCE BD. OF CAN., SIX QUICK HITS FOR CANADIAN COMMERCIALIZATION: LEADERS' ROUNDTABLE ON COMMERCIALIZATION 1 (2005). For an analogous discussion of the importance of industrial intellectual property incentives in national productivity and prosperity in the United States, see generally COUNCIL ON COMPETITIVENESS, INNOVATE AMERICA: NATIONAL INNOVATION INITIATIVE SUMMIT AND REPORT (2005).

63. *See* HEALTH CANADA, PLF CONCEPT PAPER, *supra* note 1, at 20. Health Canada states,

> In keeping with the proposed life-cycle approach, maintenance of market authorisation could require a continuing favourable benefit-risk profile for the authorised conditions of use throughout the product's lifespan. The favourable benefit-risk profile would be based on the same elements required for initial market authorisation with some possible additions, i.e., substantial evidence of efficacy, safety, and quality; substantial evidence

el remedies (termed "flexible departure") and enhanced post-market oversight of safety, efficacy, and benefit-risk, with the possibility of revocation of initial approval if the terms of initial approval are not met. Unlike Priority Review, continuing approval after initial regulatory approval is contingent upon whether pharmaceutical sponsors meet the terms and conditions assigned to the NOC/c.[64] At first glance, emphasis on NOC/c over Priority Review may seem inconsistent with the lifecycle approach. However, fast-tracking eligible NDSs and SNDSs via Priority Review results in faster approval without a change in the amount of scientific evidence required prior to market entry.[65] The process remains front-loaded in that it does not demand that sponsors conduct post-marketing studies as a means to maintain approval status. In comparison, the NOC/c mechanism demands that sponsors are subject to legal scrutiny beyond initial market authorization in exchange for faster approval. The process is considerably more back-loaded in this regard and thus is more consistent with the lifecycle approach. It is reasonable to conclude therefore that NOC/c approvals are an excellent proxy for lifecycle regulation compared with Priority Review or approval via conventional NDS and SNDS pathways.

While lifecycle models have several advantages over existing approval models,[66] concerns persist that releasing drugs into the market earlier may be misguided, given evidence that pharmaceutical firms typically do not meet conditions associated with approval once in the market in the absence of legislation compelling them to do so.[67] Moreover, concerns have been expressed

---

> for a favourable overall benefit-risk profile regarding the product and evidence of other important benefit-risk considerations relating to the impact of market authorisation on external decision-makers.

Id. Health Canada then clarifies the balance between the uncertainties of drug development and the importance of bringing new drugs to market as fast as reasonably possible:

> When a manufacturer is considering departing from the baseline requirement for substantial evidence of efficacy and safety for initial market authorisation, a more flexible approach regarding the underlying efficacy and safety evidence is envisaged when there is a compelling reason. While the regulatory requirement for a favourable benefit-risk profile for the drug's use under the proposed conditions would remain, initial requirements for substantial evidence of efficacy and safety may be counterbalanced against other, important evidence concerning contextual benefit-risk considerations. For example, the potential benefits of bringing the drug to market are deemed to outweigh the relatively increased uncertainty regarding the safety and efficacy.

Id. at 20–21.

64. *See generally* sources cited *supra* note 1.

65. Health Canada Personal Communication, *supra* note 37.

66. *See generally* Eichler et al., *supra* note 47.

67. UNION OF CONCERNED SCIENTISTS, VOICES OF SCIENTISTS AT FDA:

over the reading-in of TRIPS-based provisions incorporating strong IPR rights,[68] and specific language contemplating incorporation into policy and regulations any relevant knowledge, documents, or information produced by industry and its trade organizations.[69] While it is reasonable to speculate that the goal of these provisions is to facilitate global regulatory harmony, there is some unease that practices such as these serve the nation's economic goals more than its public health mandate.[70] This interpretation is bolstered by statements from various branches of government.[71]

PROTECTING PUBLIC HEALTH DEPENDS ON INDEPENDENT SCIENCE 1 (2006), http://www.ucsusa.org/assets/documents/scientific_integrity/Voices_of_Federal_Scientists.pdf. The Union of Concerned Scientists stated,

> From 2005 to 2007, the Union of Concerned Scientists (UCS) conducted five surveys of federal scientists to evaluate how U.S. agencies use—and misuse—science to make policy decisions . . . The results reveal extensive political interference in federal science, with serious and wide-ranging consequences for our health, safety, and environment. This interference has weakened the federal scientific enterprise and impaired the ability of U.S. agencies to serve the public interest, with the potential for long-lasting harm to the federal scientific work force.

Id.; *see also* Daniel Carpenter et al., *Drug-Review Deadlines and Safety Problem*s, 358 NEW ENG. J. MED. 1354 (2008); David B. Ross, *The FDA and the Case of Ketek*, 356 NEW ENG. J. MED. 1601 (2007) (discussing the illustrative case of the drug Ketek); Gardiner Harris, *FDA Scientists Accuse Agency Officials of Misconduct*, N.Y. TIMES, Nov. 18, 2008, at A15 (describing a letter sent by the FDA scientists on October 14, 2008 to Congress alleging FDA is engaged in "serious misconduct" by approving unsafe or ineffective medications); Susan Okie, *What Ails the FDA?*, 352 NEW ENG. J. MED. 1063, 1065–66 (2005) .

68. Bill C-51, 2nd Sess. 39th Parl., cl. 11 § 30(3) (Can. 2008). This bill states,

> Without limiting or restricting the authority conferred by any other provisions of this Act for carrying into effect the purposes and provisions of this Act, the Governor in Council may make the regulations that the Governor in Council considers necessary for the purpose of implementing, in relation to drugs, Article 1711 of the North American Free Trade Agreement or paragraph 3 of Article 39 of the Agreement on Trade-related Aspects of Intellectual Property Rights set out in Annex 1C to the WTO Agreement.

Id.

69. *Id.* at cl. 11 § 30(7)(b).

70. Janice Graham, *Smart Regulation: Will the Government's Strategy Work?*, 173 CAN. MED. ASS'N J. 1469, 1469 (2005).

71. *See, e.g.*, HEALTH PRODS. AND FOOD BRANCH, HEALTH CAN., CLINICAL TRIALS REGULATORY REVIEW—STAKEHOLDER WORKSHOP 6 (2007), *available at* http://www.hc-sc.gc.ca/dhp-mps/alt_formats/hpfb-dgpsa/pdf/prodpharma/ctrf_o_eccr_a_2007-03-26-eng.pdf [hereinafter HEALTH CANADA, STAKEHOLDER WORKSHOP]; HEALTH CANADA, BLUEPRINT, *supra* note 1, at 8–9; HEALTH CANADA, PLF CONCEPT PAPER, *supra* note 1, at 21; Reg Alcock, President, Treasury Bd., Speech Accompanying the Launch of the Government of Canada's Implementation Plan for Smart Regulation (Mar. 24, 2005) (transcript available at http://www.tbs-sct.gc.ca/media/ps-dp/2005/0324_e.asp); Peterson, *supra* note 60; *see also* Lemmens & Bouchard, *supra* note 9.

C. LINKAGE REGULATIONS

One of the most strongly contested aspects of pharmaceutical policy concerns the role of intellectual property and regulatory rights in providing economic incentives to firms and in shaping the agenda for basic medical research.[72] "Intellectual property rights" usually refers to traditional patent rights, while "regulatory rights" encompasses the growing cache of exclusivity periods (e.g., data, market, and pediatric) attached to drug-approval data. The combination of both is referred herein as "IPR rights."

A relatively new addition to the basket of IPR rights is a novel form of legal ordering referred to as "linkage regulations." So named because they tie patent protection for marketed pharmaceuticals to the drug approval process, linkage regulations enable brand name pharmaceutical firms to list as many patents as are deemed relevant to a marketed product on a patent register.[73] In Canada, this occurs under the aegis of the Patented Medicines (Notice of Compliance) Regulations.[74] Each patent must be demonstrated in litigation to be invalid or not infringed for generic market entry.

Linkage regulations are critical to the maintenance of monopoly pricing by brand name pharmaceutical firms as blockbuster drugs near the end of their conventional patent protection (patents on NCEs or NASs). This is because patents listed on the patent register effectively allow for a second "term" of patent protection, provided that patents are deemed relevant to the already marketed product. As such, linkage regulations represent a primary mechanism by which regulators promote drug development in exchange for private IPR rights.

Given the pivotal nature of the relevance requirement, it is not surprising that legislators and the courts have battled intensely over the issue. Early Federal Court of Appeal jurisprudence rejected the notion of a strict relevance requirement, opting instead for a broad statutory reading to the effect that patents need only be relevant to a medicine rather than the drug form specifically approved by regulators.[75] In other words, patents could be listed

72. David H. Guston, *Innovation Policy: Not Just a Jumbo Shrimp*, 454 NATURE 940 (2008).

73. Ron A. Bouchard, *Should Scientific Research in the Lead-up to Invention Vitiate Obviousness Under the Patented Medicines (Notice of Compliance) Regulations: To Test or Not to Test?*, 6 CAN. J. L. TECH. 1, 1–27 (2007) [hereinafter Bouchard, *Scientific Research*]; Ron A. Bouchard, *Living Separate and Apart is Never Easy: Inventive Capacity of the PHOSITA as the Tie that Binds Obviousness and Inventiveness in Pharmaceutical Litigation*, 4 U. OTTAWA L. & TECH. J. 1 (2007), *available at* http://ssrn.com/abstract=958927 [hereinafter Bouchard, *PHOSITA*]; Edward Hore, *A Comparison of United States and Canadian Laws as They Affect Generic Pharmaceutical Market Entry*, 55 FOOD & DRUG L.J. 373 (2000).

74. Patented Medicines (Notice of Compliance) Regulations SOR/1993-133 (Can.).

75. Eli Lilly Can. Inc. v. Canada, [2003] F.C.A. 24, ¶ 32, 34–35 (Can.).

generally for a drug rather than against a specific drug submission. This made it comparatively easier for brand name firms to extend patent monopolies via linkage regulations. In 2006, amendments made to the NOC Regulations required listed patents to contain at least one claim to the medical ingredient, formulation, dosage form, or use for which approval was granted.[76] This was supported by the Supreme Court of Canada in *AstraZeneca*.[77] Shortly afterward, the Federal Court of Appeal reversed its position, holding that specific relevance is required between the patent sought to be listed and the drug submission against which it was listed.[78] The intense volleying back and forth between litigants, legislators, and the courts over the issue of relevance suggests that framing a system of pharmaceutical innovation around the nexus between continuing patenting activity on drugs that have already been approved and related drugs is a contentious model of innovative drug development contingent upon strong IPR rights.

Prior to the NOC Regulations coming into force, the Supreme Court of Canada noted that patent protection and regulatory approval of pharmaceuticals were governed by different statutes as well as different policy goals and objectives. Given the specific language employed,[79] it is reasonable to con-

---

76. Patented Medicines (Notice of Compliance) Regulations SOR/1993-133 (Can.).

77. AstraZeneca Canada v. Can. Inc., [2006] 2 S.C.R. 560, ¶ 21–23 (Can.) The court stated,

> I emphasize the words in s. 4(5) that in the case of patents added afterwards, "the first person must identify the submission to which the patent list or the amendment relates, including the date on which the submission was filed". In addition, s. 3(3) provides that "[n]o information submitted pursuant to section 4 shall be included on the register until after the issuance of the notice of compliance in respect of which the information was submitted." These provisions, it seems to me, provide an important key to understanding the scheme. Entry of the "Patent list" does not destroy the linkage between the patent and the submission(s) to which it relates, nor to the NOC to which the submission(s) are directed. Specific patents are associated with one or more NDS, ANDS or SNDS, which in turn (if approved) give rise to specific NOCs, which in turn approve a specific manufacturer's product, which a generic manufacturer may seek to copy.).

Id. at ¶ 21.

78. Wyeth Can. v. Ratiopharm, Inc., [2008] 1 F.C. 447, ¶ 30 (Can).

79. AstraZeneca Can. Inc. v. Canada, [2006] 2 S.C.R. 560, ¶ 12 (Can.). The court noted that

> [t]he *NOC Regulations* lie at the intersection of two regulatory systems *with sometimes conflicting objectives*. First, is the law governing approval of new drugs, which seeks to ensure the safety and efficacy of new medications before they can be put on the market. The governing rules are set out in the *Food and Drugs Act*, R.S.C. 1985, c. F-27 (*FDA*) and the *Food and Drug Regulations*, C.R.C. 1978, c. 870. The *FDA* process culminates (if success-

clude that the court was referring to the previously divergent goals of public health policy and industrial/economic policy. The language employed by the court further suggests that these two policy branches have formally converged in the form of the NOC Regulations and that private IPR rights are viewed as a primary driver of this convergence. Indeed, government Regulatory Impact Analysis Statements (RIAS) have forged a clear policy objective of stimulating innovation in the pharmaceutical sector predicated on industrial IPR rights, including via linkage regulations.[80]

The Canadian NOC Regulations were modeled after the U.S. Hatch Waxman linkage regime,[81] which ties patent protection under the Patent Act[82] to drug approval under the Food, Drug & Cosmetic Act[83] via patent listings in the Orange Book.[84] While the United States and Canada are currently the only two jurisdictions formally employing linkage regulations to stimulate innovation, there is movement afoot to institute linkage regulation regimes in other jurisdictions, and the United States is moving toward including linkage provisions in its international trade agreements.[85] Parallel developments have

---

> ful) in the issuance of a NOC to an applicant manufacturer by the Minister of Health on the advice of his officials in the Therapeutic Products Directorate. The *FDA* objective is to encourage bringing safe and effective medicines to market to advance the nation's health. The achievement of this objective is tempered by a second and to some extent overlapping regulatory system created by the *Patent Act*, R.S.C., 1985, c. P-4. Under that system, in exchange for disclosure to the public of an invention, including the invention of a medication, the innovator is given the exclusive right to its exploitation for a period of 20 years. *Until 1993, the two regulatory systems were largely kept distinct and separate.*

Id. (emphasis added).

80. Bouchard, *supra* note 61, at 123; Bouchard, *PHOSITA*, *supra* note 73, at 46–51.

81. Drug Price Competition and Patent Term Restoration Act of 1984, Pub. L. No. 98-417, 98 Stat. 1585 (1984) (codified as amended at 21 U.S.C. § 355 (2000)) (commonly known as the Hatch-Waxman Act).

82. Patent Act, 35 U.S.C. §§ 1–376. (2006).

83. Federal Food, Drug, and Cosmetic Act, 21 U.S.C. § 355(j)(7)(A) (2006).

84. 21 U.S.C. § 355(j)(7)(A) (establishing a list of "Approved Drug Products with Therapeutic Equivalence" commonly known as the "Orange Book"); *see also* Andrew A. Caffrey, III & Jonathan M. Rotter, *Consumer Protection, Patents and Procedure: Generic Drug Market Entry and the Need to Reform the Hatch-Waxman Act*, 9 VA. J.L. & TECH. 1, 4–7 (2004) (describing the Orange Book in the context of patent litigation and drug development); Rebecca S. Eisenberg, *Patents, Product Exclusivity, and Information Dissemination: How Law Directs Biopharmaceutical Research and Development*, 72 FORDHAM L. REV. 477, 483 (2003) ("Holders of approved NDAs are required to disclose all patents that they believe would be infringed by unauthorized sales of the approved drug, and the FDA publishes the list in a publication called the Orange Book.").

85. *See, e.g.*, Judit Rius Sanjuan, *Patent-Registration Linkage*, CPTECH, Apr. 3, 2006, http://www.cptech.org/publications/CPTechDPNo2Linkage.pdf.

also taken place in other segments of the medical product development landscape. For example, both the United States National Institutes of Health (NIH) and Canadian Institutes for Health Research (CIHR) have stated that strong industrial and economic rights for biomedical firms are a fundamental linchpin for governments to fulfill their public health mandates.[86] Further evidence for convergence of domestic public health and economic policy is provided by the fact that although drug approval and drug policy have historically been controlled by Health Canada, drug patenting, drug approval-linkage, and innovation policy have become increasingly under the control of Industry Canada—setting up potential tension between the two branches of government.[87] A similar "push-pull" between public health and economic concerns is found in legislation and policy that underpins publicly funded medical research, technology transfer, and related commercialization activities in the United States and Canada.[88]

The specific platform of legal rights associated with pharmaceutical products has critical public health ramifications, not only because firms and policy-makers view it as a major economic driver for innovation in the life sciences,[89] but also because the rate and direction of innovation in the phar-

86. *See, e.g.*, EUROPEAN MEDS. AGENCY, THE EUROPEAN MEDICINES AGENCY ROAD MAP TO 2010: PREPARING THE GROUND FOR THE FUTURE, EMEA/H/34163/03/Final (2005) [hereinafter EMEA ROAD MAP]; U.S. DEPT. OF HEALTH & HUMAN SERVS., FOOD & DRUG ADMIN., INNOVATION OR STAGNATION: CHALLENGE AND OPPORTUNITY ON THE CRITICAL PATH TO NEW MEDICAL PRODUCTS (2004); Alan Bernstein, *Toward Effective Canadian Public-Private Partnerships in Health Research*, 168 CAN. MED. ASS'N J. 288 (2003); Eichler et al., *supra* note 47, at 819; Elias Zerhouni, *The NIH Roadmap*, 302 SCI. 63 (2003). For example, the EMEA ROAD MAP stipulates that the agency uses a "two-pillar approach" to make safe and effective therapeutic products available to the public. EMEA ROAD MAP, at 36. These are to facilitate more rapid access to safe and effective medicines via amendment to the existing regulatory licensing framework and to facilitate industrial innovation. While EMEA does not provide a definition of "innovation" nor a "map" of how it will facilitate innovative drug development in either its road map or its follow-up report, it can be plausibly assumed that the main economic drivers for this process will be a combination of intellectual property and regulatory rights. EUROPEAN MEDICINES AGENCY, SECOND STATUS REPORT ON THE IMPLEMENTATION OF THE EMEA ROAD MAP, EMEA/359050/2007 (2007). Eichler et al. point out that "regulators acknowledge the need to facilitate innovation and the fact that a lack of efficacious therapies *is a public health issue*." Eichler et al, *supra* note 47, at 819 (citing EMEA ROAD MAP) (emphasis added).

87. *See generally* Ron A. Bouchard & Trudo Lemmens, *Privatizing Biomedical Research—a Third Way*, 26 NATURE BIOTECHNOLOGY 31 (2008) (examining the tension between for-profit entities and the public interest in biomedical research); Bouchard, *supra* note 61.

88. For discussion of the tension between public and private interests in publicly funded medical research, see Bouchard, *supra* note 61; *see also* SHELDON KRIMSKY, SCIENCE IN THE PRIVATE INTEREST (2003).

89. *See* CANADA'S RESEARCH-BASED PHARMACEUTICAL COMPANIES (RX&D), INFORMATION GUIDE 2002, SECTION 2: INDUSTRY ISSUES (2002), *available at*

maceutical industry may be shaped *antecedently* by IPR rights incentives.

D. IPR RIGHTS AND INNOVATION POLICY

IPR rights and public policy promoting innovation have strong historical associations. Public policy in most developed nations still tends to assume basically a linear model of innovation, i.e., a product "pipeline" that begins in basic research, moves on through private research and development activities, and then to commercialization in the form of products and services.[90] This model implies a strong imperative to legally protect knowledge that has been reduced to practice as it flows through the system in the form of limited-term monopolies. For pharmaceutical innovation, the process is complicated by regulatory requirements to gain market authorization for new drugs, which is perceived as the terminus for the innovation pipeline. Accordingly, there is a considerable body of established science policy that identifies IPR rights as the major economic driver of innovation, national productivity, and translational research in the medical sciences.[91]

Despite its entrenched nature, however, the theoretical and empirical case for linear models of innovation contingent on strong IPR rights is weak. Since the 1960s, much scholarly work on innovation has indicated a highly complex, iterative process of individual and organizational learning that typi-

---

http://www.canadapharma.org/Industry_Publications/Information_Guide/section2_e.html; *see also* ASTRAZENECA CAN., THE PATENT ACT & LINKAGE REGULATIONS: ESSENTIAL TOOLS FOR THE ADVANCEMENT OF MEDICAL SCIENCE IN CANADA (2009), http://www.astrazeneca.ca/documents/en/aboutus/PatentActLinkageRegulations.pdf.

90. *See* Bouchard & Lemmens, *supra* note 87, at 35; *see generally* VANNEVAR BUSH, SCIENCE: THE ENDLESS FRONTIER (1945); DONALD STOKES, PASTEUR'S QUADRANT (1997); Benoît Godin, *The Linear Model of Innovation: The Historical Construction of an Analytical Framework*, 31 SCI. TECH. HUM. VALUES 639 (2006).

91. In its "Roadmap for Medical Research," the U.S. National Institutes of Health (NIH) defines "translational research" as research that successfully makes the transition translated from the laboratory bench to the patient bedside: "To improve human health, scientific discoveries must be translated into practical applications. Such discoveries typically begin at 'the bench' with basic research—in which scientists study disease at a molecular or cellular level—then progress to the clinical level, or the patient's 'bedside.'" NIH Roadmap for Medical Research, http://nihroadmap.nih.gov/clinicalresearch/overview-translational.asp (last visited Nov. 6, 2009). Similarly, the Canadian Institutes for Health Research (CIHR) has embedded the concept of "knowledge translation" into its statutory mandate: "The objective of the CIHR is to excel, according to internationally accepted standards of scientific excellence, in the creation of new knowledge and its translation into improved health for Canadians, more effective health services and products and a strengthened Canadian health care system." Canadian Institutes of Health Research Act § A, 2000 S.C., ch. 6 (Can.). For discussion of research in the specific context of commercialization of publically funded medical research, see generally Bouchard, *supra* note 61; KRIMSKY *supra* note 88; EXPERT PANEL ON COMMERCIALIZATION, *supra* note 62; GUTHRIE & MUNN-VENN, *supra* note 62; THE COUNCIL ON COMPETITIVENESS, *supra* note 62.

cally involves an array of public and private sector inputs with many feedbacks.[92] This body of work suggests that innovation is a dynamic combinatory process in which the probability of innovation is linked closely to the capacity to create new combinations of knowledge, resources, and skills.[93] Other empirical studies have failed to demonstrate a conclusive link between strong IPR rights policies and generally increased levels of innovation.[94] These studies suggest that the dynamics of innovation can embrace IPR rights in some circumstances, but that these rights need not comprise an essential element for innovation to occur or to increase. The implications of this scenario are especially important for innovative product development in the medical sciences, given the vast array of public health and cost considerations involved in new drug development and regulation.[95]

---

92. *See, e.g.*, HENRY CHESBROUGH, OPEN INNOVATION: THE NEW IMPERATIVE FOR CREATING AND PROFITING FROM TECHNOLOGY (2003); DOMINIQUE FORAY, THE ECONOMICS OF KNOWLEDGE (2004); RICHARD NELSON & SYDNEY WINTER, AN EVOLUTIONARY THEORY OF ECONOMIC CHANGE 27–29, 277 (1982); W. Brian Arthur, *Competing Technologies, Increasing Returns, and Lock-in by Historical Events*, 99 ECON. J. 116 (1989); Wesley M. Cohen & Daniel A. Levinthal, *Absorptive Capacity: A New Perspective on Learning and Innovation*, 35 ADMIN. SCI. Q. 128 (1990); Giovanni Dosi, *Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change*, 11 RES. POL'Y 147, 157–58 (1982); Henry Etzkowitz & Loet Leydesdorff, *The Dynamics of Innovation: From National Systems and "Mode 2" to a Triple Helix of University–Industry–Government relations*, 29 RES. POL'Y 109 (2000); Paul Nightingale, *A Cognitive Model of Innovation*, 27 RES. POL'Y 689 (1998).

93. COHEN & LEVINTHAL, *supra* note 92; W. Brian Arthur, *The Structure of Invention*, 36 RES. POL'Y 274 (2007); ERIC VON HIPPEL, DEMOCRATIZING INNOVATION (2005); C. Freeman, *Technological Infrastructure and International Competitiveness*, 13 INDUS. & CORP. CHANGE 541 (2004).

94. David C. Mowery et al., *The Growth of Patenting and Licensing by U.S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980*, 30 RES. POL'Y 99 (2001); Mariko Sakakibara & Lee Branstetter, *Do Stronger Patents Induce More Innovation? Evidence from the 1988 Japanese Patent Law Reforms*, 32 RAND J. ECON. 77 (2001); Adam B. Jaffe, *The U.S. Patent System in Transition: Policy Innovation and the Innovation Process*, 29 RES. POL'Y 531 (2000); Roberto Mazzoleni & Richard R. Nelson, *The Benefits and Costs of Strong Patent Protection: A Contribution to the Current Debate*, 27 RES. POL'Y 273 (1998). For a recent review of empirical studies, see JAMES BESSEN AND MICHAEL J. MEURER, PATENT FAILURE (2008).

95. *See* Comm'n of Patents v. Fabwerka Hoechst, [1964] S.C.R. 49, 56 (Can.). In emphasizing that courts must scrutinize pharmaceutical patents carefully in order to determine if they properly merit the grant of a monopoly privilege in light of the significant public interest at stake, the court noted that

> [i]n the particular class of case with which we are here concerned dealing with drugs and medicines, there is considerable public interest at stake, and the Commissioner should most carefully scrutinize the application to see if it merits the grant of monopoly privileges, and to determine the scope of the monopoly available.

Id.; *see generally* Catherine De Angelis et al., *Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors*, 351 NEW ENG. J. MED. 1250 (2004).

In light of the increasing disparity between the claim that an effective and efficiently operating public health system is contingent upon IPR rights[96] and the evidence disputing the legitimacy of this model,[97] therapeutic product development represents an excellent target for empirical studies of the relationship between legal incentives for innovation and product development. As noted by Jaffe, robust conclusions regarding the consequences for technological innovation of changes in patent policy are few and far between, in large part owing to a fundamental lack of empirical data.[98] The combination of recently established linkage and lifecycle regulation based models of drug de-

96. *See, e.g.*, Bouchard & Sawicka, *supra* note 29, at 65 n.168; Eichler et al., *supra* note 47.

97. *See, e.g.*, Jaffe, *supra* note 94; Mazzoleni & Nelson, *supra* note 94; Keith Pavitt, *National Policies for Technical Change: Where are the Increasing Returns to Economic Research?*, 93 PROC. NAT'L ACAD. SCI. U.S.A. 126 (1996); *see generally* MICHELE BOLDRIN & DAVID K. LEVINE, AGAINST INTELLECTUAL MONOPOLY (2008); KRIMSKY, *supra* note 88.

98. Jaffe, *supra* note 94. Jaffe notes that it is possible that the R&D boom in the late 1970s and early 1980s would not have been so large or lasted so long without enhanced IPR rights, and that it is "disquieting, however, that there is so little empirical evidence that what is widely perceived to be a significant strengthening of intellectual property protection had significant impact on the innovation process." *Id.* Jaffe further observes,

> Overall, there is a noticeable gap between the highly developed theoretical literature on patent scope and the limited empirical literature. This is due partially to the infrequency of changes in patent regimes like the one examined by Sakakibara and Branstetter. Part of the difficulty also lies in the weakness of the connection between the model constructs and quantifiable aspects of a patent regime.

Id. at 588. Finally, Jaffe comments that

> [t]his limited success is due partially to the difficulty of measuring the parameters of patent policy, and partly due to the difficulty of discerning statistically significant effects when many things have been changing at the same time. But it should surely be viewed as a challenge to researchers to try to do more.

Id. at 554. Similar conclusions were drawn by Mazzolini and Nelson and more recently by Boldrin and Levine. Mazzoleni & Nelson, *supra* note 94. Mazzolini and Nelson stated,

> The range of arguments about the positive social value of patents is obviously much wider than the area of strong empirical studies explored to date. An analyst, citing earlier studies that appear to show limited value, obviously is vulnerable to the argument that those studies do not provide evidence on some of the possibly most important functions patents serve. We cannot present here an empirically supported and intellectually persuasive argument on this broad question. The important empirical research that needs to be done in order to map out the basic facts simply has not been done yet.

Id. Boldrin & Levine, *supra* note 97, at 189–90. In a meta-analysis of empirical studies of whether introducing or strengthening patent protection leads to greater innovation, Boldrin and Levine note: "We have identified twenty three economic studies that have examined this issue empirically. The executive summary: these studies find weak or no evidence that strengthening patent regimes increases innovation; they find evidence that strengthening the patent regime increases . . . patenting!" *Id.* at 216–17; *see also* Pavitt, *supra* note 97, at 126.

velopment, therefore, provide a unique and time-sensitive opportunity to develop a domestically based yet globally relevant methodology and database for the study of pharmaceutical innovation.

Considerations, such as the aforementioned, led to the current study. Our ultimate goal is to develop an independent empirical methodology to identify patterns in the rate and direction of innovative activity by pharmaceutical firms and to analyze these data in relation to well defined regulatory incentives for pharmaceutical innovation via provision of strong IPR rights. The work is specifically designed to probe the functional and structural link between drug approval, drug patenting, drug litigation, and pharmaceutical innovation. Here, we present data from our study on the relationship between drug approval, drug regulation, and drug innovation in the domestic Canadian market. We analyzed drug approvals over the period 2001–2008, with a particular focus on the types of drugs being approved and how approvals were consistent with emerging lifecycle models of drug regulation.[99] The second major aspect of the work is a pilot study on the legal nexus between drug approval, drug patenting, and litigation, which we propose reflects trends in the broader influence of government regulation on innovation in the global pharmaceutical industry. Specifically, we argue that the global pharmaceutical industry is leaning away from the development of new drugs and towards incremental changes in existing drugs as a result of firms locking in to discrete IPR rights targets provided for by law.

II. METHODS

A. DRUG APPROVAL

Statistical analysis of drug approvals issued in Canada from January 1, 2001 to December 31, 2008 was performed as described previously.[100] Absolute numbers and fractional percentages of various types of drug approvals were calculated for each year during the eight-year test period in annual, quarterly, monthly and daily increments. Drug approvals used for calculation include NDS, SNDS, ANDS, SANDS submissions, those directed to an NAS, First in Class drugs, Me Too drugs, and drugs approved via the two expedited review streams (Priority Review and NOC/c).

For the present purposes, "new" drugs were those that were either approved through the New Drug Submission (NDS) stream, contained a New Active Substance (NAS), or directed to First in Class (FIC) drugs. In con-

99. For discussion of lifecycle drugs, see generally Eichler et al., *supra* note 47; Yeates, et al., *supra* note 1.

100. *See* Sawicka & Bouchard, *supra* note 35, at 92.

trast, "follow-on" drugs were either brand name drugs approved through the Supplementary New Drug Submission (SNDS) stream, approved via the SNDS stream directed to FIC therapies, approved via the SNDS stream containing an NAS, or generic drugs approved via either the standard Abbreviated New Drug Submissions (ANDS) stream or the follow-on Supplementary ANDS (SANDS) stream. The classification system is summarized for convenience in Table 1.

Table 1: Classification Scheme for New and Follow-on Drugs

Firm Type	New Drugs	Follow-on Drugs
A. Brand Name	NDS	SNDS
	NDS FIC	SNDS FIC
	NDS NAS	SNDS NAS
B. Generic	-	ANDS
	-	SANDS

B. DRUG PATENTING

We also conducted a study on the relationship between drug approval, drug patenting, and drug litigation. This involved statistical analysis of patenting patterns associated with sixteen of the most profitable drug products in Canada.[101] We chose the top sixteen drugs for our initial study given that this cohort was likely to display the strongest patenting and patent listing patterns. This is because pharmaceutical companies have a vested interest in protecting the market on their most profitable drugs, and the primary means of doing so is via patenting. Each of the drugs studied under the patent analysis was approved in Canada between 2001 and 2008 and were analyzed as part of the drug approval data shown in Figures 1–3. Unlike the drug approval study, the drug patenting study was not restricted to a certain time period. This was necessary because many of the patents on drug products for which approval was granted during the 2001–2008 test period were filed and issued before 2001.

A detailed patent search of the Canadian Intellectual Property Office (CIPO) database was conducted for each drug approved and analyzed in Sa-

101. Andrew Humphreys, *MedAdNews 200—World's Best-Selling Medicines*, MEDADNEWS, July 2007. The drugs analyzed were: Lipitor,™ Advair,™ Plavix,™ Nexium,™ Norvasc,™ Zyprexa,™ Diovan,™ Risperdal,™ Effexor,™ Pantoloc,™ Singulair,™ Seroquel,™ Prevacid,™ Crestor,™ Prilosec,™ and Altace.™ Note that the list does not correspond literally to that in the United States. Rather, we chose for initial study, working backwards from number one, a group of 16 drugs that were on the U.S. list and which also had approval dates between 2001 and 2008 as identified in Sawicka & Bouchard, *supra* note 35.

wicka & Bouchard.[102] The CIPO search employed broad search terms for each drug in question with an effort to cast the widest possible net so that even patents with a remote possibility of being relevant would be returned by the search engine and made available for analysis and classification. The search was designed to return all patents owned by or assigned to the drug's manufacturer—including those owned by its parent company, subsidiaries, and partners—that made claims regarding the specific medicinal ingredients associated with the drug or claims regarding the general therapeutic class(es) to which the drug belongs. The patent search for each drug comprised two search strings: (a) a specific search string that returned patents likely to be relevant to the specific drug in question; and (b) a general search string that returned patents likely to be relevant to the general therapeutic class associated with the drug in question. Both are provided in Table 2.

Table 2: Search Strings for Data Collection and Analysis.

Search String	Boolean Operators
A. Specific Search String	((chemical name)<OR>(code name)<OR>(brand name)<OR>(chemical class)<OR> (chemical formula))<AND>(owners<IN> OWNER)<AND>(PAPD>=1867-07-01)<AND> (PAPD<=study start date)
B. General Search String	((*therapeutic class*)<OR>(*active site*))<AND><NOT>(*chemical name*)<AND><NOT> (*code name*)<AND><NOT> (*brand name*)<AND><NOT>(*chemical class*)<AND><NOT> (*chemical formula*)<AND>(*owners*<IN> OWNER)<AND>(PAPD>=1867-07-01)<AND> (PAPD<=*study start date*)

The specific search string used Boolean operators to return all patents owned by the drug manufacturer or its affiliates that mention either the drug's chemical name(s), code name(s), brand name(s), chemical class(es), or chemical formula(e), and have priority dates between the date of Canada's Confederation and the start date of the study. Databases such as CIPO and the Canadian Patent Register (CPR), their American counterparts, the U.S. Patent and Trademark Office (USPTO) and Orange Book (OB) databases, as well as secondary sources, were used to acquire an exhaustive list of all possible chemical names, codes names, brand names, and chemical classes associated with a particular drug. In determining the chemical formula, precedence was given to formulae expressed in patents found on CIPO and USPTO databases. The owners referred to within the search string refer not only to a drug's manufacturer but also to its possible parent company, sub-

102. *See* Sawicka & Bouchard, *supra* note 35.

sidiary and partner(s). This list of owners was cross-referenced using CIPO, CPR, USPTO, and OB databases as well as searches of case law and secondary sources where necessary.

The general search string used Boolean operators to return all patents owned by the drug manufacturer or its affiliates, not previously found by the specific search string, that mention the therapeutic class(es) to which the drug belongs or make specific reference to the drug's active site. The therapeutic class and active site of a drug are obtained by reference to CIPO, CPR, USPTO, and OB databases, and secondary sources such as company websites and Internet searches. These sources were used to acquire an exhaustive list of all possible chemical names, codes names, brand names, and chemical classes associated with a particular drug.

Combined, the two search strings return a broad list of potential patents owned or assigned to the Canadian manufacturer or its subsidiaries or partners. The legitimacy of the search terms was confirmed using Health Canada's drug approval data, as well as manufacturer, securities, and exchange websites, from which ownership histories were ascertained. Patents were individually inspected and pruned when deemed irrelevant to drugs in the study. The USPTO database, which provides a history of prior art, was also used as a means of cross-referencing patents for relevance. Relevant patents were sorted by priority date and cross-referenced with the patents registered on the CPR pursuant to linkage regulations.

C. PATENT LISTING AND LITIGATION

We quantified patents identified using the search method that were also listed on the Canadian Patent Register (CPR) under the NOC Regulations. Patents listed on the register can be litigated numerous times since they can be listed for multiple Drug Identification Numbers (DINs) under the NOC Regulations. Patent listing is a critical step to potential extension of patent monopolies for drugs coming off patent protection because generic firms must demonstrate in litigation that each patent on the list is either invalid or not infringed by the generic product to obtain market approval. For our purposes, only the date of first instance (the earliest date on which the patent was registered) for each patent was collected and analyzed.

In addition to analyzing patents listed on the patent register, we also investigated the case law pertaining to patents litigated under the NOC Regulations. We assessed the number and types of trials, the number of patents litigated in these trials, the number and types of legal decisions on listed and litigated patents (motions, trial and appellate decisions), whether listed patents were valid and infringed (brand name victory) or invalid and not in-

fringed (generic victory), and the theoretical and actual extension of patent monopolies via the operation of linkage regulations.

D. ANALYTICAL MODEL

The third element of the study was the synthesis of empirical data from approval, patenting, and litigation studies into an analytical model. The focus of the analysis is on the impact of regulatory incentives designed to facilitate breakthrough pharmaceutical innovation by providing strong IPR rights to firms. Throughout Part III, we compare data relating to the time courses of varying types of drug approvals with concomitant drug patenting, patent listing, and litigation data. Particular attention was given to the synchronization, if any, of approval, patenting, listing, and litigation data to the times for establishment of the NOC Regulations and proposal of the Progressive Lifecycle Framework, as both were intended to facilitate enhanced access to novel therapeutic products in exchange for strong IPR rights. However, since most of the data relate to the period before the Progressive Lifecycle Framework was fully integrated into Canadian law, the majority of the analysis relates to the linkage regulations. The analysis has been cast in terms of a complex adaptive innovation ecology in Section IV.D of the Article.

E. DATA ANALYSIS

Drug approval, patenting, patent listing, and litigation data were collected, statistically analyzed, and graphed as described previously[103] using a combination of Excel® (Microsoft. Corp., Redmond, WA), GraphPad Prism® (Graphpad Software Inc. La Jolla, CA), and SigmaPlot® (Systat Software, Inc. San Jose, CA). Legal decisions relating to listed and litigated patents were obtained using Quicklaw™ (Lexis Nexis®) and Westlaw® (Thomson Reuters®). Economic data relating to prescribed pharmaceuticals were obtained with permission from IMS Health Inc. (Canada) and from published reports from the Canadian Institute for Health Information (CIHI). Patent data were obtained from Canadian (CIPO) and U.S. (USPTO) patent databases.

III. RESULTS

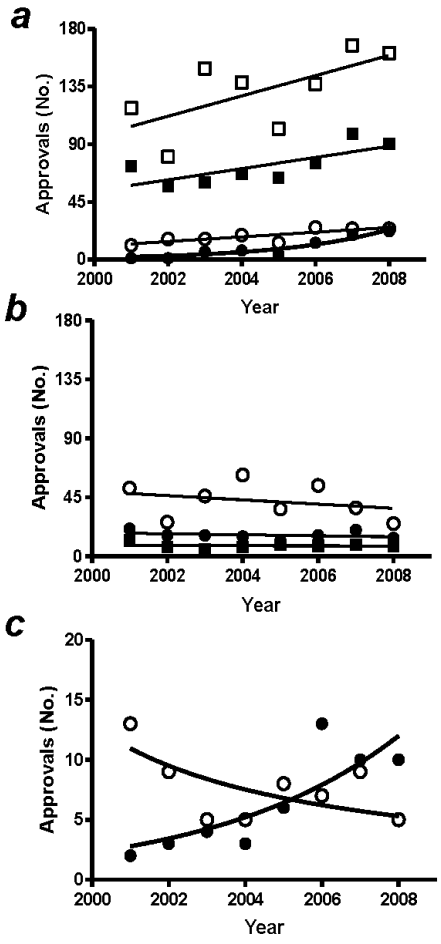
A. DRUG APPROVAL

To empirically investigate the relationship between drug regulation and innovative therapeutic product development, we first reviewed market authorizations for pharmaceuticals in Canada over the period from 2001–2008

103. *See id.*

(test period).[104] 2001 was taken as the starting point for analysis, as major amendments to the nation's food and drug legislation and regulations were made at that time which affected both the goals and mechanism of national drug regulation.[105] Market authorizations in Canada are referred to as Notices of Compliance (NOCs). We analyzed a total of 3,837 NOCs. Of these 45% were administrative in nature, e.g., product manufacturer or name change. This left 2,122 approvals for detailed analysis. These approvals were attached to 608 marketed drug products, amounting to an average of 3.5 approvals per product.

Figure 1: Shifting Patterns of Drug Approval and Drug Regulation During the Period 2001–2008



104. *See also* Bouchard & Sawicka, *supra* note 29; Sawicka & Bouchard, *supra* note 35.
105. Sawicka & Bouchard, *supra* note 35, at 107.

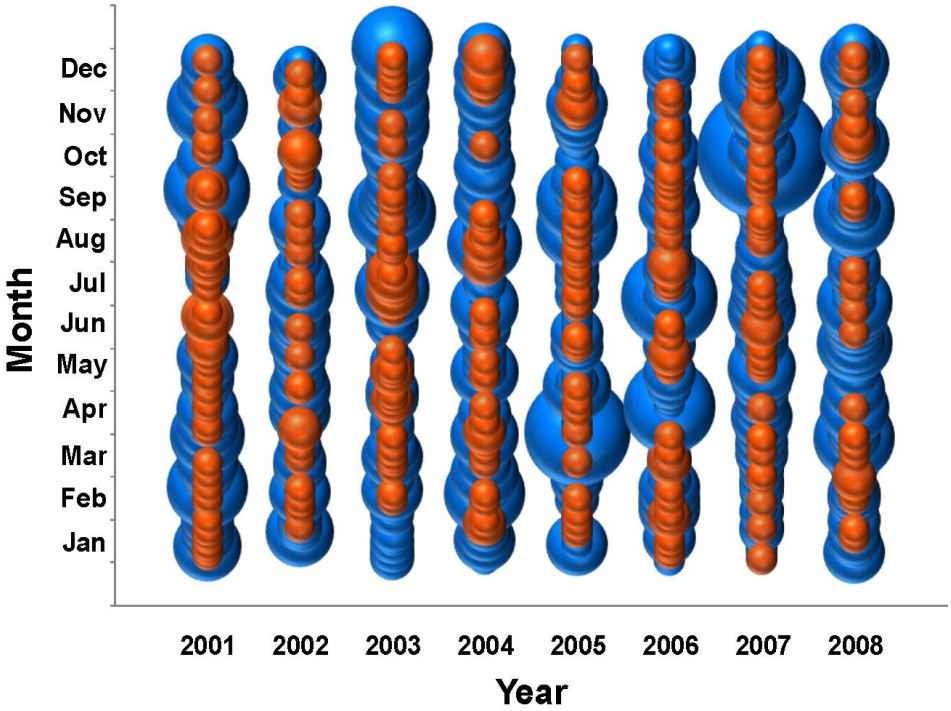
a. Market authorizations for several types of "follow-on" drug increased over the 2001–2008 test period. This includes brand name Supplemental New Drug Submission (SNDS: □) and SNDS First In Class (SNDS FIC: ●) approvals, and generic Abbreviated New Drug Submission (ANDS: ■) and follow-on Abbreviated New Drug Submission (SANDS: ○) approvals. ***b.*** In contrast, approvals granted to brand name firms for "new" drug submissions declined from a smaller baseline over the same period. This included approvals from New Drug Submission (NDS: ○), New Active Substance (NAS: ●) and NDS First In Class (NDS FIC: ■) streams. ***c.*** Expedited review pathway for drug approval is shifting towards probationary-type approval consistent with emerging lifecycle models of regulation: Expedited drug approvals with no post-market evidentiary obligations (Priority Review: ○) decreased over the 2001–2008 test period while those with significant post-market obligations conditions (NOC/c: ●) increased steeply over the same time frame.

Using the classification scheme described in Part II and summarized in Table 1, we found that follow-on drugs constituted the vast majority of drugs approved in the domestic market over the period 2001–2008. For example, in 2001 the total number of follow-on approvals was 2.39 times greater than that for total new drug approvals. This constituted 70.5% of all approvals in Canada over the test period and 65.33% of approvals granted to brand name pharmaceutical firms. As shown in Figure 1a, this trend intensified over the test period. By 2008, the number of follow-on approvals was 6.32 times greater than new drug approvals. This constituted 86.4% of all approvals in Canada and 86.02% of brand name approvals over the same time frame. Approvals directed to line extension drugs (SNDS: □) accounted for 34% in 2001, increasing to 47% in 2008. By comparison, the more innovative supplementary first in class drugs (SNDS FIC: ●) made up the smallest fraction of all follow-on approvals (5.4% in 2001). While the number of SNDS FIC approvals was small, it nevertheless increased sharply over time, from 1 in 2001 to 22 in 2008. As shown in Figure 1a, follow-on approvals granted to generic firms based on bioequivalence to previously marketed products also increased significantly over the test period. Both standard (ANDS: ■) and supplementary (SANDS: ○) generic approvals increased by 28.6% and 118.2%, respectively, over 2001 values. Therefore, all four categories of follow-on drugs increased over the test period.

Figure 1b demonstrates opposite trends for all new drug categories over the course of the test period, and that these changes took place from a smaller baseline. Approvals granted for all new drugs combined declined from 29.5% of total approvals in 2001 to 13.69% in 2008. Similarly, approvals granted to brand name pharmaceutical firms decreased from 34.67% of total approvals in 2001 to 13.44% 2008. These data represent a reduction of 55% and 48% in total approvals and approvals granted to brand name firms respectively over the eight year test period. As the regression lines illustrate, the approvals for all three new drug metrics (NDS: ○; NAS: ●; NDS FIC: ■)

declined steadily over the course of the test period.

Figure 2: Time Series Distribution for New and Follow-on Drug Approvals During the Period 2001–2008



Bubbles represent approvals granted per day for "new" (●) and "follow-on" (●) drugs as defined in the text accompanying Fig. 1. Bubble diameter is a linear representation of the number of approvals granted per day distributed over the course of the test period expressed yearly and monthly. The data illustrate that both new and follow-on drug approvals were well spaced out over the course of the test period rather than being aggregated in a given month or year, particularly when viewed over the course of the entire eight year test period.

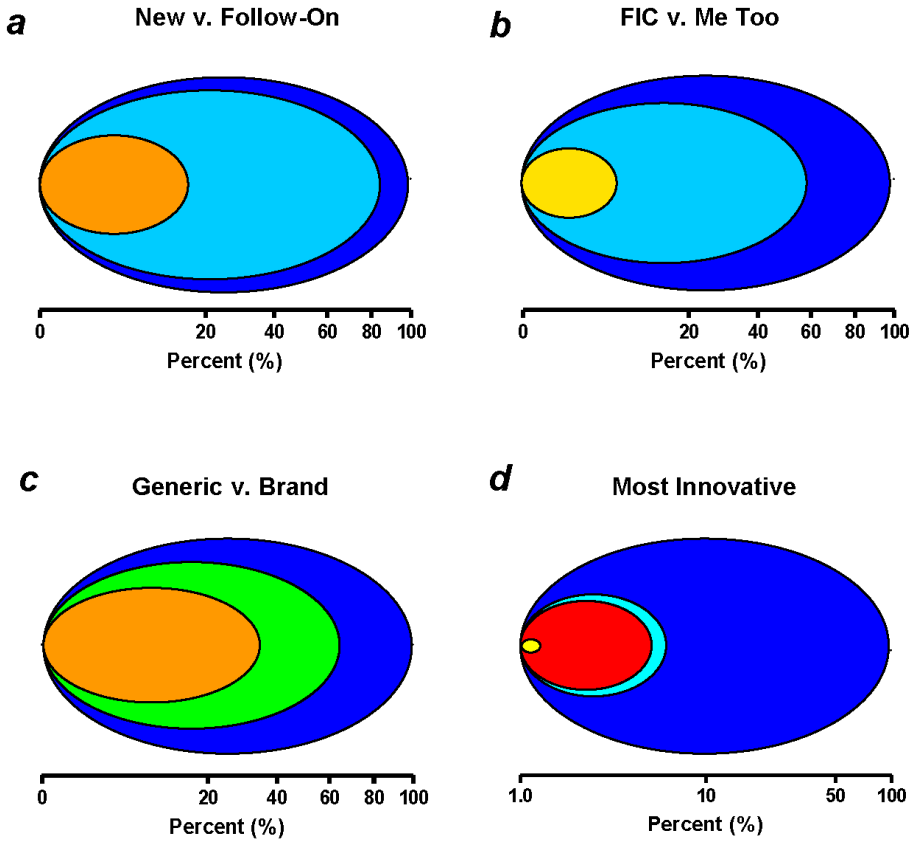
Time series plots for new and follow-on approvals are presented in Figure 2. Approval data for both classes of drugs are expressed as approvals per day plotted by month and year over the test period. The data illustrate that both new (●) and follow-on (●) drug approvals were well spaced out over the course of a given year rather than aggregated in a given month or year, particularly when viewed over the course of the entire test period. Therefore, there was no daily or monthly variation skewing yearly averages as discussed in relation to Figures 1 and 3. Comparative data for all new and follow-on approval categories in 2001 and 2008 are provided in Table 3.

Table 3: Comparison of 2001 and 2008 Drug Approval Data

Drug Type	2001		2008		Δ (%)
	N=	% Total	N=	% Total	
A. New Drugs	52	20.4%	25	83.3%	-51.9%
NDS	52	100.0%	25	100%	-51.9%
NDS FIC	12	23.1%	8	32.0%	-33.3%
NDS NAS	21	40.4%	14	56.0%	-33.3%
B. Follow-On Drugs	203	79.6%	275	91.7%	35.5%
SNDS	118	58.1%	161	58.5%	36.4%
SNDS FIC	1	0.5%	22	8.0%	2100.0%
ANDS	73	36.0%	90	32.7%	23.3%
SANDS	11	5.4%	24	8.7%	118.2%

Figures 3a–d are area diagrams illustrating *cumulative* approval data for various categories of new and follow-on drug products. As shown in Figure 3a, only 16% (n=338) of the 2,122 drugs approved over the period 2001–2008 were deemed to be "new" drugs. This refers to NDS submissions, including those directed to FIC therapies and those including an NAS. By contrast, 84% (n=1,784) of approved drugs were "follow-on" in nature, including brand name SNDS approvals, generic ANDS and SANDS approvals, and brand name SNDS approvals directed to FIC therapies.

Figure 3: Profile of Pharmaceutical Innovation in Canada between 2001–2008



a. New v. follow-on approvals. Of total drugs approved over the test period, 15% constituted New Drug Submissions (NDS: ●) while 84% were for "follow-on" drugs (SNDS, ANDS and SANDS: ●). ***b.*** Types of follow-on approvals. Of follow-on approvals, 6.1% were for supplementary "First in Class" (SNDS FIC: ●) drugs while 59% were for "Me-Too" drugs (●). ***c.*** Brand name v. generic approvals. Of all drugs approved during the test period, 65.5% of approvals were granted to brand name drug companies (NDS and SNDS: ●) and 34.5% to generic companies (ANDS and SANDS: ●). ***d.*** Most innovative drugs. While 6.5% of approvals during the test period were directed to New Active Substances (NAS: ●) and 5.3% of all NDS and SNDS submissions were approved under an expedited review process (Priority Review and NOC/c: ●), only 1.23% of all drugs approved over the period 2001–2008 were also directed to FIC therapies and contained an NAS (●). Areas are approximations of calculated means for the entire test period. Note that area scales are linear for panels ***a–c*** and log for panel ***d***.

Figure 3b shows the results of a more nuanced analysis of follow-on drugs, this time focusing on comparison of Me Too and FIC drugs. Of all drugs approved between 2001 and 2008, 59% (n=1,252) were Me Too. Of note, the fraction of Me Too drugs was substantially greater than all FIC

drugs, irrespective of whether they were NDS or SNDS (6.5%; n=138). The requirements for NDS and SNDS FIC and Me Too drugs are summarized for convenience in Table 4.

Table 4: Classification Scheme for First in Class and Me Too Drugs

Route	FIC	Me Too
A. NDS	New Chemical Form -or- New Use/Indication	Change in Benefit:Risk
B. SNDS	New Chemical Form -and- New Use/Indication	Change in Chemical Form -and- Change in Benefit:Risk

Generic drugs were the final follow-on category to be assessed. The split between total brand name and generic drugs approved from 2001 to 2008 is shown in Figure 3c. Of all drugs approved over the test period, 65.5% were directed to brand name products while the remaining 34.5% were directed to generic products.

Data for the most innovative drugs approved during the test period are given in Figure 3d. Only a small fraction (6.1%) of drugs approved (n=130) during the test period contained an NAS. Similarly, of 2,122 drugs approved, only 5.3% (n=112) went through the two expedited approval streams (Priority Review or NOC/c), and of these only a small number (n=26) were also directed to FIC therapies and contained an NAS. This amounted to 1.23% of total drug approvals over the eight-year test period and 1.87% of total brand name approvals over the same period. These results illustrate that the typical drug approved by Canadian regulators over the period 2001–2008 was most likely to be a drug approved via the SNDS stream rather than a new drug approved via either the NDS stream or either expedited stream (Priority Review or NOC/c). The likelihood that a drug approved during the test period satisfied the most stringent requirements for a breakthrough drug was close to zero (1.23%).

As discussed supra, there are two forms of expedited drug approval in Canada: "Priority Review" and approval via the "NOC with conditions" (NOC/c) pathway.[106] Priority Review allows appropriate candidates to be shifted forward in the approval queue without a change in evidentiary requirements for safety and efficacy required for conventional NDS approval. Drug candidates must be directed to treatment of a serious, life-threatening,

106. *Id.* at 87.

or severely debilitating disease with an unmet medical need or for which a substantial improvement in the benefit-risk profile is demonstrated.[107] By contrast, the NOC/c pathway allows a drug to gain market access prior to completion of traditional Phase 3 clinical trials, provided that it is directed to a serious, life-threatening, or severely debilitating disease for which no drug is marketed or where the candidate presents a better overall benefit-risk profile than existing therapies. Unlike the Priority Review stream, continuing approval via the NOC/c stream is contingent upon whether pharmaceutical sponsors meet the conditions assigned to the NOC/c. For this reason, NOC/c approval is a reasonable proxy for emerging lifecycle models of drug regulation.[108]

Data in Figure 1c suggest that Canadian regulators may be shifting away from Priority Review as the dominant mechanism for expedited review towards the NOC/c pathway. Priority Review approvals (○) decreased from 14 in 2001 to a low of 6 in 2008, declining 57% over the eight year test period. By comparison, the number of NOC/c approvals (●) escalated sharply over time, from a minimum of 2 in 2001 to a maximum of 13 in 2006 (stabilizing at 10 in 2007–2008). Compared to the 57% decline in the number of Priority Review approvals, peak NOC/c approvals increased by 650%. The totals for both streams over the test period were not dissimilar; 61 and 51 for Priority Review and NOC/c, respectively. However, as illustrated by the data and fits in Figure 1c, the trends for the two pathways crossed in 2005.

Of interest, the legal basis for Priority Review and NOC/c approval are not expressly provided for under the current Food and Drugs Act[109] or regulations.[110] Rather, both are grounded in administrative instruments known as "guidance documents" that do not have the force of law.[111] Data described in Figure 1c therefore demonstrates that Canadian regulators are already "anticipating" the lifecycle regulatory framework proposed in Bill C-51,[112] along with its recalibrated balance of pre-market and post-market access, safety, and efficacy. Together, the data in Figures 1–3 suggest that Canadian regulators are focusing on faster approval with enhanced post-market surveillance, whiles approval is geared more towards follow-on rather than towards breakthrough drug development.

---

107. HEALTH CANADA, GUIDANCE FOR INDUSTRY, *supra* note 30, at 1–2.

108. For a discussion of this issue see Section I.B.1 of this Article and Bouchard & Sawicka, *supra* note 29, at 105–06.

109. *See* Food and Drugs Act, R.S.C., ch. F-27 (1985).

110. *See* Food and Drug Regulations, C.R.C., ch. 870 (2009).

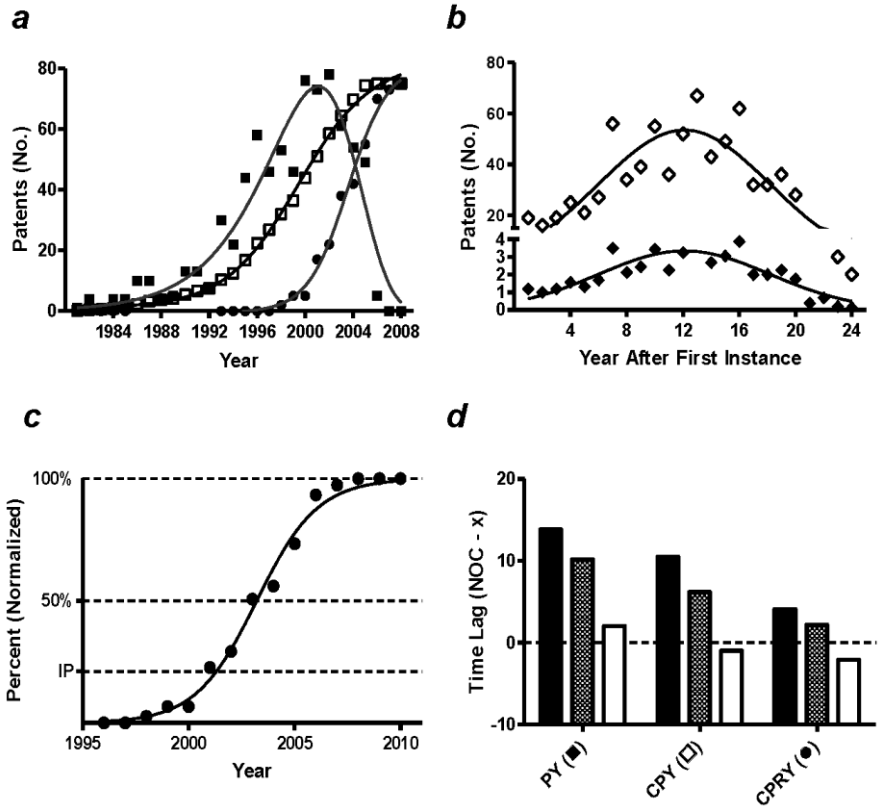
111. *See* Bouchard & Sawicka, *supra* note 29, at 52.

112. Sawicka & Bouchard, *supra* note 35, at 117.

B. DRUG PATENTING

Figure 4 shows data relating to drug patenting and patent listing of drugs approved for sale in Canada during the period 2001–2008. The data are for 16 of the most profitable drugs sold in Canada for which an NOC was granted during the test period (approval subset). The list parallels the top 16 drugs sold in the United States during the same period.[113]

Figure 4: Patenting and Patent Listing Patterns Associated with Drug Approval



a. Total patents issued by year associated with a sub-set of sixteen top selling drugs (■); cumulative number of patents associated with the sub-set (□); and cumulative number of patents listed on the patent register under linkage regulations associated with the sub-set (●). Note the strong convergence of total and listed patents over the course of the test period. ***b.*** Total (◇) and average (◆) number of patents on approved drugs within the subset plotted as a function of the time after the priority date on which the first patent on the subset was issued. ***c.*** Method used to calculate the temporal gap between the date of mean drug approval on the patent subset (2005) and the inflection point (IP), 50th, and 100th percentile of normalized maximum drug patenting and approvals. Data are from the cumulative number of patents (●) above. ***d.*** Graph expressing the temporal relationship between drug approval

113. Humphreys, *supra* note 101. For more information, see Part I.

and the IP, 50th, and 100th percentile of maximal normalized patents granted per year (PY), cumulative patents per year (CPY), and cumulative patents listed on the patent register per year (CPRY). Time points are calculated as the difference between the date of drug approval (NOC) and the date of the IP, 50th, and 100th percentile (NOC-x). The data suggest that drug patent listing may be a better proxy for drug approval than drug patenting.

As illustrated in Figure 4a, total patents granted on the approval subset had a bell-shaped distribution (Gaussian; R^2=0.91), peaking in 2001 (■). There were 772 patents on 16 products, corresponding to an average patent per product ratio of 48:1. The calculated inflection point, representative of the take-off point from baseline, for total patents issued yearly occurred about 1991 (1991.35). This was just before the linkage regulations came into force in 1993. That the inflection point preceded the NOC Regulations is not surprising in light of the significant negotiations leading up to TRIPS and the coming into force of the linkage regulations regime. Cumulative patents for the subset rose over time in a manner that was well described by a sigmoidal function (□; R^2=0.99), peaking at about 2004. The calculated inflection point (1994.70) was slightly later than that calculated for total patents, occurring just after the linkage regulations came into force. Figure 4b (top) gives the same data re-plotted as a function of the year after the first patent was issued. Patents on approved drugs were granted over a relatively long term of 25 years (◇), peaking at 77 patents per year on the 12th year after the first patent was granted. As illustrated in Figure 4b (◆), this amounted to an average of 3.34 patents per product per year.

C. PATENT LISTING AND LITIGATION

Over the last decade, there have been increasing claims to the effect that the linkage regulation regime is used as more of a sword than a shield by brand name pharmaceutical firms.[114] Figure 4a illustrates the manner in which

114. *See generally* ROY J. ROMANOW, COMM'N ON THE FUTURE OF HEATH CARE IN CAN., BUILDING ON VALUES: THE FUTURE OF HEALTH CARE IN CANADA: FINAL REPORT 208–210 (2002), *available at* http://dsp-psd.pwgsc.gc.ca/Collection/CP32-85-2002E.pdf (discussing the negative impact of evergreening by pharmaceutical corporations to extend the life of their patents and the costs associated with preemptory litigation on patent protection disputes); Bouchard, *PHOSITA*, *supra* note 73, at 46–52 (arguing that the procedures associated with linkage regime have led to unoriginal line extension patents resulting in reduced competition between firms and restricted consumer access to essential medications); Bouchard, *Scientific Research*, *supra* note 73 (arguing that that obviousness analysis by courts in NOC litigation allows pharmaceutical corporations to maintain a monopoly over their patented chemical compounds by setting a low bar for what is "nonobvious"); Caffrey & Rotter, *supra* note 84 (discussing the use of the Hatch-Waxman Act procedures by pharmaceutical corporations to maintain elevated prices, and the needs for reform that favors consumer interests); Hore, *supra* note 73, at 8–10 (discussing that NOC regulations lead to extended litigation between pharmaceutical corporations and potential generic produces that often leave infringe-

patents for the approval subset were listed on the patent register over the test period. The time course for cumulative listed patents (●) was well described by a sigmoid function (R^2=0.99), with a relatively steep slope, an inflection point near 2001 (2001.10), and an apparent peak in 2008. Importantly, the curves for cumulative patents (□) and the fraction of patents that were listed on the patent register (●) converged strongly over time. This result supports the conclusion that brand name firms are listing patents they obtain on the patent register in a timely fashion in order to delay generic entry.[115]

Of 772 patents granted on the approval subset, 77 were listed on the patent register between 1998 and 2008. On average there were 4.81 listed patents per product. As indicated by the difference between the average number of patents per year (3.34) and the average number of listed patents per product (4.81), domestic linkage regulations allow patents to be listed on more than one product. Unlike drug patenting, which occurs only in an anterograde direction (i.e., patents must be new, non-obvious, and have utility over the prior art), patents may be listed on the patent register in either an anterograde or retrograde direction. For example, originating patents relating to proton pump blockers may be listed not just for first generation product Losec® (racemic mixture of R and S omeprazole Mg^{2+}) but also the second generation product Nexium® (S enantiomer, esomeprazole Mg^{2+}), and vice versa.

We next investigated the temporal relationship between NOC grant, patent issue and patent listing. From each of the curves in Figure 4a, we calculated three values: (a) the inflection point at which the data deviated most strongly from baseline values (closed bars), and the point at which each curve reached the (b) 50th (hatched bars) and (c) 100th percentile (open bars) of normalized maximum values. The inflection point was calculated as the zero point of the second derivative of the data trendlines. Each of the three values was then plotted as a function of the average date on which the subset received marketing approval (2005). This was done to obtain a measure of the delay between drug approval and drug patenting and listing. The procedure is demonstrated for cumulative patent listing data in Figure 4c (●).

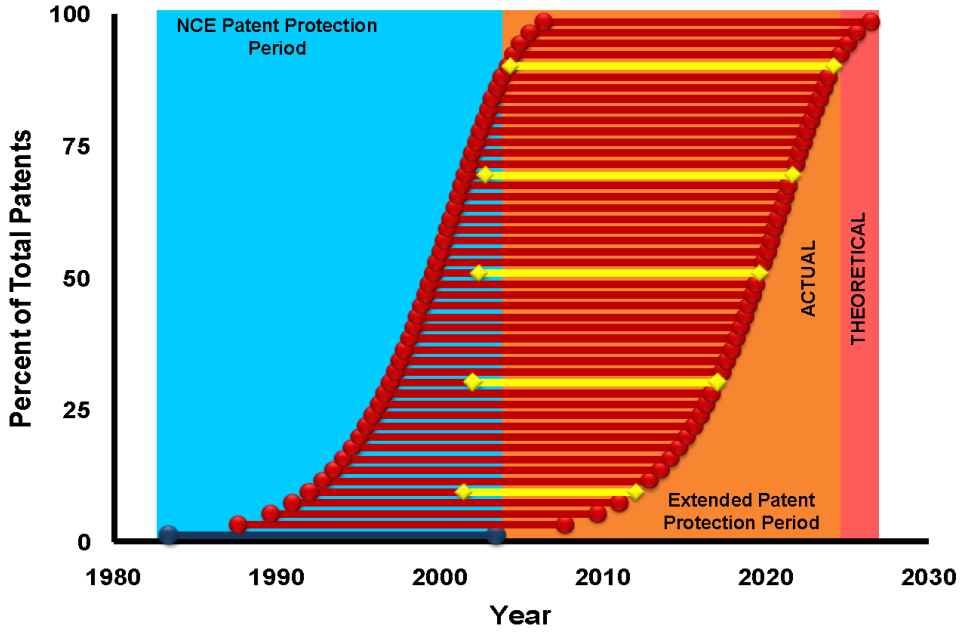
---

ment claims unresolved); Jaffe, *supra* note 94.

115. AstraZeneca Can. Inc. v. Canada, [2006] 2 S.C.R. 560, ¶ 29 (Can.) (discussing the ambiguity of NOC Regulations and that requirements for filing prohibition motions could be interpreted so as to force generic producers "to address new patents as fast as [pharmaceutical corporations] could have them added to the . . . list"); *see generally* Bouchard, *PHOSITA*, *supra* note 73, at 50 (discussing how the low standard for "incremental line extension" patents allows the corporations to register patents for uninventive products after minimal investment, keeping generic products off the market for longer).

As illustrated by the bar graphs in Figure 4d, there was a significant lag between the date on which NOCs were granted and the dates on which patents on the same drug product were granted. This pattern was observed independent of whether patents were expressed by year of grant (Patent per Year, or PY) or cumulatively (Cumulative Patents per Year, or CPY). This is not surprising in light of the regulatory lag between drug patenting and drug approval. The data were different however for patent listing (Cumulative Patents Registered per Year, or CPRY). As shown in Figure 4d, average data for both the inflection point and 50th percentile exceeded the null point by only 4 and 2 years, respectively. This can be compared with 10 and 8 years for corresponding data for cumulative patents (CPY). The lag between drug approval and patent listing was even greater for patenting data expressed as a function of year of grant (PY). Of interest, the calculated values for the 50th percentile and peak patent listing for CPRY were 1–2 years on either side of the null point. This result indicates there was virtually no significant lag between drug approval and patent listing as the test period progressed. While the data obtained do not provide conclusive evidence for a causal relationship between drug regulation and drug development, they demonstrate that patent listing is a substantially better proxy for drug approval than drug patenting.

Figure 5: Extension of Patent Monopoly for Marketed Drugs via Operation of Linkage Regulations



Period of extended patent protection for averaged drugs in the subset (n=16). Left and right sigmoid curves represent cumulative patent protection start and end dates. The term of pa-

tent protection was deemed to begin on the priority date. Terms are shown for the "originating patent" on the New Active Substance/New Chemical Entity (●; n=1) and all "subsequent patents" (●; n=21). The date on which patents were listed on the register is also shown (◆; n=5). The duration of theoretical and actual patent protection under linkage regulations associated with originating and subsequent patents are illustrated by representative horizontal lines and shading along the time axis. Note the period of patent protection associated with originating patents lasted about 20 years (■), from 1983–2003. In comparison, the duration of extended patent protection associated with all "subsequent patents" was much longer (■), lasting from about 1987 to 2028. Of the 48 patents granted per drug, an average of 5 were listed on the patent register. The term of protection associated with these patents ran from 1993–2025 (■). This yielded an actual extended period of patent protection of 22 years beyond that afforded by the originating patent. Note that due to strategic listing of patents on the patent register (◆), there was little difference between theoretical and actual patent protection under linkage regulations.

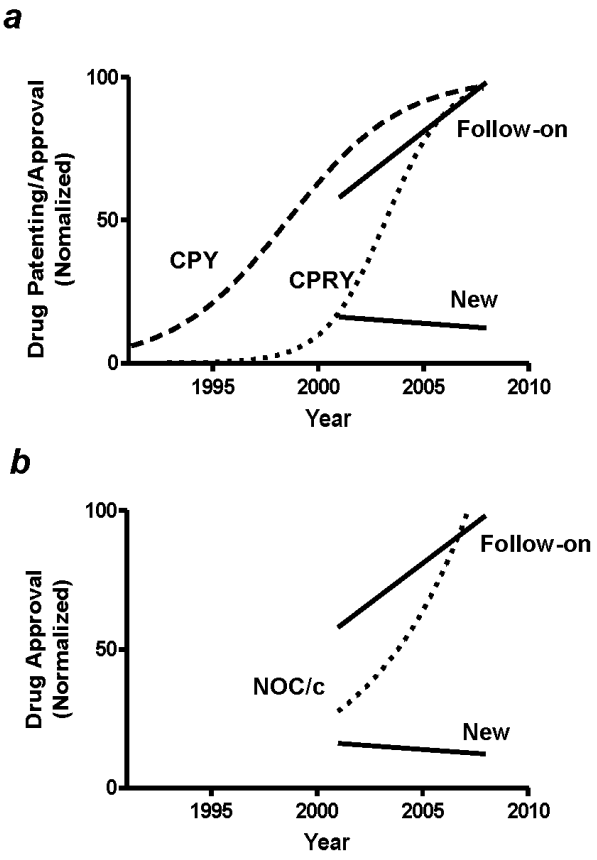
Given the results in Figure 4, we further probed the nexus between drug approval, drug patenting, and patent listing, particularly as it relates to potential extension of the term of patent protection afforded to drugs that are already approved in Canada. Figure 5 shows a comparison of potential and actual periods of extended patent protection for the average drug product in the approval subset due to operation of linkage regulations. Here "potential" is used to refer to the hypothetical extension of patent protection under patent legislation and linkage regulations if all patents granted were in fact listed on the patent register. In comparison, the "actual" term of extended patent protection refers to the extension of the duration of patent protection beyond that afforded by the originating patent alone as a result of those patents actually registered on the patent list. The sigmoid curves represent the start and end dates for the potential term of patent protection as a function of patents associated with approved drugs. The term starts with the priority date of the "originating patent," e.g., the first patent on the drug, typically that on the NAS/NCE (●), and ends 20 years from the filing date of the originating patent plus the cumulative terms of all "subsequent patents" (●) associated with the marketed drug. This is illustrated by the corresponding horizontal lines and shading in Figure 5. Patents actually listed on the register are represented by appropriate symbols (◆) and horizontal patent term lines.

The average period of patent protection associated with originating patents was about 20 years, from 1983–2003. This represents an average of patent terms before (17 years from date of grant) and after (20 years from filing date) amendments made to the Patent Act pursuant to TRIPS. In comparison, the duration of potential extended patent protection associated with subsequent patents was about 2-fold longer, lasting from about 1987 to 2026. This yields a term of extended patent protection due to operation of linkage regulations of about 43 years per drug on average. However, this calculation

does not reflect the actual period of extended patent protection, which would only be a function of cumulative terms for patents actually listed under the NOC Regulations on the register. Of the average of 48 patents per product on the approval subset, 10% (4.81 patents per product) were actually listed on the register.

Termination of listed patents was spaced fairly evenly between 2010 and 2025 rather than being clumped together at the front end of the data set. The even distribution of listing resulted in the extension of the term of patent protection from the end of the NAS patent in 2003 to termination of the latest listed patent in 2025. The extension of the average patent term owing to linkage regulations amounts to an increase of 22 years, representing a doubling of the duration of patent protection beyond that associated with the originating patent. As illustrated by the appropriate symbols (◆) and shading in the figure, there was little difference between potential and actual terms of patent extension for the approval subset under the NOC Regulations. This was due to strategic timing of patent listing by brand name drug firms e.g., firms stagger the registration of their strongest patents to obtain the longest period of protection. Of note, comparison of data in Figures 3 and 5 demonstrate that while the average drug approved in Canada over the test period has an arguably low innovative value, the average period of patent protection afforded to products in the approval subset is in fact quite substantial.

Figure 6: Comparison of the Timing of Trends for Drug Innovation, Lifecycle Regulation, Patent Grant and Patent Listing



a. Overlay of time courses of fits to normalized cumulative patents per year (CPY; long dash), cumulative listed patents per year (CPRY; short dash), new drug approvals (New; downward linear) and follow-on drug approvals (Follow-on; upward linear). Data for "new" and "follow-on" innovations were calculated from NDS and SNDS/ANDS/SANDS curves in Figs. 1a and 1b. Drug patenting and listing data are from Fig. 4a. ***b.*** Overlay of new drug approvals and follow-on drug approvals from panel *a* and life-cycle-based NOC/c approvals (NOC/c; short dash). Data for expedited review were taken from Fig. 1c. Comparison of these curves suggests that steep time-dependent changes in patent grant, patent listing or NOC/c approval as proxy for lifecycle regulation appear to be poorly correlated with, and thus to provide poor incentives for, breakthrough drug development as measured by new drug approval data.

The importance of the timing of shifts in innovation profiles, expedited drug approval, drug patenting, and patent listing is underscored by the data in Figure 6. In this analysis, drug patenting and listing represent patent incentives for innovation, whereas expedited drug approval is taken as a measure of lifecycle-based regulatory incentives for innovation. The data for "new"

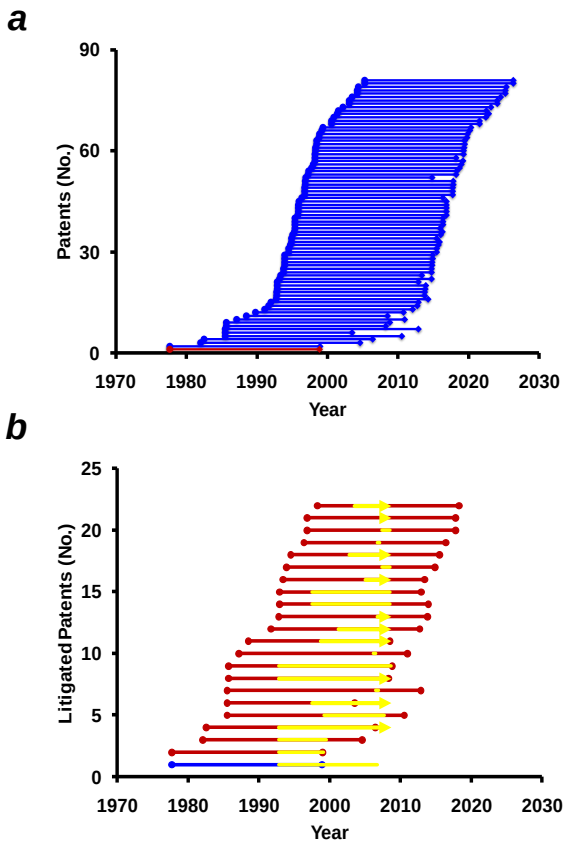
and "follow-on" innovations represent the fits to NDS data and SNDS, ANDS, and SANDS data from Figures 1a and 1b. Data for lifecycle regulation were taken as the fit to NOC/c data from Figure 1c. Drug patenting and patent listing curves are those from Figure 4a.

A comparison of fitted curves in Figure 6 indicates that neither the steep time-dependent changes in patent grant and patent listing preceding (Figure 6a) nor the NOC/c approval in the midst of relatively linear trends for new and follow-on drugs (Figure 6b) appear to provide a measurable incentive for pioneering drug development, at least as reflected by the data and fits to new and follow-on drugs. That these trends (1) occurred before and during the comparatively linear changes in new and follow-on drug approval by regulators (Figures 1 and 6) and (2) were observed independent of the temporal association of drug approval, drug patenting, and patent listing (Figure 4d) suggests that the current basket of IPR rights targets provides a much stronger incentive for follow-on rather than pioneering drug development. The results demonstrate that pharmaceutical firms, when they so desire, are capable of responding rapidly and strongly to regulatory incentives in the context of drug regulation, but that this responsiveness has not extended to increasing the production of new and innovative drugs.

Finally, there has been sharp criticism of the practice of "evergreening" drug products via linkage regulations in the United States and Canada.[116] Evergreening refers to extending the market monopoly on a drug facing originating patent expiration through listing of further relevant patents on the patent register for minor modifications to the marketed drug. An example of this phenomenon from our data set is presented in Figure 7.

116. Pfizer, Inc. v. Apotex, Inc., 480 F.3d 1348 (Fed. Cir. 2007); Bristol-Myers Squibb Co. v. Canada, [2005] 1 S.C.R. 533 (Can.); Apotex Inc. v. Merck Frosst Canada Inc., [1998] 2 S.C.R 193 (Can.), ¶ 33; *see also* Caffrey & Rotter, *supra* note 84.

Figure 7: Example of Extension of Patent Monopoly for Omeprazole



a. Relative to the forms of drugs marketed between 2001–2008, 82 patents were granted in relation to Losec® and Nexium.® As observed for averaged data on the subset of 16 drugs, the timing of the grant and the duration of cumulative patents followed a sigmoidal course, with patent protection beginning in 1978 and extending to about 2025. The first regulatory approval for omeprazole was in 1989. Data for the first New Chemical Entity (●) and all subsequent patents (●) are provided. ***b.*** Of the 82 patents granted on the two drugs, 22 were listed on the patent register and litigated under linkage regulations. Priority dates and patent terms are represented by appropriate symbols (●) and lines. Initiation, duration and termination of litigation on individual patents are represented by orange lines. Completely solid orange lines represent completed litigation. Right-facing arrows (➞) represent litigation which is still ongoing e.g., where it has not yet been determined that the listed patent was valid and infringed (brand name victory) or invalid and not infringed (generic victory). For details of individual trials, see text.

Omeprazole, marketed in Canada as Losec® (Prilosec® in the United States) and the second generation product Nexium® are widely considered to be two of the most profitable drugs developed over the last several decades.

Not surprisingly, they have also been the subject of prolonged and highly contentious litigation in both the United States and Canada. The chemistry and mechanism of action of both drugs is highly similar. Indeed, as illustrated in Table 5, their chemical names and formulae are almost identical. The difference between the compounds, as alleged in litigation in both jurisdictions,[117] is that the magnesium salt form of omeprazole (Losec®) undergoes a chemical shift following ingestion that converts a portion of the racemic mixture that is potentially inactive to the fully active chemical form (Nexium®).[118] This chiral shift has been claimed to double the effective drug concentration.[119]

Table 5: Comparison of Omeprazole (Losec®) and Esomeprazole (Nexium®)

Band-Name	Formula	Chemical Name
Losec®	$C_{17}H_{19}N_3O_3S$	6-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl) methylsulfinyl)-1*H*-benzo[*d*]imidazole
Nexium®	$C_{17}H_{19}N_3O_3S$	(*S*)-5-methoxy-2-[(4-methoxy-3,5-dimethylpyridin-2-yl) methylsulfinyl]-*H*-benzoimidazole

Setting aside the scientific veracity of this claim for the moment, the question arises of how pharmaceutical firms are able to strategically employ minor, but potentially significant changes to already patented and marketed compounds in order to maintain market share, through either "blocking pa-

117. No final trial or appeal decisions relating to omeprazole enantiomers have been released to date. For a notation of the seven ongoing applications under the NOC Regulation pertaining to esomeprazole (Nexium®); *see* AstraZeneca Can. Inc. v. Apotex Inc., [2008] F.C. 537 (Can.). There are a number of Canadian and U.S. appeal decisions regarding enantiomers under the NOC Regulations. *See, e.g.*, AstraZeneca AB v. Ranbaxy Pharms., Inc., 2008 U.S. Dist. LEXIS 102097 (D.N.J. Dec. 15, 2008); Dr. Reddy's Labs, Ltd., v. AstraZeneca AB, No. 08-2496, 2008 U.S. Dist. LEXIS 66176 (D.N.J. Aug. 28, 2008); Ivax Pharms., Inc. v. AstraZeneca AB, No. 08-2165, 2008 U.S. Dist. LEXIS 66177 (D.N.J. Aug. 28, 2008); AstraZeneca v. Ranbaxy Pharms., Inc., No. 05-5553, 2008 U.S. Dist. LEXIS 6337 (D.N.J. Jan. 25, 2008); Pfizer Canada Inc. v. Minister of Health and Ranbaxy Laboratories Ltd. [2008] F.C.A. 108; Janssen-Ortho Inc. v. Novopharm Ltd., [2007] F.C. 809 (Can.); Apotex Inc. v. Sanofi-Synthelabo Can. Inc., [2006] F.C.J. 1945 (Can.). For additional judicial consideration of the anti-competitive and/or fraudulent nature of such patenting and marketing strategies, see: Walgreen Co. v. AstraZeneca Pharm. L.P., 534 F. Supp. 2d 146 (D.D.C. 2008); Pa. Employees Benefit Trust Fund v. Zeneca Inc., 499 F.3d 239 (3d Cir. 2007).

118. T. Lind et al., *Esomeprazole Provides Improved Acid Control vs. Omeprazole in Patients With Symptoms of Gastro-oesophageal Reflux Disease*, 14 ALIMENTARY PHARMACOLOGY & THERAPEUTICS 861 (2000). For a review of chirality in sulphur compounds, see Ronald Bentley, *Role of Sulfur Chirality in the Chemical Processes of Biology*, 34 CHEM. SOC. REV. 609 (2005).

119. Bentley, *supra* note 118.

tents" (inactive patents that nevertheless serve as a barrier to market entry) or via patents that are listed on the patent register specifically in order to deter or initiate litigation.

We identified 82 patents, granted over a period of 20 years, associated with the two drugs. Together, the patents had a 50 year cumulative term of patent protection. As shown in Figure 7a, the time course and duration of patent protection were sigmoidal, similar to the averaged data in Figure 5a. Data are given for the first NCE patent (●) and all subsequent patents (●) identified using the methodology described supra. The priority dates for the first and final patent were 1978 and 2005, respectively. Therefore, the period of hypothetical patent protection on the omeprazole group ran from 1975 to about 2025. In comparison, the first NOC for omeprazole (Losec®) was granted on June 13, 1989, yielding a regulatory gap of close to 10 years. Of 82 patents that were deemed relevant to omeprazole, 22, or 27% of all relevant patents, were listed on the patent register. If not listed on the register at some future point in time, the remaining 73% were deemed to function as blocking patents or fodder for future patent listing efforts. As noted supra, patents could be listed on more than one drug product provided they are legally relevant to the marketed product. This is reflected in the fact that the 22 patents were the subject of 75 individual legal determinations (many of which aggregated in a single trial).

All 22 listed patents have or continue to be disputed at trial in some form or another. This is shown in Figure 7b, which illustrates listed and litigated patents (●) and the timing and duration of ongoing (➞) and final (—) litigation. Litigation over 15 of the 22 patents lasted in excess of 2 years, with 14 final trial decisions to date. (Some patents were litigated multiple times, as indicated infra.) Final decisions were at the Federal Court of Canada, Federal Court of Appeal, or Supreme Court of Canada. As might be expected with so many patents being litigated multiple times, decisions on the merits were not harmonious from one decision to the next.[120] Indeed, there were numerous instances (n=11) where a court at one level decided that patents were invalid or not infringed with one set of litigants, while a different court at the same level decided that the same patents were valid and infringed with different litigants. In addition to litigation under the NOC Regulations, there were also 3 related patent infringement actions involving listed patents, one of which is ongoing (data not shown).

Figure 7 does not include data relating to individual trials. While this would have provided a better sense of just how extensive the litigation was

120. Bouchard, *PHOSITA*, *supra* note 73.

over these two drugs, it would have complicated the figure unnecessarily. For example, over the period 1993–2009, there were 61 separate trials on 22 listed patents, including 310 motions (mean=5.08 per trial) and 25 final trial decisions. Of final decisions, 14 were appealed to the Federal Court of Appeal and 8 went on to the Supreme Court of Canada. Litigation occurred over a term of sixteen years, essentially from the time the linkage regulations came into force in 1993 until the present. Four trials on 12 patents are currently ongoing.

Given that the NDS patent expired in 1999, extended patent protection on omeprazole has been ongoing for at least 10 years. But this does not necessarily equate to a decade of prohibited generic entry under the linkage regulations, owing to the requirement that generics must first obtain approval for market entry themselves *and* demonstrate in litigation that all relevant patents are invalid or not infringed by their product.[121] Of 25 final decisions levied by the courts, there were 13 cases where patents listed on the register were judged to be invalid or not infringed. The average date of the first automatic injunction for all litigants was February 2001. This represented the date on which drug approvals granted, or to be granted, to generic firms were "put on hold." The average date on which the group of 13 trials ended, and thus the date of "reactivation" of average generic approval was December 2003. Therefore, litigation over patents relating to Losec® and Nexium® resulted in a delay of market entry of close to three (2.83) years for the group. According to IMS Health,[122] sales of the two drugs in drugstores and hospitals over the same time frame were CN $1.4 billion. In comparison, total spending on prescription pharmaceuticals rose from CN $11.7 billion in 2001 to CN $17.97 billion in 2004,[123] representing an increase of 92%. This includes an increase in out-of-pocket consumer spending from CN $2.56 billion to CN $3.36 bil-

121. Merck Frosst Canada Ltd. v. Apotex Inc., [2009] F.C.A. 187 (Can.); Apotex Inc. v. Merck Frosst Canada Inc., [1998] 2 S.C.R 193 (Can.).

122. INTERCONTINENTAL MARKETING SERVICES HEALTH INC., 2003 ANNUAL REPORT TO SHAREHOLDERS (2003), http://www.imshealth.com/portal/site/imshealth; INTERCONTINENTAL MARKETING SERVICES HEALTH INC., 2002 ANNUAL REPORT TO SHAREHOLDERS (2002), http://media.corporate-ir.net/media_files/NYS/RX/reports/ar2002.pdf; INTERCONTINENTAL MARKETING SERVICES HEALTH INC., 2001 ANNUAL REPORT TO SHAREHOLDERS (2001), http://www.imshealth.com/portal/site/imshealth; *see also*, INTERCONTINENTAL MARKETING SERVICES HEALTH INC., CANADIAN PHARMACEUTICAL INDUSTRY REVIEW 2001 (on file with author); INTERCONTINENTAL MARKETING SERVICES HEALTH INC., CANADIAN PHARMACEUTICAL INDUSTRY REVIEW 2002 (on file with author); INTERCONTINENTAL MARKETING SERVICES HEALTH INC., CANADIAN PHARMACEUTICAL INDUSTRY REVIEW 2003 (on file with author).

123. CAN. INST. FOR HEALTH INFO., *supra* note 8. Note, the 2008 value was forecasted by the report, and is not an actual value.

lion. It is reasonable however to speculate that 'but for' the existence of the linkage regime that generic entry may have occurred closer to expiry of the NCE patent, with an accordingly shorter period of delayed entry. Either way, the linkage regulations regime has proved to be a highly effective mechanism for extending market monopolies on profitable pharmaceuticals.

D. LIMITATIONS

The strength of conclusions from our pilot study is tempered by two limiting factors. The first is that the time frame for the drug approval study is smaller (2001–2008) than that for the patenting (1978–2008) and patent listing (1993–2008) studies. This owes to the fact that our initial work on drug approval was done prior to undertaking the patent study. The year 2001 was chosen as our starting point in the drug approval study because substantial amendments to Canadian drug regulation were made at this time that affected both the mechanisms and speed of approval.[124] It will therefore be important for future work to include approval data from before the domestic linkage regulations came into force in 1993.

The second, and related, limitation is that the approval data set was for 608 drugs while our pilot study on drug patenting and linkage regulations was only for 16 drugs. For reasons given in Part II: Methods, this made good sense for the pilot study. We attempted to extrapolate the approval data back in time. However, given the yearly scatter in the data set and resulting confidence levels, this was not feasible. We obtain some assurance from the consistent nature of the daily and monthly scatter of approval data described in Figure 2. More importantly, we have now increased our data set to 95 of 608 approved drugs between 2001 and 2008 in a follow-up study. The results (data not shown) indicate that all major patterns for drug patenting and patent listing shown in Figures 4–7 are repeated not just for the entire 'most profitable' data set, but for three different sub-groups (most profitable, n=33; NOC/c, n=22; Priority Review, n=46). In particular, there was no substantial difference in the patenting data in Figures 4a–4d (n=16) and the 2-fold larger data set of most profitable drugs in the expanded study (n=33). Even so, future research must complete the patenting data for not only the 608 drugs approved during the period 2001–2008, but also for the broader drug ap-

124. HEALTH CANADA, STAKEHOLDER WORKSHOP, *supra* note 71, at 6 (explaining that the objectives of the 2001 regulations were to "[s]horten application review times without endangering health and safety; [i]mprove safety mechanisms for research subjects; [r]egulator to be more involved in clinical trial monitoring and follow-up; [r]emove obstacles to additional R&D; [i]mprove access to innovative therapies and advice from Canadian physicians with research experience").

proval data as it grows to encompass and back-date the coming into force of the linkage regulations in 1993 and beyond.

IV. DISCUSSION

A. TRENDS IN GLOBAL DRUG DEVELOPMENT

Data in Figure 1 demonstrate that the number of "new" drug approvals is decreasing significantly over time, while the number of follow-on approvals is increasing. Cumulative data in Figure 3 show that the number of truly innovative drug products is vanishingly small (1.23% of total and 1.87% of brand name approvals) over the eight-year test period. In general, our qualitative findings relating to pharmaceutical innovation parallel those observed in other jurisdictions.[125] That is, the multinational pharmaceutical industry is leaning away from breakthrough drug development towards less innovative products referred to variously as follow-on, incremental, supplemental, line extension, me too, and bioequivalent drugs.

While our data do not speak directly to claims that diminished innovation is due to the loss of "low hanging fruit"[126] or to the spiraling costs of drug development,[127] the data regarding the nexus between drug approval and patenting provide a third plausible explanation for the diminution of breakthrough product development. The results shown in Figures 1–7 suggest that innovation policy and drug regulation contingent on IPR rights can profoundly shape the rate and direction of innovative activity by multinational pharmaceutical firms antecedently, towards incentives provided for by law and away from truly breakthrough products under conditions where the two do not necessarily coincide.

Depending on the source and degree of industry affiliation, published de-

125. *See, e.g.*, NAT'L INST. FOR HEALTH CARE MGMT., CHANGING PATTERNS OF PHARMACEUTICAL INNOVATION 7 (2002); Domenico Motola et al., *An Update on the First Decade of the European Centralized Procedure: How Many Innovative Drugs*, 62 BRIT. J. OF CLINICAL PHARMACOLOGY 610 (2006); Editorial, *European and French Pharmaceutical Market Assessed by Prescrire in 2005: Mainly Bogus Innovation*, 30 FARMACIA HOSPITALARIA 68 (2006); John Abraham & Courtney Davis, *A Comparative Analysis of Drug Safety Withdrawals in the UK and the US (1971–1992): Implications for Current Regulatory Thinking and Policy*, 61 SOC. SCI. & MED. 881 (2005); *Drugs in 2001*, *supra* note 17; Kenneth I. Kaitin et al., *Therapeutic Ratings and End-of-Phase II Conferences: Initiatives To Accelerate the Availability of Important New Drugs*, 31 J. CLINICAL PHARMACOLOGY 17 (1991); *New Medicines in 2007: Regulatory Agencies and Policy Makers Leave Public Health in the Hands of the Pharmaceutical Industry*, 17 PRESCRIRE INT'L 78 (2008).

126. Fredric J. Cohen, *Macro Trends in Pharmaceutical Innovation*, 4 NATURE REVS. DRUG DISCOVERY 78, 82 (2005).

127. *See generally* Joseph A. DiMasi et al., *The Price of Innovation: New Estimates of Drug Development Costs*, 22 J. HEALTH ECON. 151 (2003).

finitions of what constitutes an "innovative drug" vary considerably, from as low a threshold as simply containing an NAS,[128] to the slightly more stringent requirements of either being directed to FIC therapies[129] (irrespective of whether approval is directed to a new or follow-on drug) or to follow-on drugs that nevertheless undergo Priority Review.[130] However, merely containing an NAS is an insufficient basis for designating a drug as pioneering or strongly innovative because there is ample room in the definition for minor changes to previously approved medical ingredients, including salts, esters, solvates, polymorphs, and enantiomers. A similar conclusion applies to drugs that are only directed to FIC therapies, as these can also be follow-on versions of previously marketed products containing slightly modified medical ingredients or directed to new uses within a therapeutic class. Moreover, where Priority Review need only be directed to drugs demonstrating moderate clinical improvement over existing therapies, it is also an insufficient proxy for strong innovation. A more reasonable definition is that truly pioneering drugs are those that are approved via the new drug approval pathway (NDS), contain an NAS or NCE, undergo some form of Priority Review, and are directed to a FIC therapy.[131] Only in combination do these requirements approach a reasonable definition for a truly breakthrough technology.

Regulatory agencies in North America have previously attempted to derive innovation indices for pharmaceuticals. For example, in 2000, the Canadian Patented Medicines Prices Review Board[132] released data to the effect that of drugs approved between 1996–2000, 44.8% were line extensions and 49.6% were new versions of marketed drugs with moderate, little, or no improvement. Only 5.5% of all drugs approved represented a substantial therapeutic advance. These results parallel data from a large-scale study of innovation in the French prescription drug market demonstrating that of drugs approved over the term 1981–2001, the most innovative drugs represented only 3% of total approvals, while drugs with some important therapeutic gain and those with little to no therapeutic gain represented 8% and 89% of total approvals, respectively.[133]

---

128. J. D. Kleinke, *Commentary: Much Ado About a Good Thing*, 325 BRIT. MED. J. 1168 (2002).

129. Cohen, *Macro trends in pharmaceutical innovation*, *supra* note 126; COMM. ON KNOWLEDGE ASSESSMENT, NAT'L RESEARCH COUNCIL, PROSPECTUS FOR NATIONAL KNOWLEDGE ASSESSMENT (1996).

130. NAT'L INST FOR HEALTH CARE MGMT., *supra* note 125.

131. *Id.*

132. PATENTED MEDICINE PRICES REVIEW BOARD, ANNUAL REPORT 2000 (2001), *available at* http://www.pmprb-cepmb.gc.ca/English/View.asp?x=113&mp=91.

133. *Drugs in 2001*, *supra* note 17, at 59; *see also New Medicines in 2007*, *supra* note 128, at

For the United States, Kaitin et al. reported data from an analysis of drugs approved by the FDA between 1978–1989 that were rated by the agency as having an important therapeutic gain, a modest gain, and little to no gain. [134] Only NCEs which excluded salts, esters, and other dosage forms of previously approved drugs were studied. The authors found that only 14.7% of approvals had the strongest innovation rating, whereas 34.5% and 49.5% were deemed modestly or weakly innovative. A more recent study by the NIHCM[135] demonstrated that of all drugs approved by the FDA during 1989–2000, 15%, 28%, and 57% were deemed to be the "most innovative" (NCE plus Priority Review), "moderately innovative" (follow-on plus Priority Review), and "modestly innovative"(follow-on), respectively.

As in Figures 1a and 1b, in the NIHCM and French studies, approvals for standard follow-on drugs increased steeply over the test periods, while data for the most innovative drugs remained flat over time. The fact that the values of 14.7% and 15% from the Kaitin and NIHCM studies represent NCEs with important therapeutic gain or drugs approved via the NDS stream, rather than drugs also undergoing Priority Review and directed to FIC therapies, suggests that the number of truly breakthrough drugs in these studies was more in line with data in Figures 1–3. With the exception of the French study, each of these indices were reported shortly after policy initiatives impacting drug development came into force, such as linkage regulations in the United States and Canada, the consolidation of U.S. patent appeals courts, and legislation facilitating technology transfer and commercialization via strong IPR rights.[136]

B. ROLE OF DRUG PATENTING AND LINKAGE REGULATIONS

Despite potential weaknesses in the empirical underpinnings of the innovation indices noted supra, it is of interest that there has never developed a parallel empirical literature relating to the social benefits of public health

---

78–82. Note that the 2001 French study included new drugs and also new indications for existing drugs already on the market. Moreover, generic drugs were rated as no improvement over existing drugs. For discussion of the 2000 Canadian and 2001 French studies in the context of the Canadian pharmaceutical marketplace see generally Lexchin, *supra* note 17.

134. Kaitin et al., *supra* note 125, at 17–24.

135. NAT'L INST. FOR HEALTH CARE MGMT., *supra* note 125, at 8.

136. Jaffe, *supra* note 94 (discussing the pros and cons of arguments favoring the stimulation of innovation through provision of strong patent rights); Bhaven N. Sampat, *Patenting and US Academic Research in the 20th Century: The World Before and After Bayh-Dole*, 35 RES. POL'Y 772 (2006) (reviewing policy, legislative, and court reforms intended to facilitate commercialization of innovative products, including via strong intellectual property rights); Bouchard, *supra* note 61 (discussing the balancing of private intellectual property rights and publicly funded research in producing and commercializing innovative medical products).

and/or innovation policy that is strongly contingent on IPR rights[137]or that comprises a regulatory preference for incremental innovation. The social benefits of innovation are raised under the linkage regulations regime through the terms of the traditional patent bargain. This refers to the grant of a limited monopoly in exchange for public disclosure of socially valuable knowledge.[138] In a public health context where drug approval and drug patenting are linked, the essence of the patent bargain may be viewed as the exchange of extended patent protection for a socially beneficial level of pharmaceutical innovation. Thus, the public expects, and should expect, something of substantial social value in exchange for extended patent protection and monopoly pricing. In other words, the empirical or other data should support a strong legal and functional nexus between public health policy and patent policy.

Undue extension of patent protection for poorly innovative drugs via linkage regulations is similar in manner to so-called "weak" patents. Weak patents are those that provide poor levels of innovation over relevant prior art.[139] Leading courts have consistently held that patents of this nature stifle innovation,[140] chill competition,[141] encroach on the legal mandate of promoting the progress of science and useful arts,[142] and encourage inefficient transfers of wealth.[143] Relevant to the pharmaceutical market, weak patents hijack the IPR rights landscape[144] and allow patentees to extract unwarranted monopoly rents when they would otherwise receive nothing for non-inventive disclosures.[145] Policies underpinning patent protection must be sufficiently worthwhile to the public to warrant the restrictive effect of the patent mo-

137. Mazzoleni & Nelson, *supra* note 94; Pavitt, *supra* note 97.

138. *See, e.g.*, KSR Int'l v. Teleflex, Inc., 550 U.S. 398, 419 (2007); Graham v. John Deere Co. of Kan. City, 383 U.S. 1, 17–18 (1966); Whirlpool Corp. v. Camco Inc., [2000] 2 S.C.R. 1067, ¶ 37 (Can.); *see also* Ron A. Bouchard, KSR v. Teleflex *Part 1: Impact of U.S Supreme Court Patent Law on Canadian Intellectual Property and Regulatory Rights Landscape*, 15 HEALTH L.J. 221 (2007).

139. *KSR Int'l*, 550 U.S at 419.

140. Hotchkiss v. Greenwood, 52 U.S. 248 (1851).

141. Whirlpool, 2 S.C.R. 1067; Free World Trust v. Electro Sante Inc., [2000] 2 S.C.R. 1024 (Can.); Royal Typewriter Co. v. Remington Rand, Inc., 168 F.2d 691 (2d Cir. 1948).

142. *KSR Int'l*, 550 U.S. at 419.

143. Anita Varma & David Abraham, *DNA is Different: Legal Obviousness and the Balance Between Biotech Inventors and the Market*, 9 HARV. J.L. & TECH. 53, 55 (1996). For general discussion of the relevance of weak patents to innovation, see also Glynn S. Lunney, Jr., *E-Obviousness*, 7 MICH. TELECOMM. & TECH. L. R. 363 (2001) and Bouchard, *PHOSITA*, *supra* note 73.

144. *See Royal Typewriter Co.* 168 F.2d at 693–94.

145. *See* Lunney, *supra* note 146, at 384.

nopoly,[146] and weak pharmaceutical patents in particular have been held to offend the public interest.[147]

The applicability of jurisprudence relating to weak patents may be particularly relevant to linkage regulations owing to two considerations that do not apply to other industries, technology-based or otherwise. The first is that the weak relevance standard for listing, which as discussed supra, provides a very broad target for patentees when aiming for the automatic injunction under both U.S. and Canadian linkage regulations.[148] This injunction, an earlier form of which has been referred to as "Draconian",[149] prevents generic firms from market entry until all patents on the register are proved in litigation to be invalid or not infringed.[150] The second is the empirical observation in both jurisdictions that litigation on the merits of contested patents under linkage regulations results in decisions where 75% of listed patents are deemed by the courts to be either invalid or not infringed by generic products.[151] It is reasonable to speculate that the administrative costs of prolonged litigation under linkage regulations are passed on to consumers in the form of extended monopoly pricing and other rent seeking behaviors.[152]

A linkage regulation regime that provides patent protection on poorly innovative drugs that extends well beyond the term of originating patents, not only has the potential to debilitate the patent system in the short term,[153] but

146. Graham v. John Deere Co. of Kan. City, 383 U.S. 1, 11 (1966).

147. *See, e.g.*, R. v. Nova Scotia Pharm. Soc'y, [1992] 2 S.C.R. 606 (Can.); Comm'n of Patents v. Farbwerke Hoechst Aktiengesellschaft, [1964] S.C.R. 49, 56 (Can.).

148. Caffrey & Rotter, *supra* note 84; Hore, *supra* note 73. For a detailed discussion of the standard for relevance, see Regulations Amending the Food and Drug Regulations (Data Protection), Regulatory Impact Analysis Statement, 140 C. Gaz. pt. II, at 1495–1502 (2006), *available at* http://www.gazette.gc.ca/archives/p2/2006/2006-10-18/pdf/g2-14021.pdf.

149. Merck Frosst Can. Inc. v. Canada, [1998] 2 S.C.R 193 ¶ 33 (Can.). This passage has been cited by the court in later decisions. *See, e.g.*, AstraZeneca Can., Inc. v. Canada, [2006] S.C.R. 560, 2006 SCC 52 ¶ 17 (Can.); Bristol-Myers Squibb Co. v. Canada, [2005] 1 S.C.R. 533 SCC 26 ¶¶ 24, 146 (Can.).

150. *Id.*

151. FED. TRADE COMM'N, GENERIC DRUG ENTRY PRIOR TO PATENT EXPIRATION: AN FTC STUDY (2002), http://www.ftc.gov/os/2002/07/genericdrugstudy.pdf [hereinafter FTC 2002]; *see* ED HORE, PATENTLY ABSURD: EVERGREENING OF PHARMACUETICAL PATENT PROTECTION UNDER THE PATENTED MEDICINES (NOTICE OF COMPLIANCE) REGULATIONS OF CANADA'S PATENT ACT 5, 11 (2004) http://www.canadiangenerics.ca/en/news/docs/patently_absurd_04.pdf (discussing data in the context of U.S. and Canadian linkage); Caffrey & Rotter, *supra* note 84, at 13–14.

152. *See* BOLDRIN & LEVINE, *supra* note 97, at 260–265 (discussing generally rent-seeking through patenting of follow-on drugs).

153. *See* Whirlpool Corp. v. Camco Inc., [2000] 2 S.C.R. 1067, ¶ 37 (Can.) (arguing that extending patent protections with follow on patents to obvious variations is counter to the patent bargain); Graham v. John Deere Co. of Kan. City, 383 U.S. 1, 17–18 (1966) (holding

also to weaken pharmaceutical innovation in the long run. Innovation is weakened because the combination of weak relevance requirements and automatic injunctions takes patent protection to a point near its logical extreme. The data reported here suggest that if linkage regimes provide fertile grounds for firms to compete at a lower level of innovation, they also discourage firms from innovating at a level of competition that would provide the greatest benefit to society.

This dilemma can be illustrated by a comparison of data in Figures 3 and 5. On the one hand, Figure 3 demonstrates that a very small fraction of drugs approved by regulators over the 8 year test period could be considered truly breakthrough in nature. This includes drugs approved via the NDS stream (16%), those containing an NAS (6.1%), total NDS and SNDS drugs directed to FIC therapies (6.5%), those that underwent one of two pathways for expedited review (5.3%), and those that met the most stringent requirements for breakthrough products (1.23%). On the other hand, Figure 5 illustrates that patent protection under linkage regulations does not discriminate between poorly or strongly innovative drugs. It offers broad and long-lasting IPR rights to pharmaceutical firms regardless of the types of products being introduced into the marketplace. This is particularly relevant for follow-on drug products, which are recognized to entail lower risks and costs to pharmaceutical firms.[154] As suggested by data in Figures 1c, 6b, and elsewhere,[155] the evolution toward a lifecycle-based regulatory approach to drug approval will likely do little to affect the rate and direction of innovative activity by firms absent shifts in legal incentives for breakthrough and follow-on drug development.

The data in Figures 4–7 further support discordance between the basket of IPR rights incentives for innovation and resulting product development. For example, the close temporal relationship between drug approval and patent listing in Figure 4d and the strong convergence in Figure 4a of patent grant and patent listing following linkage regulations coming into force provide evidence that patent listing evolved into a more effective target, and thus a better proxy, for drug approval than drug patenting per se once the linkage regime came into effect. Other evidence for this conclusion comes from data in Figure 6, which demonstrate that steep time-dependent changes in drug patenting, patent listing, and the evolution toward lifecycle regulation appeared to have occurred independently of concomitant trends for new and follow-on drug approvals. The outcome of this dynamic, supported by aver-

that patents on obvious variations of known objects are invalid).

154. Cohen, *supra* note 126, at 79.

155. Sawicka & Bouchard, *supra* note 35.

aged data for sixteen drugs (Figure 5) and the single example of omeprazole (Figure 7), is that when given the opportunity, pharmaceutical firms will leverage government policy and regulation in order to maintain market share for drugs coming off patent instead of developing new blockbuster drugs. The results are not dissimilar to studies of complex political systems, where "yardsticks" designed to measure progress reorient behavior narrowly towards fulfillment of yardstick metrics.[156] The implication of this scenario is that firms are aiming ex ante at legal targets which provide the most return on investment rather than the most benefit to the public.

C. CONVERGENCE OF ECONOMIC AND PUBLIC HEALTH POLICY

The data reported here challenge two key assumptions that have underpinned public health policy and economic/industrial policy in industrialized nations for at least the past half century. The first is that strong IPR rights protection is essential to motivate and increase the amount of innovation that occurs in the economy. The second is that public health goals can best be met by encouraging innovation in private industry, essentially by merging public health goals with industrial development goals, buttressed by the IPR rights regime. Importantly, our findings do not indicate abnormal behavior by pharmaceutical companies, but rather failure of government policy and regulations to produce a specifically desired effect, namely, the increased production of truly innovative drugs. It is entirely understandable that pharmaceutical firms avail themselves of regulatory incentives allowing product evergreening after the original patent has expired where it maximizes the benefit and minimizes the risk to shareholders.[157]

Our findings suggest that blending of industrial and health policy goals may be ineffective and possibly counterproductive in terms of public health outcomes. They also suggest that although new lifecycle regulatory regimes have great potential to increase the efficiency of public health provision by

---

156. *See generally* ROBERT JERVIS, SYSTEM EFFECTS: COMPLEXITY IN POLITICAL AND SOCIAL LIFE (1997).

157. AstraZeneca Can., Inc. v. Canada, [2006] S.C.R. 560, 2006 SCC 52 ¶ 39 (Can.). Discussing the "general" relevance requirement articulated by the Federal Court of Appeal in *Eli Lilly Canada Inc. v. Canada* [2003] FCA 24, Justice Binnie stated,

> Given the evident (and entirely understandable) commercial strategy of the innovative drug companies to evergreen their products by adding bells and whistles to a pioneering product even after the original patent for that pioneering product has expired, the decision of the Federal Court of Appeal would reward evergreening even if the generic manufacturer (and thus the public) does not thereby derive any benefit from the subsequently listed patents.

Id.

placing new remedies in clinical environments sooner, the efficacy of this approach can be weakened through inadequate monitoring and supervision, such that pharmaceutical firms perceive a higher incentive to exploit existing patented technologies in new ways rather than increasing the flow of new technologies. At a more general level, the data lend empirical support to an emerging consensus that, in many circumstances, IPR rights may be an inhibitor of innovation.

Although our study was based on domestic Canadian data, we argue that the results are significant within the global context of drug regulatory reform and innovation policy. First, efforts have been underway for some time to harmonize the goals and mechanisms of drug regulation globally. Second, in most developed nations, university-based translational research, firm research and development activities, and national science and technology policy are closely integrated and likewise mirror one another. Third, qualitative trends in approval of new and follow-on drugs track one another fairly closely in most major jurisdictions, and the drug patents that we analyzed represent the most profitable drugs not only in Canada, but also in U.S. and E.U. markets. Given that a small number of multinational pharmaceutical corporations are responsible for drug innovation globally[158] and are doing so increasingly in partnership with drug regulators,[159] it is reasonable to speculate that drug development and regulation in OECD economies is steadily converging upon a risk management philosophy whereby critical benefit-risk calculations are made not only for drug approval, but also for drug development.

We conclude that policy and legislative incentives designed to stimulate innovation in the pharmaceutical industry have had the opposite effect, and that shifting to a lifecycle regulatory model is unlikely to alter this scenario absent effort to balance legal and regulatory incentives for breakthrough and follow-on drug development. Our findings do not suggest abnormal behavior by pharmaceutical firms, but rather a failure of policy incentives intended to induce the desired result.[160]

158. BOLDRIN & LEVINE, *supra* note 97, at 241–42.

159. *See generally* Mary E. Wiktorowicz, *Emergent Patterns in the Regulation of Pharmaceuticals: Institutions and Interests in the United States, Canada, Britain, and France*, 28 J. HEALTH POL., POL'Y & L. 615 (2003).

160. *See* Barry Bozeman & Daniel Sarewitz, *Public Values and Public Failure in US Science Policy*, 32 SCI. AND PUB. POL'Y 119 (2005); Barry Bozeman, *Public Value Failure: When Efficient Markets May Not Do*, 62 PUB. ADMIN. REV. 145 (2002); John D. Sterman, *All Models Are Wrong: Reflections on Becoming a Systems Scientist*, 18 SYS. DYNAMICS REV. 501 (2002).

D. ANALYTICAL MODEL: PHARMACEUTICALS AND THE ECOLOGY OF INNOVATION

In this Article, we describe qualitatively and quantitatively various elements of the legal landscape governing biomedical innovation in a way that indicates that it functions as a strongly networked *innovation ecology*. We have referred to this ecology in previous work as a regulated Therapeutic Product Lifecycle (rTPL).[161] A highly simplified rTPL diagram is shown in Figure 8, which represents the lifecycle of therapeutic product development and regulation as a system ecology where the "whole is greater than the sum of its parts." Here, innovation is not depicted as a linear "pipeline," or process, moving from basic medical research in universities to applied research in firms, and then on to commercialized products.[162] Rather, the grouping of network nodes in the figure (arbitrarily but functionally connected) are interconnected and interdependent in an iterative manner over time. The functioning of the system cannot be understood from analysis of the properties of individual nodes.[163] Important for the present study, strongly altering the function of one element in the system, in this case the basket of IPR rights intended to stimulate innovation, has the potential to alter the behavior of the entire system.[164]

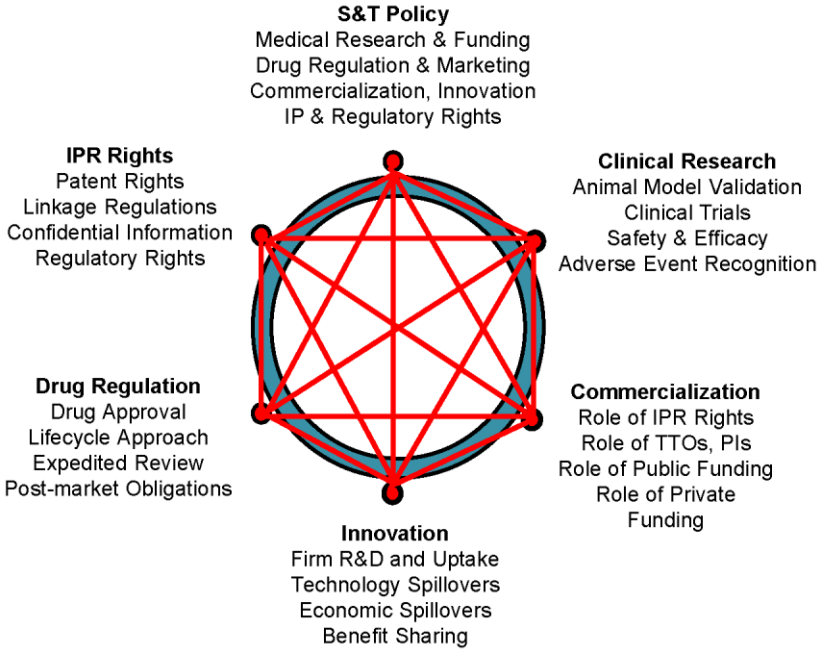
161. Ron A. Bouchard, *Reflections on the Value of Systems Models for Regulation of Medical Research and Product Development*, 17 HEALTH L. REV. 28 (2008) [hereinafter Bouchard, *Reflections*]; Sawicka & Bouchard, *supra* note 35.

162. Godin, *supra* note 90; STOKES, *supra* note 90.

163. Dean Rickles et al., *A Simple Guide to Chaos and Complexity*, 61 J. EPIDEMIOLOLOGY & COMMUNITY 933 (2007).

164. JOHN H. MILLER & SCOTT E. PAGE, COMPLEX ADAPTIVE SYSTEMS: AN INTRODUCTION TO COMPUTATIONAL MODELS OF SOCIAL LIFE 9 (2007).

Figure 8: Systems Model of a regulated Therapeutic Product Lifecycle (rTPL) Innovation Ecology



Innovation is represented as an iterative process over time involving several functional groupings, including national science and technology (S&T) policy, clinical research, university and firm commercialization, innovation by private firms, drug regulation by national governments, and intellectual property and regulatory (IPR) rights covering both drug submissions and marketed products. Large red nodes represent functional groupings, and include sub-functions enumerated in the figure. Red lines are multi-directional between nodes and sub-functions and are independent of time (acknowledging that the process generally moves clockwise).

It occurs to us that feedbacks between the various nodes in this innovation ecology are indicative of phenomena associated with complex adaptive systems, in which positive and negative feedback governs system learning, growth, and self-regulation.[165] In both biological and social systems, it has

165. Feedback interactions in complex systems have received increased attention in recent years. *See generally* ALBERT-LASZLO BARABASI, LINKED: HOW EVERYTHING IS CONNECTED TO EVERYTHING ELSE AND WHAT IT MEANS (Plume 2003) (investigating the role of feedback in biological and social networks, including corporations and living organisms, producing system fitness); JAMES GLEICK, CHAOS: MAKING A NEW SCIENCE (1988) (describing order and chaos generally and how complex systems balance the two through adaptation and positive and negative feedback loops); JOHN H. HOLLAND, ADAPTATION IN NATURAL AND ARTIFICIAL SYSTEMS (1992) (outlining the importance of adaptive mechanisms in natural and artificial systems to the growth and destruction of complex systems);

been demonstrated that strong positive feedback has the potential to move a system away from fitness or operational efficiency, even to the point of inducing the system to collapse.[166] In a complex system, 'order' can help the functioning of the system, but hinder it in others. For example, it has been observed in a range of natural and biological systems that imposition of too much order can yield a system that is inflexible.[167] Moreover, this inflexibility has the potential to move the system away from a state of fitness, in this case the production of breakthrough drugs. Once major patterns and institutions have been fully explored in a highly regulated system, the system may transition into what Kaufmann refers to as "detail mode"[168] where its further evolution is limited to modest improvements on increasingly optimized designs. Indeed, some evidence suggests that the more complicated the system, the more autonomous the agents in the system become, thus reducing the levels of control that it is possible to wield over them without stifling fitness or efficiency.[169]

JOHN H. HOLLAND, HIDDEN ORDER: HOW ADAPTATION BUILDS COMPLEXITY (1995) (discussing adaptation in complex adaptive systems and how order and disorder are often balanced at subtle levels in these systems); STEVEN JOHNSON, EMERGENCE: THE CONNECTED LIVES OF ANTS, BRAINS, CITIES, AND SOFTWARE (Scribner 2002) (discussing the characteristics of emergent systems, including the role of positive and negative feedback loops in governing de-centralized system growth and adaptation); STUART KAUFFMAN, AT HOME IN THE UNIVERSE: THE SEARCH FOR THE LAWS OF SELF-ORGANIZATION AND COMPLEXITY (Oxford University Press 1995) (investigating the conditions that give rise to the growth and destruction of complex adaptive systems and describing how optimal complex adaptive systems are balanced on the edge of chaos); GREGOIRE NICOLIS & ILYA PRIGOGINE, EXPLORING COMPLEXITY: AN INTRODUCTION (1990) (addressing the problem of complexity in using mathematical modeling and the role of essentially irreducible uncertainty in complex systems); M. MITCHELL WALDROP, COMPLEXITY: THE EMERGING SCIENCE AT THE EDGE OF ORDER AND CHAOS (Simon & Schuster 1992) (discussing the role of the interrelation and inter-dependence of players, including individuals and institutions, in complex adaptive systems and showing that systems of this nature are never in stasis, but rather always continually evolving); Brian W. Arthur, *Positive Feedbacks in the Economy*, 262 SCI. AM. 92, 92–99 (1990) (discussing the presence of feedback in producing order and simplicity even in the most complex economic systems).

166. *See, e.g.*, Robert M. May et al., *Complex Systems: Ecology for Bankers*, 451 NATURE 893 (2008) (explaining that catastrophic changes in financial systems can be attributed to its feedback mechanisms). For a look at the role of feedback in policy failure, see generally Bozeman & Sarewitz, *supra* note 160, (discussing how unintended consequences can result in "policy failure") and Sterman, *supra* note 160 (discussing the contribution of uncertainty and the unintended consequences of an action related to inadequate problem articulation). For a review of feedback in complex international political systems, see generally JERVIS, *supra* note 159; COMPLEXITY IN WORLD POLITICS: CONCEPTS AND METHODS OF A NEW PARADIGM (Neil E. Harrison ed., SUNY Press, 1997).

167. KAUFFMAN, *supra* note 165, at 26.

168. *Id.* at 14.

169. JOHNSON, *supra* note 165, at 186.

Based on the reasoning above, it would seem to be a reasonable conjecture that a complex adaptive innovation ecology, such as we have depicted as an rTPL, may then be one with a large degree of potential creativity and productivity balanced by an equal degree of uncertainty and instability and effected through positive and negative feedback loops, including those initiated by law. This discussion has potentially significant implications for the interpretation of existing pharmaceutical policy, regulation, and literature, including the data reported here. In our previous work on pharmaceutical innovation and litigation,[170] we suggested that regulatory preferences that do not respect the complex nature of the system they seek to regulate (including over-regulation masquerading as under-regulation) have the potential to harm the innovative outputs of the system. This result can be affected by either allowing undue capture of resources or benefits into the hands of discrete actors or through loss of innovative capacity relative to practical considerations of use, including those incentivized through regulatory preferences.

The model in Figure 8 envisions all steps in the innovation process as interdependent, particularly over the longer horizon.[171] At the 'beginning' of the process, national science and technology policies are negotiated and initiated to drive national innovation priorities.[172] These policies set the balance between economic and public health goals and expenditures.[173] The next point is represented by publicly funded medical research,[174] which policymakers now desire to be strongly 'translational' in nature and therefore underpinned by strong IPR rights.[175] The mid-point of the process is where clinical trial results become increasingly available, at which point firms identify attractive technologies and begin to layer more substantial IPR rights over them, particularly patent rights. These patent rights, and the various spin-out firms they can create (e.g. from technology transfer), then become metrics, which in turn are used to determine what constitutes effective and efficient national science and technology policies and practices.[176] Finally, we move

170. For general discussion of the problems inherent in linear legislative and jurisprudential models of pharmaceutical innovation and how they may be mitigated by systems models of innovation, see generally Ron A. Bouchard, KSR v. Teleflex *Part 2: Impact of U.S. Supreme Court Patent Law on Canadian and Global Systems-Based Innovation Ecologies*, 15 HEALTH L.J. 247 (2007) [hereinafter Bouchard, *Systems*]; Bouchard, *Reflections*, *supra* note 161; Bouchard, *PHOSITA*, *supra* note 73.

171. Arthur, *supra* note 93.

172. Bouchard, *Systems*, *supra* note 170 at 248–50; *see also* Bouchard & Sawicka, *supra* note 29, at 57–58.

173. Bouchard & Lemmens, *supra* note 87.

174. *See id.*; Bouchard, *supra* note 61.

175. *See id.* at 2–3.

176. For discussion of the failure of linear models of "basic" and "applied" research and

towards the perceived terminus of the process, where products are at or near the regulatory approval point and firms have identified targets for either novel breakthrough products or incremental innovations with strong evergreening potential.[177] At this point, and especially at later points in the rTPL, [178] linkage regulations and regulatory rights become dominant forms of IPR rights protection.[179] However, as noted earlier, the mid-point and end-point of the pharmaceutical innovation system are increasingly merging, as regulators move towards lifecycle regulatory models which allow for early or flexible departure of drugs prior to completion of tradition Phase 3 trials, with greater post-market surveillance. Moreover, both pharmaceutical, and more recently biotechnology, firms operating under the linkage regime can now layer IPR rights on products at all stages of development, including those about to come off patent, those in regulatory review, and those in development. Recent data[180] indicate that the linkage regime operating in conjunction with established patent law and the drug approval regime allows firms to produce a substantial number and array of patent classifications, which can in turn be used to list on the patent register in order to prohibit generic entry on older drugs and to support follow-on drug development submissions, thus further collapsing the drug development cycle. The present study therefore supports the need to extend and broaden the innovation analysis to include the entire landscape of interconnections between drug approval, patenting, and litigation, as well as the nexus between broader national science and technology policies and the effects thereof on the rate and direction of firm innovation.[181]

Schumpeter noted that innovations of different magnitudes tended to appear in cycles of varying lengths, geared largely to the rate at which advantages from innovations declined over time through increasing use and imita-

---

development to account for the innovation process, see generally STOKES, *supra* note 90; Godin, *supra* note 90.

177. Bouchard, *Scientific Research*, *supra* note 73, at 13; Bouchard, *PHOSITA*, *supra* note 73, at 22–23.

178. For discussion of regulatory rights, how they relate to traditional food and drug law, and the points in the drug development cycle at which they come into play, see generally Caffrey and Rotter, *supra* note 84; Eisenberg, *supra* note 84; Hore, *supra* note 73; HORE, *supra* note 151.

179. Bouchard, *PHOSITA*, *supra* note 73, at 48; Bouchard & Sawicka, *supra* note 29, at 63–64.

180. Ron A. Bouchard et al., *Empirical Analysis of Drug Approval-Patenting Linkage for High Value Pharmaceuticals*, 8 NW. J. TECH. & INTELL. PROP. (forthcoming 2010).

181. Bouchard, *Systems*, *supra* note 170, at 258–62; Bouchard, *Reflections*, *supra* note 161, at 38–39.

tion.[182] For policies and regulations aimed at stimulating innovation, the risk is always that they may catch one of these cycles at the wrong moment, thus contributing more to the declining phase of an existing cycle than to the development phase of a new one. They may do this by damaging the incentives that drive new entrants, or by preserving practices that have become inefficient.[183] Clearly this applies to inefficient or ineffective regulatory policies that lead to increasingly poor performance as judged by the goals and objectives of policy-makers, in this case an increased supply of truly innovative remedies.

Based on data here and elsewhere,[184] we propose that the current lifecycle of pharmaceutical development *and* regulation may be nearing a point of exhaustion such as Schumpeter would have recognized. Evidence for this includes: a strongly increasing trend towards ever-smaller incremental innovation in the last decade (Figure 1); an increase of low level of innovation being supported by a combination of weak patents and linkage regulations (Figures 4–7); a decreasing number of truly breakthrough drugs as well as drugs containing NCEs and NASs (Figures 1 and 3); a substantial number of patents per drug (Figure 4); the fact that many patents under linkage regulations are either invalid or infringed when tested on the merits,[185] and the growth in both the scope and depth of IPR rights associated with poorly innovative drug products over the last 20 years.

As noted by many commentators, the basket of IPR rights afforded to pharmaceuticals has grown to encompass an astounding array of mechanisms, which may be interpreted as micro levels of order or detail as per the discussion supra. These include increased patent terms, decreased standards for obviousness, utility, and subject matter requirements for patenting, allowance for listing of weak patents via linkage regulations, the automatic stay provision barring generic entry, loss of compulsory licensing provisions, and the ever growing basket of regulatory rights associated with drug submis-

182. JOSEPH A. SCHUMPETER, BUSINESS CYCLES: A THEORETICAL, HISTORICAL AND STATISTICAL ANALYSIS OF THE CAPITALIST PROCESS (1939); *see also* Gert-Jan Hospers, *Joseph Schumpeter and His Legacy in Innovation Studies*, 18 KNOWLEDGE TECH. & POL'Y 20 (2005) (reviewing the relevance of Schumpeter's work for current innovation theory and practice).

183. *Id.* at 32.

184. For a review of cumulative empirical studies of pharmaceutical innovation and patenting, see BOLDRIN & LEVINE, *supra* note 97, particularly Chapter 8 ("Does Intellectual Property Increase Innovation?") and Chapter 9 ("The Pharmaceutical Industry").

185. *See, e.g.*, FTC 2002, *supra* note 151; Caffrey & Rotter, *supra* note 84; Valerie Junod, *Drug Marketing Exclusivity Under United States and European Union Law*, 59 FOOD & DRUG L.J. 479 (2004); Hore, *supra* note 73.

sions.[186] It is not just a coincidence that the basket of IPR rights attached to pharmaceutical products is growing in both scope and depth at a time when innovation is widely considered to be faltering.

Even if the current rTPL is not near the point of exhaustion, data such as those reported here should provide useful information for jurisdictions contemplating some form of linkage regulations or other types of linkages between public health and economic policies. In jurisdictions that maintain that IPR rights are integral to innovation, the results may offer an opportunity to correct or fine-tune existing policies underpinning innovation, including adjusting economic incentives in accordance with the degree of innovation and accompanying social benefits[187] based on a growing body of empirical data.[188]

---

186. Regulations Amending the Food and Drug Regulations (Data Protection), Regulatory Impact Analysis Statement, 140 C. Gaz. pt. II, at 1495–1502 (2006), *available at* http://www.gazette.gc.ca/archives/p2/2006/2006-10-18/pdf/g2-14021.pdf (demonstrating that "regulatory rights," such as market data and pediatric exclusivity, add up to a term of market exclusivity in various jurisdictions ranging from 5.5 to 11.5 years and that this period of market exclusivity exists independent and alongside patent protection via traditional patent legislation and emerging linkage regulations.) While regulatory rights spread globally via provisions to this effect in TRIPS and other U.S.-based trade agreements, they have been the subject of increasing scrutiny recently, including within the United States. For example, Senator Bernie Sanders (I-VT) recently put forward an amendment to the health care reform bill that would eliminate data exclusivity where duplicating clinical trials involving human subjects violates Article 20 of the Declaration of Helsinki on Ethical Principles for Medical Research Involving Human subjects pertaining to clinical trial ethics. James Love, *Senator Sanders Amendment 2858 Would Replace Data Exclusivity with Cost Sharing, If New Trials Violate Medical Ethics*, KNOWLEDGE ECOLOGY INT'L, Dec. 9, 2009, http://keionline.org/node/707.

187. Michael Abramowicz, *Perfecting Patent Prizes*, 56 VAND. L. REV. 115 (2003); Amitava Banerjee et al., *The Health Impact Fund: Incentives for Improving Access to Medicine*. 375 THE LANCET 166 (2010); Paul Grootendorst, *Patents, Public-Private Partnerships or Prizes: How Should We Support Pharmaceutical Innovation?* (Soc. & Econ. Dimensions of an Aging Population, Paper No. 250), *available at* http://socserv2.socsci.mcmaster.ca/~sedap/p/sedap250.pdf; Aidan Hollis, *Optional Rewards for New Drugs for Developing Countries*, (April 5, 2005) (unpublished manuscript, on file with the World Health Organization), *available at* www.who.int/entity/intellectualproperty/submissions/Submissions.AidanHollis.pdf; Aidan Hollis, *An Efficient Reward System for Pharmaceutical Innovation*, (Jan. 17, 2005) (unpublished manuscript, on file with the University of Calgary), *available at* http://econ.ucalgary.ca/fac-files/ah/drugprizes.pdf; Joseph Stiglitz, *Scrooge and Intellectual Property Rights: A medical prize fund could improve the financing of drug innovations*, 333 BRIT. MED. J. 1279 (2006).

188. For example, existing price control methodologies employed by governments (e.g., Patented Medical Prices Review Board of Canada) or insurers and other institutional payers may be modified to incorporate an "innovation index" factor such as that described in Figure 3d in their pricing algorithms. Prices could be increased or decreased in accordance with empirical assessment of whether approved drugs, when compared with existing drugs, are highly innovative (NDS+NAS+ER+FIC or NDS+NAS+FIC), moderately innovative (NDS+ER+NAS or NDS+NAS), less innovative (SNDS+ER+FIC or SNDS+FIC), poorly innovative (SNDS) or not innovative at all (ANDS; SANDS).

E-HEALTH HAZARDS: PROVIDER LIABILITY AND ELECTRONIC HEALTH RECORD SYSTEMS

Sharona Hoffman[†] *& Andy Podgurski* [††]

ABSTRACT

In the foreseeable future, electronic health record (EHR) systems are likely to become a fixture in medical settings. The potential benefits of computerization could be substantial, but EHR systems also give rise to new liability risks for health care providers that have received little attention in the legal literature. This Article features a first of its kind, comprehensive analysis of the liability risks associated with use of this complex and important technology. In addition, it develops recommendations to address these liability concerns. Appropriate measures include federal regulations designed to ensure the quality and safety of EHR systems along with agency guidance and well crafted clinical practice guidelines for EHR system users. In formulating its recommendations, the Article proposes a novel, uniform process for developing authoritative clinical practice guidelines and explores how EHR technology itself can enable experts to gather evidence of best practices. The authors argue that without thoughtful interventions and sound guidance from government and medical organizations, this promising technology may encumber rather than support clinicians and may hinder rather than promote health outcome improvements.

TABLE OF CONTENTS

---

† Professor of Law and Bioethics, Co-Director of Law-Medicine Center, Case Western Reserve University School of Law; B.A., Wellesley College; J.D., Harvard Law School; LL.M. in Health Law, University of Houston.

†† Professor of Electrical Engineering and Computer Science, Case Western Reserve University. B.S., M.S., Ph.D., University of Massachusetts. The authors wish to thank Jessica Berg, Henry Bloom, Shawneequa Callier, David Hyman, Richard Krueck, Maxwell J. Mehlman, Cassandra Robertson, and Robert Strassfeld for valuable comments on drafts of this paper. We are also grateful for the very skillful research assistance of Michael Hill.

I. INTRODUCTION

The American Recovery and Reinvestment Act of 2009 (ARRA),[1] better known as President Obama's stimulus legislation, was enacted to rescue an

1. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009).

ailing economy in early 2009.[2] One of its many goals was "to increase economic efficiency by spurring technological advances in science and health."[3] To that end, ARRA dedicated nineteen billion dollars to the promotion of health information technology.[4]

The ARRA's goal is to computerize all Americans' health records by 2014.[5] Currently, only a small minority of health care practices use electronic health record (EHR) systems, including perhaps seventeen percent of doctors and ten percent of hospitals.[6] In order to comply with this mandate and avoid penalties for non-compliance,[7] health care providers will need to increase their rate of EHR system adoption dramatically.

Comprehensive EHR systems will have a pervasive influence on medical care and serve multiple functions beyond storing medical files. They electronically transmit diagnostic test images and results, laboratory reports, and radiologic images and reports to physicians so that these can be quickly reviewed and shared with patients.[8] The systems feature computerized provider-order entry (CPOE), which allows providers to send electronic orders, such as those for laboratory tests and medications, to appropriate parties.[9] They also feature decision support tools, among which are clinical guidelines, clinical reminders, drug-allergy and drug interaction alerts, and drug-dose support.[10] EHR systems may also provide a secure messaging feature to help physicians communicate with patients confidentially.[11] Ideally, EHR systems should be interoperable and thus be able to automatically

2. *Id.* at § 3 (stating that the purpose of the Act is "to preserve and create jobs and promote economic recovery" and "to assist those most impacted by the recession").

3. *Id.* § 3(a)(3).

4. David Blumenthal, *Stimulating the Adoption of Health Information Technology*, 360 NEW ENG. J. MED. 1477 (2009).

5. American Recovery and Reinvestment Act § 3001(c)(3)(A)(ii).

6. *Id.* (noting that these figures represent practices using basic systems, not necessarily sophisticated or comprehensive systems); Catherine M. DesRoches et al., *Electronic Health Records in Ambulatory Care: A National Survey of Physicians*, 359 NEW ENG. J. MED. 50, 54 (2008); Ashish K. Jha et al., *Use of Electronic Health Records in U.S. Hospitals*, 360 NEW ENG. J. MED. 1628, 1631 (2009).

7. Blumenthal, *supra* note 4, at 1477–78 (noting that "[p]hysicians who are not using EHRs systems meaningfully will lose 1% of their Medicare fees in 2015, then 2% in 2016, and 3% in 2017").

8. Jha et al., *supra* note 6, at 1632.

9. *Id.*

10. *Id.*

11. Catherine Chen et al., *The Kaiser Permanente Electronic Health Record: Transforming And Streamlining Modalities of Care*, 28 HEALTH AFF. 323, 325 (2009) (describing the secure messaging system implemented by Kaiser Permanente Hawaii in September 2005).

incorporate records and process information from EHR systems developed by different vendors.[12]

The potential benefits of computerization are considerable.[13] In short, EHR systems can facilitate access to patients' medical records, improve the quality of care and the accuracy of treatment decisions, achieve cost savings, and promote clinical research.[14] Some health care providers with EHR systems already report better outcomes, fewer complications, lower costs, and fewer malpractice claim payments.[15] Without discounting any of these potential benefits, this Article focuses on the risks of EHR systems and on liability concerns associated with their use. It argues that despite the promise of this technology, the implementation of EHR systems must proceed with both caution and appropriate government oversight.

In recent years, more than a few startling EHR-related stories have surfaced. For example, software glitches in the U.S. Department of Veterans Health Administration's EHR system exposed veterans to excessive,

---

12. BIOMEDICAL INFORMATICS: COMPUTER APPLICATIONS IN HEALTH CARE AND BIOMEDICINE 952 (Edward H. Shortliffe & James J. Cimino eds., 2006) [hereinafter BIOMEDICAL INFORMATICS] (explaining that interoperable systems can communicate with each other, exchange data, and operate seamlessly and in a coordinated fashion across organizations).

13. We have discussed them extensively in prior work. *See* Sharona Hoffman & Andy Podgurski, *Finding A Cure: The Case for Regulation and Oversight of Electronic Health Record Systems*, 22 HARV. J.L. & TECH. 103, 112–19 (2008) (discussing the benefits of EHR systems).

14. *Id.*; *see also* Richard J. Baron et al., *Electronic Health Records: Just Around the Corner? Or Over the Cliff?* 143 ANNALS INTERNAL MED. 222, 225–26 (2005) (discussing the benefits of an EHR system in a small practice); Stephen T. Parente & Jeffrey S. McCullough, *Health Information Technology And Patient Safety: Evidence From Panel Data*, 28 HEALTH AFF. 357, 357–58 (2009) (utilizing four years of inpatient data from Medicare patients and finding that EHRs have "a small, positive effect on patient safety"); Julie Weed, *If All Doctors Had More Time to Listen*, N.Y. TIMES, June 7, 2009, at BU1 (praising EHR systems and arguing that they save physicians time and money). *But see* Yong Y. Han et al., *Unexpected Increased Mortality After Implementation of a Commercially Sold Computerized Physician Order Entry System*, 116 PEDIATRICS 1506, 1510–12 (2005) (noting that the mortality rate among children increased from 2.80% to 6.57% after computerized physician order entry implementation and asserting that further evaluation of this evolving technology is needed).

15. Ruben Amarasingham et al., *Clinical Information Technologies and Inpatient Outcomes*, 169 ARCH. INTERN. MED 108, 111–12 (2009) (reporting on a survey that involved 167,233 patients at 41 urban Texas hospitals); Anunta Virapongse et al., *Electronic Health Records and Malpractice Claims in Office Practice*, 168 ARCH. INTERN. MED. 2362, 2365 (2008) (presenting a survey of 1,345 Massachusetts physicians and stating that although the study's results were inconclusive, they suggest that "physicians with EHRs appear less likely to have paid malpractice claims"). *But see* Steve Lohr, *Little Benefit Seen, So Far, in Electronic Patient Records*, N.Y. TIMES, Nov. 16, 2009, at B3 (reporting on research that revealed that EHR systems have "not yet had a real impact on the quality or cost of health care").

potentially life-threatening dosages of the blood-thinner heparin.[16] In a different incident, a hospital pharmacy's computer program generated erroneous medication order lists, leading to the delivery of the wrong drugs to patients in many wards.[17] A May 2009 article featured the alarming title " 'Nearly Killed' by E-Records Data Model" and described the distressing experience of a patient in an intensive care unit with an EHR system that did not allow doctors and nurses to access critical medical information or obtain medication from the pharmacy in a timely fashion.[18] The liability risks of EHR systems, however, have received little attention in the legal literature.

Along with the potential to enhance health outcomes, this new technology may bring novel responsibilities, burdens, and complexities for medical practices. Historically, medical innovations, such as anesthetics and x-rays, have generated increased tort litigation as patients quickly came to expect better care while physicians struggled to perfect their use of challenging technologies.[19] The same phenomenon may well occur with EHR systems. This Article details specific liability risks associated with EHR systems and explores strategies to alleviate liability concerns.[20] For the sake of simplicity, we use the terms EHR and EHR systems to designate electronic health records and the systems in which they operate. We mean the term EHR to be synonymous with what others call the electronic medical record (EMR).[21]

---

16. Hope Yen, BlueCross BlueShield Association, Veterans Exposed to Incorrect Drug Doses, (Jan. 13, 2009), http://www.bcbs.com/news/national/veterans-exposed-to-incorrect-drug-doses.html.

17. Richard I. Cook & Michael F. O'Connor, *Thinking About Accidents and Systems*, *in* IMPROVING MEDICATION SAFETY 80, 80–82 (Kasey Thompson & Henri R. Manasse eds., 2005) (explaining that the problem was rooted in a backup tape that was incomplete and corrupted).

18. Tony Collins, *"Nearly Killed" by E-Records Data Model*, COMPUTERWEEKLY.COM, (May 21, 2009), http://www.computerweekly.com/Articles/2009/05/21/236128/nearly-killed-by-e-records-data-model.htm.

19. James C. Mohr, *American Medical Malpractice Litigation in Historical Perspective*, 283 J. AM. MED. ASS'N 1731, 1733–34 (2000); Mark F. Grady, *Why Are People Negligent? Technology, Nondurable Precautions, and the Medical Malpractice Explosion*, 82 NW. U. L. REV. 293, 297–301, 314–15 (1988) (explaining that many "believe that new technology adds to the number of negligence claims" and analyzing the reasons for this phenomenon).

20. *See infra* Part IV.

21. There is confusion in the literature about the terms EHR and EMR. For example, the HITECH Act defines an EHR as "an electronic record of health-related information on an individual that is created, gathered, managed, and consulted by authorized health care clinicians and staff." American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, 123 Stat. 115 (2009) (to be codified at 42 U.S.C. § 17921(5)). However, one commentator notes that the HITECH Act's definition of EHR is "confusingly . . . one that is generally associated with an EMR." Nicolas P. Terry, *Personal Health Records: Directing More Costs and*

With their wealth of capabilities, EHR systems are likely to raise the public's expectations concerning clinicians' performance and to affect the standard of care to which clinicians are held for medical malpractice purposes.[22] The systems make unprecedented volumes of information available to physicians.[23] With computers connecting them to a local, regional, and perhaps even national health information network,[24] doctors could have access to every detail of the patient's medical history from birth until the present time and be expected to consider all relevant information in their treatment decisions. EHR systems also provide doctors with sophisticated decision support tools,[25] which will raise the public's expectations concerning the quality of medical treatments. More common use of e-mail and secure messaging for patient-doctor communication and improved access to clinical data through personal health records[26] may further increase patient demands and expectations.

Physicians who have more complete records and better decision support and communication tools, but who do not have the time or skill to assimilate the unprecedented amount of available data and to optimize their use of technology, may face medical malpractice claims that would have never emerged in the past.[27] Clinicians who mishandle EHR systems and thereby cause injury to patients could also in rare cases face disciplinary action initiated by state licensing boards and even criminal prosecution.[28] Health care organizations such as hospitals may likewise face reaccreditation challenges and lawsuits based on vicarious liability and other negligence theories.[29]

Risks to Consumers?, 1 DREXEL L. REV. 216, 257 (2009).

22. *See infra* notes 69–79 and accompanying text for a discussion of medical malpractice and the standard of care.

23. Jha et al., *supra* note 6, at 1633 (discussing the various capabilities of comprehensive EHR systems).

24. American Recovery and Reinvestment Act § 3002(b)(1) (articulating the goal of establishing a "nationwide health information technology infrastructure that permits the electronic exchange and use of health information").

25. Jonathan A. Handler et al., *Computerized Physician Order Entry and Online Decision Support*, 11 ACAD. EMERGENCY MED. 1135, 1135–36 (2004).

26. *See* Paul C. Tang et al., *Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption*, 13 J. AM. MED. INFORMATICS ASS'N 121, 121 (2006) (explaining that personal health records provide a "repository for patient data," provide capabilities that "assist patients in managing chronic conditions," and generally allow individuals to be more active in their own health care).

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

28. *See infra* Section III.C.

29. *See infra* Sections III.A.1 & III.C.

In addition, computerization and electronic distribution of private health information could lead to privacy breach claims. Electronic data is vulnerable to improper disclosure through hacking, laptop theft, inadvertent disclosure, or deliberate leaks.[30] Once electronic information is accessed by unauthorized personnel, it can be rapidly distributed to a worldwide audience through the Internet, potentially causing humiliation, ruining careers, or causing other serious harms.[31]

This Article provides a first of its kind, comprehensive analysis of the liability risks associated with EHR systems, which may soon become a fixture in all medical settings. It considers the mandates of the Health Information Technology for Economic and Clinical Health Act (HITECH Act),[32] the portion of the ARRA that focuses on health information technology. Part II describes EHR systems and how they function in the contemporary medical practice setting. Part III analyzes new liability risks associated with EHR systems.

Part IV then formulates recommendations to address liability concerns. In particular, we argue that EHR systems, which are currently an unregulated technology,[33] must be regulated by the federal government in order to achieve quality control.[34] In addition, agency guidance and clinical practice guidelines should assist providers in optimizing EHR system use.[35] This Article explores how the standard of care should be established with respect to an emerging technology with a very limited use history. It proposes a new, uniform process for the development of clinical practice guidelines that is coordinated by a central professional organization and is based on field evaluation. It also suggests that EHR systems' own audit trails[36] and electronic search capabilities could contribute much to the formulation of sound guidelines concerning operating standards. Regulations, agency guidance, and widely accepted, authoritative clinical practice guidelines would all constitute admissible evidence of the standard of care and provide some degree of predictability for litigation purposes at the same time that they help clinicians maximize the benefits and minimize the risks of EHR system use.[37]

---

30. Sharona Hoffman & Andy Podgurski, *In Sickness, Health, and Cyberspace: Protecting the Security of Electronic Private Health Information*, 48 B.C. L. REV. 331, 332–34 (2007).

31. *Id.* at 332.

32. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13001, 123 Stat. 115, 226 (2009).

33. *See* Hoffman & Podgurski, *supra* note 13, at 126.

34. *See infra* Section IV.A.1.

35. *See infra* Section III.A.2 & IV.B.2.

36. *See infra* notes 336–39 and accompanying text.

37. *See infra* Section IV.B.

II. EHR SYSTEM ATTRIBUTES

An EHR can be defined as a "repository of electronically maintained information about an individual's lifetime health status and health care."[38] An EHR system is the "addition to an electronic health record of information management tools."[39] Comprehensive EHR systems provide a broad range of functions.[40] They assist providers in managing health information and data by displaying laboratory test results, allergies, lists of other medications the patient is taking, medical and nursing diagnoses, patient demographics, and providers' notes.[41] EHR systems also transmit results from laboratory tests, radiology procedures, and other diagnostic examinations electronically so that providers can quickly and efficiently access needed information.[42] Many systems allow clinicians to submit computerized medication orders and other care instructions, which can reduce or eliminate lost orders, duplicate orders, mistakes caused by illegible handwriting, and delays in filling orders.[43]

Of particular importance and complexity are EHR systems' decision support features. Automatic reminders and prompts can improve preventive care, diagnosis, treatment, and disease management.[44] For example, an EHR system can remind providers that a patient needs a vaccination or mammogram or that the patient is allergic to a medication that the doctor wishes to prescribe. More sophisticated systems might even analyze entered data and suggest appropriate diagnostic tests, diagnoses, or treatment plans.[45]

EHR systems can optimize connectivity and communication.[46] They can facilitate online communication among medical team members, between clinicians and other providers such as laboratories or pharmacies, and between caregivers and their patients. Communication can be achieved through e-mail, web messaging, integrated health records within and across treatment settings, telemedicine,[47] and home telemonitoring.[48] Once in place,

---

38. BIOMEDICAL INFORMATICS, *supra* note 12 at 937.

39. *Id.*

40. INSTITUTE OF MEDICINE, KEY CAPABILITIES OF AN ELECTRONIC HEALTH RECORD SYSTEM 7–9 (2003) [hereinafter KEY CAPABILITIES].

41. *Id.* at 7.

42. *Id.* at 7–8.

43. *Id.* at 8.

44. *Id.* at 8–9.

45. Handler et al., *supra* note 25, at 1135–36 (discussing systems that assist in diagnosis and therapeutic decisions).

46. KEY CAPABILITIES, *supra* note 40, at 9.

47. Telemedicine is "the delivery of health care at a distance, increasingly but not exclusively by means of the Internet." BIOMEDICAL INFORMATICS, *supra* note 12, at 991.

48. Home telemonitoring can be defined as "an automated process for the transmission of data on a patient's health status from home to the . . . health care setting."

an EHR system may become the primary means of communication among clinicians.

As stated in the Health Information Technology for Economic and Clinical Health (HITECH) Act, the federal government's goal is to achieve interoperability by building a "nationwide health information technology infrastructure that permits the electronic exchange and use of health information."[49] "Interoperability" means the ability of two or more systems to exchange data and to operate in a coordinated fashion.[50] With interoperability, authorized personnel would be able to access patient records regardless of where they are stored and by whom the patient was previously treated, including records compiled by providers in distant locations and other health care networks.[51] This capability would allow doctors to discover information about a new patient's medical history, drug lists, allergies, and other critical matters for which they currently must depend upon the patient's memory. Furthermore, emergency room personnel treating unconscious or uncommunicative patients would no longer need to operate in complete ignorance of crucial medical facts.[52] However, interoperability will dramatically expand the amount of information clinicians must read and consider in treating their patients. It will also increase the risk of inappropriate disclosure because individuals across the country may be able to access a patient's records.

One component of some EHR systems that is particularly appealing to patients is the personal health record (PHR). A PHR has been defined as "an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment."[53]

Guy Paré et al., *Systematic Review of Home Telemonitoring for Chronic Diseases: The Evidence Base*, 14 J. AM. MED. INFORMATICS ASS'N 269, 270 (2007).

49. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3002(b)(1), 123 Stat. 115, 234 (2009) (to be codified at 42 U.S.C. § 300jj-12(b)(1)).

50. BIOMEDICAL INFORMATICS, *supra* note 12, at 952.

51. Hoffman & Podgurski, *supra* note 13, at 112–13.

52. They could also have immediate access to important documents such as a living will or durable power of attorney for health care.

53. Tang et al., *supra* note 26, at 122 (citing MARKLE FOUNDATION, CONNECTING FOR HEALTH: THE PERSONAL HEALTH WORKING GROUP FINAL REPORT (2003), http://www.connectingforhealth.org/resources/final_phwg_report1.pdf; *see also* American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13400(11), 123 Stat. 115, 259 (2009) (to be codified at 42 U.S.C. § 17921(11)), (defining a PHR as an "electronic record of . . . health information . . . on an individual that can be drawn from multiple sources and that is managed, shared, and controlled by or primarily for the individual").

PHRs are found in various forms. Some are web pages that allow patients to enter their own health information; others are physician-provided patient portals that allow patients to access part or all of their EHRs; and still others are constructed by employers or insurers and enable patients to review their claims data.[54] For example, an independent vendor, Epic Systems, has developed a PHR called MyChart that allows patients to read their EHRs, including lab test results, and to communicate electronically with physicians through secure messaging, but it does not provide access to progress notes.[55] MyChart, which is integrated with an EHR system, is hosted by medical practices, and is used by 2.4 million U.S. patients, according to recent estimates.[56]

A second PHR model is PatientSite, a hospital-based system built by Beth Israel Deaconess Medical Center. PatientSite allows patients to access their lists of problems, medications, allergies, visit schedules, and laboratory and other diagnostic test results.[57] In addition, patients can add information to their PHRs, such as readings from home-administered tests, records of over-the-counter drugs that they take, and personal notes.[58] Furthermore, the system provides for secure messaging, automated appointment scheduling, prescription renewals, and specialist referrals.[59]

Still other PHRs are independent, personally-controlled products.[60] The patients decide who can review, write, or change these health records.[61] Such PHRs could exchange data with EHR systems or function as stand-alone records that are supplied on smart cards, CDs, or USB drives.[62]

Only a minority of medical practices currently use EHR systems to a significant extent. According to a recent survey, only 2.9 percent of U.S. hospitals have comprehensive EHR systems "across all major clinical units."[63] An additional 7.9 percent of hospitals have basic systems that include electronic clinicians' notes in at least one clinical unit, and 11.3 percent have basic systems that do not include electronic clinicians' notes.[64]

---

54. John D. Halamka, *Early Experiences with Personal Health Records*, 15 J. AM. MED. INFORMATICS ASS'N 1, 1 (Jan./Feb. 2008).

55. *Id.* at 1–2.

56. *Id.* at 1.

57. *Id.* at 2.

58. *Id.*

59. *Id.*

60. *Id.* at 2–3.

61. *Id.* (describing Indivo, a personally controlled health record).

62. *Id.* at 3; Tang et al., *supra* note 26, at 122.

63. This figure includes Veterans Health Administration hospitals. Jha, *supra* note 6, at 1631.

64. *Id.*

A basic system includes electronic notes concerning patient demographics, medical problem lists, medication lists, discharge summaries, laboratory reports, radiologic reports, and diagnostic test results, but excludes clinicians' notes.[65] Seventy-five percent of hospitals have adopted electronic laboratory and radiologic test result reporting, and seventeen percent have CPOE.[66] An earlier survey focusing on ambulatory care concluded that only four percent of physicians had comprehensive EHR systems, while thirteen percent had basic systems.[67]

New technologies have the potential to improve health outcomes and patient satisfaction dramatically. However, they may also create significant challenges for clinicians by generating increased workloads, unrealistic patient expectations, privacy breaches, and the likelihood of computer-related mishaps that endanger patient welfare. Consequently, health care providers may be faced with new litigation vulnerabilities that did not emerge during the era of paper medical records.[68]

III. LIABILITY CONCERNS

This section provides a comprehensive overview of the liability risks faced by EHR system users. Contemporary EHR system technology has significant limitations, and if these cause harm, aggrieved individuals and enforcement entities have many legal resources. Plaintiffs whose alleged injuries are associated with EHR systems could sue health care providers for medical malpractice. Those who believe that their records have been improperly disclosed to third parties could assert privacy violation claims. In rare circumstances, providers accused of negligent EHR system use could face disciplinary proceedings (initiated by professional organizations), government enforcement actions, or criminal prosecutions. Each of these potential claims and penalties will be addressed below.

A. MEDICAL MALPRACTICE CLAIMS

Patients who feel that their care givers were negligent in treating them may assert medical malpractice claims. To prevail, the plaintiff must establish

65. *Id.* at 1633.

66. *Id.* at 1631.

67. DesRoches et al., *supra* note 6, at 50, 54. The difference between basic and fully functional EHR systems is discussed *id.* at 52.

68. Shana Campbell Jones et al., *The Interoperable Electronic Health Record: Preserving Its Promise by Recognizing and Limiting Physician Liability*, 63 FOOD & DRUG L.J. 75, 81 (2008) (noting that physicians may eschew EHR system adoption if they are alarmed by the prospect of "expanded professional liability exposure").

the four elements of negligence:[69] (1) a duty of care owed by the defendant to the plaintiff, (2) breach of that duty through conduct that fails to meet the applicable standard of care, (3) harm or injury, and (4) a causal link between the injury and the breach of duty.[70] The standard of care in each case is determined based on an assessment of whether the defendant "proceed[ed] with such reasonable caution as a prudent man would have exercised under such circumstances."[71] Thus, in medical malpractice cases, plaintiffs must prove that "the professional failed to conform to the generally recognized and accepted practices in his profession."[72]

As evidenced by the phrase "accepted practices," medical malpractice jurisprudence establishes that the legal standard of care is determined by professional custom.[73] Deviation from custom can constitute conclusive proof of negligence.[74] Physicians are not required to provide *optimal* care in order to avoid liability, but rather they are required to provide the level of care that could ordinarily be expected.[75]

One further question is whether professional custom should be judged based on practices in a narrow geographic location, such as the defendant's own community, or whether the area of focus should be broader, perhaps even national.[76] Although early decisions adhered to a "strict locality" rule,

69. Eleanor D. Kinney, *Administrative Law Approaches to Medical Malpractice Reform*, 49 ST. LOUIS U. L.J. 45, 49 (2004).

70. PROSSER & KEETON ON THE LAW OF TORTS (W. Page Keeton et al. eds., 5th ed. 1984).

71. Vaughan v. Menlove, (1837) 132 Eng. Rep. 490, 492 (C.P.) (affirming a jury verdict for the plaintiff who was injured when a fire that began in the defendant's haystack burnt down his house).

72. Doe v. Am. Red Cross Blood Servs., 377 S.E.2d 323, 326 (S.C. 1989).

73. RESTATEMENT (THIRD) OF TORTS § 13 cmt. b (Proposed Draft No. 1, 2005) ("In professional-malpractice cases, the malpractice standard is to a significant extent defined in terms of professional standards and customs."); Michelle M. Mello, *Of Swords and Shields: The Role of Clinical Practice Guidelines in Medical Malpractice Litigation*, 149 U. PA. L. REV. 645, 654–58 (2001) (discussing the role of professional custom in standard of care analysis in medical malpractice cases).

74. RESTATEMENT (THIRD) OF TORTS, *supra* note 73, § 13 cmt. b; Mello, *supra* note 73, at 658. *Cf.* Hood v. Phillips, 554 S.W.2d 160, 165 (Tex. 1977) ("A physician who undertakes a mode or form of treatment which a reasonable and prudent member of the medical profession would undertake . . . shall not be subject to liability."); Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 612 (1998) ("Doctors who have followed customary medical procedure are not to be considered negligent."). *But see* Tim Cramm et al., *Ascertaining Customary Care in Malpractice Cases: Asking Those Who Know*, 37 WAKE FOREST L. REV. 699, 707–10 (2002) (discussing an "incipient trend towards modifying custom as conclusive" and moving closer to the traditional reasonable standard for negligence cases).

75. Cramm et al., *supra* note 74, at 702.

76. *Id.* at 705–07; BARRY R. FURROW ET AL., HEALTH LAW CASES, MATERIALS AND

most states currently follow a "similar locality" or national standard.[77] Interoperable EHR systems would make a national professional custom rule more sensible because records will be nationally accessible and transmittable and because decision support could provide clinicians across the country with state of the art information and support.[78] Interoperability would not preclude defendants from presenting evidence that they had more limited resources at their particular institutions since establishing the standard of care requires consideration of the specific circumstances at issue.[79]

1. Liability of Health Care Entities: Corporate Negligence and Vicarious Liability

Medical malpractice claims can be asserted against health care entities such as hospitals and clinics under the theories of corporate negligence and vicarious liability.[80] In corporate negligence cases, health care organizations can be held liable for failing to safeguard their patients' safety and welfare.[81] Hospitals have the following four duties:

> (1) a duty to use reasonable care in the maintenance of safe and adequate facilities and equipment; (2) a duty to select and retain only competent physicians; (3) a duty to oversee all persons who practice medicine within its walls as to patient care; and (4) a duty to formulate, adopt and enforce adequate rules and policies to ensure quality care for the patients.[82]

In establishing a prima facie case of corporate negligence, plaintiffs must show (1) that the hospital deviated from the standard of care; (2) that the hospital has actual or constructive knowledge of the flaws or procedures that

---

PROBLEMS 338–39 (6th ed. 2008).

77. Cramm et al., *supra* note 74, at 705–07; FURROW ET. AL., *supra* note 76, at 338 (explaining the concern that a strict locality rule would make it very difficult for plaintiffs to find expert witnesses because physicians would be reluctant to testify against colleagues in their own communities).

78. Cramm et al., *supra* note 74, at 706 (noting that the globalization of information supports moving away from a locality rule).

79. *See supra* note 71 and accompanying text; FURROW ET AL., *supra* note 76, at 338 (citing RESTATEMENT (SECOND) OF TORTS § 299A cmt. d (1965)). ("A country doctor cannot be expected to have the equipment, facilities, experience, knowledge or opportunity to obtain it, afforded him by a large city.").

80. Darling v. Charleston Cmty. Mem'l Hosp., 211 N.E.2d 253 (Ill. 1965) (recognizing a cause of action for corporate negligence); Alexander v. Mount Sinai Hosp. Medical Center, 484 F.3d 889, 903 (7th Cir. 2007) (explaining how plaintiff could sustain a medical malpractice claim against the hospital based on vicarious liability).

81. Thompson v. Nason Hosp., 591 A.2d 703, 707 (Pa. 1991).

82. *Id.*

caused the injury; and (3) that a causal link exists between the conduct and the harm.[83]

Organizations can also be held liable for the actions of their employees through the vicarious liability theories of respondeat superior and ostensible agency. The doctrine of "respondeat superior," which literally means "let the superior answer," establishes that employers are responsible for the acts of their employees in the course of their employment.[84] Thus, hospitals may be held liable for inappropriate EHR system uses by nurses, residents, interns, or other health professionals. However, in many instances, hospitals are shielded from liability for physicians' acts because physicians are considered independent contractors rather than employees.[85] Nevertheless, courts have found that a hospital's imposition of workplace rules and regulations upon staff physicians is enough to undercut the doctors' independent contractor status and expose the hospital to liability.[86] Therefore, hospitals that establish EHR-use protocols and policies may be responsible for clinicians' negligent operation of these systems.

An alternative theory of liability is ostensible agency. A hospital can be liable for an independent contractor's wrongdoing if the individual is deemed to be the hospital's "ostensible agent."[87] A court can find ostensible agency if (1) the patient looks to the entity rather than the specific physician for care, and (2) the hospital "holds out" the doctor as its employee.[88] The ostensible agency theory is particularly applicable to emergency room care, because patients generally seek medical treatment from emergency departments rather than from individual attending physicians.[89]

---

83. Rauch v. Mike-Mayer, 783 A.2d 815, 827 (Pa. Super. Ct. 2001).

84. BLACK'S LAW DICTIONARY 1338 (8th ed. 2004) (defining the term to mean that employers are responsible for the acts of their employees in the course of their employment).

85. *See, e.g.*, Kashishian v. Port, 481 N.W.2d 277, 280 (Wis. 1992) (holding that even though a physician was a member of the hospital's staff and was required to comply with hospital policies, no master-servant relationship existed); Albain v. Flower Hosp., 553 N.E.2d 1038, 1044 (Ohio 1990) (finding that the physician's staff privileges did not make the hospital vulnerable to respondeat superior liability for his actions).

86. Mduba v. Benedictine Hosp., 384 N.Y.S.2d 527, 529 (N.Y. App. Div. 1976) (finding that a physician was a hospital employee rather than an independent contractor because the hospital controlled the way he operated its emergency room); *see generally* Martin C. McWilliams, Jr. & Hamilton E. Russell III, *Hospital Liability for Torts of Independent Contractor Physicians*, 47 S.C. L. REV. 431 (1996).

87. *See* Simmons v. St. Clair Mem'l Hosp., 481 A.2d 870, 874 (Pa. Super. Ct. 1984).

88. *Id.*; Burless v. W. Va. Univ. Hosps., Inc., 601 S.E.2d 85, 95–96 (W. Va. 2004) (discussing ostensible agency theory and proof criteria).

89. *See* Torrence v. Kusminsky, 408 S.E.2d 684, 692 (W. Va. 1991) (stating that "where a hospital makes emergency room treatment available to serve the public as an integral part

Through corporate negligence or vicarious liability theories, health care entities could be held liable for injuries caused by equipment defects or by their employees' misuse of sophisticated technology. The remainder of this Part will focus largely on clinicians' use of EHR systems and the potential problems they might experience.

2. Clinician Liability

Use of EHR systems could generate negligence claims against providers for a variety of reasons. EHR system operation can be time-consuming and burdensome, and increased work demands could cause rushed physicians to make medical mistakes. Greater access to existing diagnostic data and economic pressures to avoid duplicating tests could lead to errors from inappropriate reliance on outdated or inadequate prior testing. Mistakes may also result from data entry errors. Clinicians may be faulted for ignoring critical prompts and alerts from decision support features. Furthermore, providers who do not thoughtfully handle communication tools such as e-mail and PHRs may face frustrated, anxious, and litigious patients. Finally, product defects that affect medication orders or alerts can cause serious harm to patients. This Section will carefully consider each of these potential liability sources.

a) Physician Time Constraints and Information Overload

The typical contemporary physician faces significant time pressures and extreme workload demands.[90] A common complaint is that EHR system use is time consuming and requires clinicians to process an impossible amount of information.[91] This challenge can lead to medical mistakes and liability exposure.

The average visit to a primary care physician lasts thirteen to eighteen minutes.[92] Doctors are not able to spend sufficient time with patients to

---

of its facilities, the hospital is estopped to deny that the physicians and other medical personnel on duty providing treatment are its agents" and that "[r]egardless of any contractual arrangements with so-called independent contractors, the hospital is liable to the injured patient for acts of malpractice committed in its emergency room, so long as the requisite proximate cause and damages are present").

90. *See infra* notes 92–97 and accompanying text.

91. *See infra* notes 100–04and accompanying text.

92. Andrew Gottschalk & Susan A. Flocke, *Time Spent in Face-to-Face Patient Care and Work Outside the Examination Room*, 3 ANNALS FAM. MED. 488, 491 (2005) (finding that the average time per patient was 13.3 minutes); Kimberly S. H. Yarnall et al., *Family Physicians as Team Leaders: "Time" to Share the Care*, PREVENTING CHRONIC DISEASE: HEALTH RES., PRAC., & POL'Y 1, 6, Apr. 2009, http://www.cdc.gov/pcd/issues/2009/apr/08_0023.htm (finding that the mean length for an acute care visit is 17.3 minutes, the mean for a chronic disease

provide the comprehensive preventive and chronic disease care that is recommended in clinical practice guidelines.[93] In addition, physicians spend up to forty-five percent of their time each day attending to tasks outside of the examination room, such as reviewing charts, completing forms, writing prescriptions, consulting colleagues, and answering staff inquiries.[94] In a 2008 survey of approximately 11,950 physicians, over forty percent indicated that they saw between twenty-one and thirty patients per day, and over seventy-five percent described their practices as either at "full capacity" or "overextended and overworked."[95] If these responses are representative,[96] most physicians would find it very difficult to accommodate additional work in their already crowded schedules.[97]

It is also unlikely that physicians will decrease the number of patients they see in order to address time pressures. The United States is facing a

---

care visit is 19.3 minutes, and the average for a preventive care visit is 21.4 minutes, and that of total clinical time spent by physicians, these comprise 45.8%, 37.4%, and 16.8% respectively); Kevin Fiscella & Ronald M. Epstein, *So Much to Do, So Little Time: Care for the Socially Disadvantaged and the 15-Minute Visit*, 168 ARCH. INTERNAL MED. 1843, 1843 (2008) ("The average office visit in the United States lasts for about 16 minutes."); Chen, *supra* note 11, at 329 (reporting that the average time spent by patients with providers during 1998–2008 was 16.4 minutes).

93. Fiscella, *supra* note 92, at 1843–44; Truls Østbye et al., *Is There Time for Management of Patients with Chronic Diseases in Primary Care?*, 3 ANNALS FAM. MED. 209, 212 (2005) ("We calculated that comprehensive high-quality management of 10 common chronic diseases require more time than primary care physicians have available for all patient care."); Yarnall et al., *supra* note 92, at 1 ("The common denominator in the failure to deliver services is probably lack of physician time."). For a discussion of clinical practice guidelines *see infra* Section IV.B.2.a).

94. Gottschalk & Flocke, *supra* note 92, at 490–91; Jeffrey Farber et al., *How Much Time Do Physicians Spend Providing Care Outside of Office Visits?*, 147 ANNALS INTERNAL MED. 693, 695–97 (2007).

95. THE PHYSICIANS' FOUNDATION, THE PHYSICIANS' PERSPECTIVE: MEDICAL PRACTICE IN 2008: SURVEY SUMMARY & ANALYSIS, 4 (2008), http://www.physicians foundations.org/usr_doc/PF_Survey_Report.pdf. More specifically, the number of patients per day seen by physicians was as follows: 7.4% saw 0–10; 31.71% saw 11–20; 41.28% saw 21–30; 13.68% saw 31–40; 3.71% saw 41–50; 0.99% saw 51–60, and 1.23% saw over 61. In describing their practices, 44.92% indicated that they were at full capacity; 31.37% were "overextended and overworked;" and 23.71% indicated that they "[h]ave time to see more patients and assume more duties." *See also* Gottschalk & Flocke, *supra* note 92, at 491 (finding that the "mean number of patients seen per day was 29.1" in a survey of eleven primary care physicians who did not use EHRs).

96. THE PHYSICIANS' FOUNDATION, *supra* note 95, at 4. A major limitation of the study is that the response rate was only four percent. *Id.* at 4. It is possible that the respondents are a self-selected group of individuals who felt particularly pressured or unhappy. Nevertheless, the report is based on answers from 11,950 physicians, which is not an insignificant number. *Id.*

97. Yarnall et al., *supra* note 92, at 1; Østbye et al., *supra* note 93, at 212.

shortage of primary care physicians,[98] so fewer doctors are available to treat a growing U.S. population. In addition, financial incentives discourage doctors from reducing the number of patients they see, and decreasing Medicare, Medicaid, and private insurance reimbursements may threaten the economic viability of some practices and require them to maintain a high volume of patient visits.[99]

EHR systems impose new demands on physicians' workdays.[100] They require clinicians to type text directly into the EHR, a task that is disfavored by some providers.[101] According to one study, using bedside or examination room computers increased physician documentation time by 17.5 percent while using centrally located desktops for CPOE rather than prescription pads increased physician documentation time by 98.1 percent to 328.6 percent.[102] Typing visit notes in accordance with EHR specifications generally takes longer than dictating notes or writing a succinct visit summary by hand.[103] EHR systems have templates that require physicians to record far more information than they have traditionally included in paper files, and not all of the information is essential or even relevant to proper patient care.[104]

---

98. THE PHYSICIANS' FOUNDATION, *supra* note 95, at 10 (reporting that 78% of physicians believe there is a shortage of primary care physicians); Kevin Grumbach & Thomas Bodenheimer, *A Primary Care Home for Americans: Putting the House in Order*, 288 J. AM. MED. ASS'N 889, 890 (2002) (stating that "primary care is endangered" because fewer medical school graduates are choosing to become primary care physicians and to practice internal medicine).

99. THE PHYSICIANS' FOUNDATION, *supra* note 95, at 3 (discussing declining Medicare and Medicaid reimbursement); Yarnall et al., *supra* note 92, at 1 (noting the problem of "inadequate insurance reimbursement"); Ming Tai-Seale et al., *Time Allocation in Primary Care Office Visits*, 42 HEALTH SERVS. RES. 1871, 1886 (2007) ("Incentives in prevailing physician payments favor procedure-based patient care over time-intensive evaluation and management care."); Leigh Ann Backer, *Strategies for Better Patient Flow and Cycle Time*, 9 FAM. PRAC. MGMT. 45 (2002), *available at* http://www.aafp.org/fpm/20020600/45stra.html (noting reduced Medicare and private insurance reimbursement and offering recommendations to maximize patient flow and cycle time in family medicine practices); Aris Sophocles, *Time Is of the Essence: Coding on the Basis of Time for Physician Services*, 10 FAM. PRAC. MGMT. 27, 27 (2003) (explaining that "CPT [current procedural terminology] lists a variety of codes that are strictly time dependent").

100. Thomas Bodenheimer, *Innovations in Primary Care in the United States*, 326 BRIT. MED. J. 796, 798 (2003) (asserting that EHR systems impose "extra demands on physicians' time").

101. C.R. Weir et al., *Direct Text Entry in Electronic Progress Notes*, 42 METHODS INF. MED. 61, 61 (2003).

102. Lise Poissant et al., *The Impact of Electronic Health Records on Time Efficiency of Physicians and Nurses: A Systematic Review*, 12 J. AM. MED. INFORMATICS ASS'N 505, 508 (2005).

103. Baron et al., *supra* note 14, at 225; Weir et al., *supra* note 101, at 66.

104. Weir et al., *supra* note 101, at 65 (noting that templates may be up to 5 pages in length); Anne Armstrong-Coben, *The Computer Will See You Now*, N.Y. TIMES, Mar. 5, 2009, at A27 (asserting that the EHR system requires her "to bring up questions in the order they

Time spent on EHR-related tasks is time not spent interacting directly with patients.[105] Physicians who have fewer minutes to speak with and examine patients may provide lower quality care. In addition, patients may resent the doctor's focus on the computer and apparent inattention to them[106] and be more apt to sue if they are dissatisfied with their health outcomes. This concern is not theoretical. Multiple studies have shown that patients most often decide to sue when they are displeased with the quality of the physician-patient relationship and feel they cannot communicate well with their doctors.[107]

Computerized records can be lengthy and cumbersome to read. Whereas having to write notes by hand encourages brevity, physicians entering notes electronically may copy large segments of information from elsewhere in the record for the sake of completeness.[108] But this practice may make it far more difficult for a provider to obtain an overview of the patient's current condition or locate a needed detail quickly.[109] With interoperability,[110] doctors may have access to records from patients' visits to numerous specialists and be expected to consider all relevant information concerning each patient's medical and treatment history.[111] The challenges of reviewing a patient's entire EHR may be compounded by data display problems. Doctors may need to scroll through numerous screens in order to find the detail they seek, information may be organized awkwardly or fragmented throughout the EHR, and all data might appear in a uniform format so that physicians seeking a particular fact cannot easily scan the data.[112]

The challenges posed by the large volumes of information contained in interoperable EHRs could be addressed in part through the work of nurses or other lower-cost providers who meet with the patient at the beginning of

---

appear [and] to ask the parents of a laughing 2-year-old if she is 'in pain' ").

105. Armstrong-Coben, *supra* note 104, at A27 (explaining that the computer interferes "with what should be going on in the exam room—making that crucial connection between doctor and patient").

106. Baron et al., *supra* note 14, at 224 (reporting that after EHR system implementation, some patients asked, "Doctor, do you find you are spending more time interacting with the computer than with your patients?").

107. Beth Huntington & Nettie Kuhn, *Communication Gaffes: A Root Cause of Malpractice Claims*, 16 BAYLOR U. MED. CENTER PROC. 157, 157–60 (2003) (reviewing studies that explore the circumstances in which patients decide to sue their physicians).

108. Weir et al., *supra* note 101, at 66.

109. Armstrong-Coben, *supra* note 104, at A27 ("In the past, I could pick up a chart and flip through it easily . . . . Now . . . important points often get lost.").

110. BIOMEDICAL INFORMATICS, *supra* note 12, at 952.

111. *See* Hoffman & Podgurski, *supra* note 13, at 112–13.

112. Ross Koppel, *Role of Computerized Physician Order Entry Systems in Facilitating Medication Errors*, 293 J. Am. Med. Ass'n 1197, 1199–1201 (2005).

the appointment. These providers could review the EHR and interview the patient before the doctor enters the examination room and then supply the physician with notes or a verbal report summarizing and highlighting the most relevant information. This approach, while potentially helpful, would raise issues of vicarious liability for physicians. Under the doctrine of respondeat superior, doctors who directly supervise and control staff members may be held liable for injuries associated with inaccurate or deficient summary reports provided by office personnel.[113]

Case law establishes that physicians can be held liable for harm that could have been averted had they more carefully studied their patients' medical records. For example, *Short v. United States* involved a patient whose doctor failed to diagnose his prostate cancer in time for it to be cured.[114] The court held that under Vermont law, the physician violated the standard of care by failing to review the patient's past visit notes, which would have elucidated the nature of his condition.[115] In *Conrad-Hutsell v. Colturi*, a court of appeals reversed a directed verdict for the defendant.[116] The court held that a question of fact existed as to whether a physician, who did not obtain a copy of the patient's medical record that would have indicated a history of narcotics overuse, should be held liable for the patient's addiction to the drugs he prescribed.[117]

With EHR systems, clinicians may find it extremely difficult to process the plethora of information that floods their computer screens.[118] Yet those who miss a critical detail, such as a past illness treated by a different specialist that might affect the doctor's therapeutic decision, could be held liable for negligence because the fact in question was likely just a few clicks away when the physician was reviewing the patient's EHR.[119] The demands of EHR

113. Carol R. Goforth, *Limiting the Liability of General Partners in LLPs: An Analysis of Statutory Alternatives*, 75 OR. L. REV. 1139, 1201–13 (1996) (discussing the respondeat superior doctrine and its application to medical malpractice cases); Franklin v. Gupta, 567 A.2d 524, 537 (Md. Ct. Spec. App. 1990) (explaining that a physician can be held liable if "the negligent actors were, in fact, under his direct supervision and control"); Harris v. Miller, 438 S.E.2d 731, 741 (N.C. 1994) (holding that the defendant physician "enjoyed authoritative control" over a nurse anesthetist who performed his job duties negligently during surgery and that the trial court erred in "refusing to submit plaintiff's vicarious liability claim to the jury").

114. Short v. United States, 908 F. Supp. 227, 231–33 (D. Vt. 1995) (explaining that the patient required a bilateral orchiectomy and was not expected to survive for long).

115. *Id.* at 236.

116. No. L-01-1227, 2002 WL 1290844 (Ohio Ct. App. May 24, 2002).

117. *Id.* at 1–2.

118. Armstrong-Coben, *supra* note 104, at A27 (stating that EHRs present "screens filled with clicked boxes," that all information is provided in the same font, and that "important points often get lost").

119. EHR systems may also make discovery more burdensome and complicated than it

system operation and the very large amounts of information that users could be expected to consider may thus lead to malpractice liability.

b) Reliance on Others' Diagnosis and Treatment Decisions

Interoperability could raise another malpractice challenge as well by providing clinicians with incentives to rely on prior tests results. Currently, patients who transfer to a new doctor or seek a second opinion may be subjected to the same battery of tests that they have already undergone elsewhere.[120] With interoperability, authorized clinicians will have direct access to the results of all prior diagnostic tests and procedures, no matter where they were conducted. In light of "government and private studies [that] have found that much of the $2.5 trillion spent on health care each year

---

was in the past. Rule 34 of the Federal Rules of Civil Procedure allows parties to request to inspect, copy, test, or sample any electronically stored information, including e-mail, image files, and material from databases. Furthermore, the producing party must present the requested data in a reasonably usable form. FED. R. CIV. P. 34 (a)–(b).

EHRs may be difficult to produce and review because they are voluminous, especially if they are interoperable and contain records from all of the patient's treating physicians, laboratories, radiologists, and other providers. In addition, their format might make them abstruse to those not carefully trained in the system because of fragmented displays and other usability problems. EHRs may also generate unique authentication problems. User access, computer programming changes, backup systems, inputs, and other aspects of EHR system operation must all be carefully controlled in order to safeguard the integrity and authenticity of all medical records. Kevin Brady et al., *E-Discovery in Healthcare & Pharmaceutical Litigation: What's Ahead for ESI, PHI & EHR?*, 9 SEDONA CONF. 167, 174–75 (2008). In addition, the integrity of EHRs could be compromised during the discovery process itself because of inappropriate search and retrieval procedures, data conversion or other forms of mishandling. *Id.* at 174–75; *In re* Vinhee, 336 B.R. 437, 444 (9th Cir. 2005) (discussing authenticity and explaining that "the record being proffered must be shown to continue to be an accurate representation of the record that originally was created"). Thus, responding to document requests involving EHRs could be time-consuming, cumbersome, and costly. *See generally* Cecily Walters, *Attorney Survey Reveals Concerns About Litigation* Costs, TRIAL, Feb. 2009, at 64 (reporting that in responding to a survey of fellows of the American College of Trial Lawyers, "more than 87 percent said that e-discovery increases litigation costs, and almost 77 percent indicated that courts 'do not understand the difficulties in providing e-discovery.' "). *But see* Thomas R. McLean, *EMR Metadata Uses and E-Discovery*, 18 ANNALS HEALTH L. 75, 109 (2009) (explaining that because medical malpractice actions often require only the records of one patient or a few patients, the volume of documents involved in e-discovery may not be significantly greater than the amount involved in "traditional paper discovery").

120. CONG. BUDGET OFFICE, EVIDENCE ON THE COSTS AND BENEFITS OF HEALTH INFORMATION TECHNOLOGY 11 (2008), *available at* http://www.cbo.gov/ftpdocs/91xx/doc9168/05-20-HealthIT.pdf (discussing the potential for duplicated testing); Jan Walker et al., *The Value of Health Care Information Exchange and Interoperability*, HEALTH AFF., Jan. 19, 2005, at W5-10, W5-13-14, http://content.healthaffairs.org/cgi/content/full/hlthaff.w5.10/DC1 (discussing redundant testing).

is wasted on the duplication of tests and unneeded procedures,"[121] providers will likely be under considerable pressure to avoid repeating tests in order to achieve cost savings.[122] However, reliance on prior test results can lead to misdiagnoses or sub-optimal treatment decisions. For example, a technician who was sloppy or not sufficiently skilled may have conducted the prior test, or the patient's condition could have changed in the intervening time.[123]

One study of one hundred cases involving diagnostic errors determined that eight were caused by "[o]verreliance on someone else's finding or opinion" and failure to verify other clinicians' diagnoses in light of current findings.[124] Such mistakes have led to litigation and large plaintiff recoveries. For example, in *Whitaker v. Frankford Hospital*, a patient suffered a massive stroke after being discharged by an emergency room doctor who relied on a radiologist's interpretation of an MRA/MRI that erroneously indicated only a "very low percentage of blockage" in the carotid arteries.[125] Both physicians were among the defendants, and the plaintiff ultimately recovered millions of dollars through a settlement with some defendants and a jury verdict against others.[126] Because interoperable EHR systems would provide easy access to previously gathered medical data, problematic reliance on other clinicians' findings may become increasingly common.[127]

---

121. Robert O'Harrow Jr., *The Machinery Behind Health-Care Reform*, WASHINGTONPOST.COM, May 16, 2009, http://www.washingtonpost.com/wp-dyn/content/article/2009/05/15/AR2009051503667.html (suggesting that EHR systems could diminish the waste generated by the duplication of tests).

122. *Id.*; CONG. BUDGET OFFICE, *supra* note 120, at 11 (discussing the avoidance of duplicate or inappropriate diagnostic tests); Rainu Kaushal et al., *Return on Investment for a Computerized Physician Order Entry System*, 13 J. AM. MED. INFORMATICS ASS'N 261, 263 tbl. 1 (2006) (discussing the financial benefits of EHR systems, including decreased laboratory tests and radiology utilization); Walker et al., *supra* note 120, at W5-16 ("Interoperability between . . . organizations would enable computer-assisted reduction of redundant tests.").

123. R. James Brenner et al., *Radiology and Medical Malpractice Claims: A Report on the Practice Standards Claims Survey of the Physician Insurers Association of America and the American College of Radiology*, 171 AM. J. ROENTGENOLOGY 19, 20–21 (1998) (discussing the association between diagnostic errors and poor image quality in various radiological tests); E. James Potchen & Mark A. Bisesi, *When Is It Malpractice to Miss Lung Cancer on Chest Radiographs?*, 175 RADIOLOGY 29, 30 (1990) (stating that "poor image quality alone may be a source of negligence").

124. Mark L. Graber et al., *Diagnostic Error in Internal Medicine*, 165 ARCHIVES INTERNAL MED. 1493, 1497 (2005).

125. Nos. 1557, 2007 Phila. Ct. Com. Pl. LEXIS 287, at *2 (Pa. C.P. 2007).

126. *Id.* at *1–4.

127. It should be noted, however, that in some cases, conducting repeated tests is not in the patient's best interest. This would be true if the initial results are accurate, and the test is very uncomfortable or exposes the patient to risk such as radiation, or if the second diagnostic procedure shows different, incorrect results upon which the doctor may erroneously rely.

Physicians will thus face difficult decisions regarding whether to re-order expensive tests to verify diagnoses. They will need to continue to balance the competing interests of patient welfare, liability risks, and cost savings.

c) Input Errors

While paper files may contain illegible handwriting, misspellings, or other errors, use of automated technology may exacerbate the problem of record inaccuracies.[128] A study of sixty patient records with 1,891 notes from the Department of Veterans Health Administration's Computerized Patient Record System (CPRS) found that eighty-four percent of notes contained "at least one documentation error," and there were an average of 7.8 documentation mistakes per patient.[129] For example, cut and paste functions are designed to save doctors time by allowing them to copy information from old clinical notes into new progress notes. If such notes are not carefully edited, old symptoms, vital signs, or test results can appear to be current, and such mistakes can create new threats to patient safety and liability exposure for clinicians.[130]

A number of other problems can also arise because of careless clinician data entry. Occasionally, notes are entered into the wrong patient's record, and such erroneous information may mislead subsequent providers who consult an EHR.[131] In one reported incident, an "AIDS patient was wrongly told he had skin cancer on his neck because a test result for another patient was associated with his electronic record."[132] Likewise, physicians may hit the wrong key or inadvertently read the wrong patient's electronic record and thus base a treatment decision on incorrect information. In addition, users utilizing electronic signatures often neglect to indicate their titles or credentials.[133] This omission could be significant in a hospital setting, where

128. Weir et al., *supra* note 101, at 61.

129. *Id.* at 62, 64.

130. *Id.* at 64–65; Kenric W. Hammond et al., *Are Electronic Medical Records Trustworthy? Observations on Copying, Pasting and Duplication*, AMIA 2003 SYMP. PROC. 269, 269, 272 (2003), *available at* http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1480345&blobtype=pdf (reviewing 243 VA patient files and finding that 9% of notes contained copied text); Eugenia L. Siegler & Ronald Adelman, *Copy and Paste: A Remediable Hazard of Electronic Health Records*, 122 AM. J. MED. 495, 495–96 (2009) (cautioning that cut and paste functions can lead to patient problem lists never changing, notes and errors being copied by multiple staff members, and loss of accurate narrative).

131. Weir et al., *supra* note 101, at 65 (finding five instances out of 1,891 in which narrative notes were typed for the wrong patient).

132. Jacob Goldstein, *Big Challenges Await Health-Records Transition*, WALL ST. J., April 21, 2009, at A4.

133. Weir et al., *supra* note 101, at 65 (finding that 53% of electronic signatures "failed to appropriately reflect the credentials and/or title of the author").

care coordinators need to determine whether a patient was visited by a particular type of clinician or whether a specific treatment decision was made at the appropriate authority level.

Providers' reliance on electronic systems to order medication and other treatments is another novel source of medical mistakes.[134] One study of a hospital's CPOE system found that it posed the following challenges, which could lead to incorrect user input and consequent dosage errors:

> (1) Cumbersome medication charting and fragmented displays make it difficult to identify the patient to whom a particular record belongs or require doctors to look at numerous screens in order to obtain the patient's full medication list;
>
> (2) Physicians may fail to enter discontinuation orders for particular drugs when they change patients' medications so that the pharmacy continues to provide the old drugs as well as the new ones;
>
> (3) Problematic log-off procedures cause physicians to order medications on the system before the previous user has fully logged out, resulting in the wrong patient receiving the newly-ordered therapy;
>
> (4) The system requires that drug orders be reactivated rather than automatically transferred when patients are moved within the hospital (e.g. from the intensive care unit to a regular hospital room) so that patients whose doctors fail to reactivate orders are deprived of needed medications; and
>
> (5) System inflexibilities significantly impede providers' ability to enter nonstandard specifications or to order non-formulary medications.[135]

Medication errors and other mistakes involving CPOE functionality could thus lead to medical malpractice litigation and physician liability if they harm patients.

d) The Challenges of Decision Support

Decision support, defined as "any information added by a system to assist the clinician's decision-making process,"[136] can come in many forms.

---

134. Joan S. Ash et al., *Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors*, 11 J. AM. MED. INFORMATICS ASS'N 104, 106 (2004) (discussing errors relating to entering and retrieving information as well as communication and coordination problems).

135. Koppel, *supra* note 112, at 1199–1201. Non-formulary medications are "[d]rugs not on a [health care] plan-approved drug list." Medicare Glossary Definitions, http://www.medicare.gov/Glossary/search.asp?SelectAlphabet=N&Language=English#Content.

136. Handler et al., *supra* note 25, at 1135.

These include prompts based on clinical practice guidelines, clinical alert systems that warn providers about problems such as drug allergies and drug interactions, data tags that elucidate test results (such as an "L" next to a low laboratory value), and recommendations for diagnostic tests and treatment modalities based on patients' symptoms and conditions.[137] Although decision support has the potential to improve the quality of health care, it can also be disruptive in some circumstances. Furthermore, evidence that a doctor ignored automated alerts or recommendations may serve as compelling proof of physician wrongdoing for plaintiffs who suffer poor outcomes because of a doctor's treatment decision.

Studies have shown that decision support can appreciably improve patient care. One study found that reminders can significantly increase the use of preventive measures such as pneumococcal and influenza vaccinations in hospitalized patients.[138] Several other articles confirm the usefulness of decision support for preventive care purposes.[139]

Other researchers, however, have found that decision support is frequently disregarded.[140] According to one article, physicians often ignored suggestions concerning disease management because they distrusted them, did not appreciate a computer telling them how to practice medicine, or were too busy to consider computerized recommendations carefully.[141] Another study found that physicians did not follow suggestions because they could be

137. *Id.* at 1135–36.

138. Paul R. Dexter et al., *A Computerized Reminder System to Increase the Use of Preventive Care for Hospitalized Patients*, 345 NEW ENG. J. MED. 965, 968 (2001) (stating that with reminders, the use of pneumococcal vaccination increased from approximately zero to approximately 35%, and the use of influenza vaccinations increased from approximately zero to approximately 50% in the hospital).

139. Clement J. McDonald et al., *The Regenstrief Medical Record System: a quarter century experience*, 54 INT'L J. MED. INFORMATICS 225, 247 (1999) (asserting that "[r]eminders increased the use of preventive interventions up to four-fold," including use of influenza vaccines, mammography, and cervical pap testing); Alex R. Kemper et al., *Adoption of Electronic Health Records in Primary Care Pediatric Practices*, 118 PEDIATRICS e20, e23 (2006) (stating that "[a]lthough prompts for preventive services can improve care, many of the EHRs in use do not provide this feature").

140. Amit X. Garg et al., *Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes*, 293 J. AM. MED. ASS'N 1223, 1231 (2005) (stating that the systems' effects on patient outcomes are not sufficiently studied and are inconsistent when they are examined); Handler et al., *supra* note 25, at 1136 (stating that the benefit of decision support during documentation is unclear and often does not seem to affect clinicians' adherence to recommended guidelines).

141. Usha Subramanian et al., *A Controlled Trial of Including Symptom Data in Computer-Based Care Suggestions for Managing Patients with Chronic Heart Failure*, 6 AM. J. MED. 375, 379 (2003).

erased without being read if the user hit the escape key.[142] However, when the escape key was disabled, provider adherence to suggestions increased significantly.[143] Some postulate that providers might be resistant to decision support concerning disease management but receptive to suggestions concerning preventive care,[144] which may be perceived as less challenging to their professional judgment.

At times, it is medically appropriate for doctors to discount decision support messages. In many instances, decision support prompts and alerts can be excessive and disruptive and, therefore, justifiably overridden.[145] For example, drug-allergy alerts often indicate merely that some patients are sensitive to the medication even though they will suffer no serious reaction, and alerts continue to appear even if the patient has tolerated a medication well.[146] Drug-allergy alerts often do not distinguish between warnings of high clinical significance and the much more routine notices of benign drug sensitivities, so that all alerts are provided in the same format and color.[147] Researchers have found that doctors accept fewer than twenty percent of drug-allergy alerts, and almost all overrides are medically appropriate and do not risk significant harm to patients.[148] However, a doctor who is accustomed

142. William M. Tierney, *Can Computer-Generated Evidence-Based Care Suggestions Enhance Evidence-Based Management of Asthma and Chronic Obstructive Pulmonary Disease? A Randomized, Controlled Trial*, 40 HEALTH SERVS. RES. 477, 491 (2005).

143. Dexter et al., *supra* note 138, at 968.

144. Subramanian et al., *supra* note 141, at 379; Tierney, *supra* note 142, at 491–92; INSTITUTE OF MEDICINE, CROSSING THE QUALITY CHASM: A NEW HEALTH SYSTEM FOR THE 21ST CENTURY (2001). (finding convincing evidence that decision support improves preventive care, patient monitoring, and appropriate drug prescriptions but a dearth of convincing evidence for its usefulness for diagnosis and disease management).

145. Saeid Eslami et al., *Evaluation of Outpatient Computerized Physician Medication Order Entry Systems: A Systematic Review*, 14 J. AM. MED. INFORMATICS ASS'N 400, 404 (2007). (concluding that alerts are "largely ignored by physicians" but that many "alerts are not applicable to the patient at hand" or "are not clinically important"); Gilad J. Kuperman et al., *Medication-related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review*, 14 J. AM. MED. INFORMATICS ASS'N 29, 30 (2007) ("Excessive drug-allergy alerting in clinically irrelevant circumstances is highly prevalent and a major disruptor of clinicians' workflows.").

146. Kuperman et al., *supra* note 145, at 404.

147. *Id.*

148. *Id.* (citing Susan A. Abookire et al., *Improving Allergy Alerting in a Computerized Physician Order Entry System*, PROC. AM. MED. INFORMATICS ASS'N SYMP. 2, 2–6 (2000), *available at* http://www2.amia.org/pubs/symposia/D200703.PDF; Tyken C. Hsieh, *Characteristics and Consequences of Drug Allergy Alert Overrides in a Computerized Physician Order Entry System*, 11 J. AM. MED. INFORMATICS ASS'N 482 (2004).

to overriding alerts may become desensitized to them and occasionally ignore a critical one.[149]

Yet despite such practices, proof that a physician overrode or ignored an alert may constitute powerful evidence of wrongdoing for injured plaintiffs in litigation. In *Jones v. Bick*, the court found that a doctor failed to meet the standard of care when he did not consider warnings contained in the Physicians' Desk Reference (PDR) concerning the anti-psychotic drug prescribed to a patient who subsequently died of cardiac arrest.[150] It is even more likely that a physician would be found liable in similar circumstances if he did not have to use a reference book such as the PDR, but rather had a warning appear on his computer screen.

Like physicians, health care entities can be sued for ignoring CPOE warnings. Already, several pharmacies have been sued for failing to contact physicians to inform them of prescription problems of which they were made aware by electronic alerts. In *Cafarelle v. Brockton Oaks CVS, Inc.*, the court denied summary judgment to a pharmacy that overrode warning prompts and filled a child's Proventil inhaler prescriptions three times more often than was appropriate.[151] *In Happel v. Wal-Mart Stores*, the court found that the pharmacy had a duty to warn the patient's physician that the drug he prescribed was contraindicated for his patient because she was allergic to aspirin.[152] The store routinely entered patients' allergy information into its computer and had allergy warnings appear when prescriptions were filled.[153] These are likely the first of many cases involving CPOE.

Decision support is designed to help clinicians achieve optimal outcomes. However, it may at times be disruptive and distracting, and it could create records of prompts and alerts that increase the risk of liability for health care providers.

---

149. Peter A. Gross & David W. Bates, *A Pragmatic Approach to Implementing Best Practices for Clinical Decision Support Systems in Computerized Provider Order Entry Systems*, 14 J. AM. MED. INFORMATICS ASS'N 25, 26 (2007) (speculating that users might "ignore the most critical interaction alerts due to 'information overload' or 'inability to recognize the needle in the haystack' ").

150. Jones v. Bick, 891 So. 2d 737, 746 (4th Cir. 2004); *see also* Fournet v. Roule-Graham, 783 So. 2d 439 (5th Cir. 2001) (affirming judgment for plaintiff who accused her physician of negligence based on his prescribing Provera despite a warning in the PDR that the drug should not be given to a patient with a history of deep vein thrombosis).

151. Cafarelle v. Brockton Oaks CVS, Inc., 5 Mass. L. Rep. 257, 257 (1996).

152. Happel v. Wal-Mart Stores, Inc., 766 N.E.2d 1118, 1121, 1125, 1128 (Ill. 2002).

153. *Id.*

e) Responsiveness to Electronic Communication

EHR systems may allow patients to communicate with physicians through secure messaging that authenticates recipients and encrypts text.[154] Such communication, however, can lead to further liability concerns if doctors do not instruct patients to avoid e-mail use when immediate care is necessary and do not limit patient expectations concerning this service. Electronic communication can increase clinicians' accessibility and decrease the need for telephone calls and ambulatory care visits as clinicians address patients' health concerns through e-mail.[155] Early evidence reveals a high level of patient satisfaction with e-mail communication.[156] Nevertheless, online messaging creates a new setting in which physicians must avoid mistakes or risk liability.[157] Doctors must determine whether to ask the patient to come to the office for a physical examination or to offer medical advice without an in-person visit. Similarly, doctors or their staff members must check e-mail

154. Chen et al., *supra* note 11, at 325 (describing Kaiser Permanente Hawaii's My Health Manager, a secure patient-physician messaging system through which members sent over 51,000 messages in 2007); Steven E. Waldren, *Email in Clinical Care*, 4 BMJ USA E325, E325 (2004), *available at* http://www.bmj.com/cgi/reprint/329/7471/E325 ("To ensure confidentiality, the recipient (patient) must be authenticated and the message itself must be transmitted in an encrypted manner.").

155. Chen et al., *supra* note 11, at 327 (finding a 26.2% percent reduction in the yearly total office appointment over 2004–2007, with face-to-face contact replaced by scheduled telephone visits and secure messaging); Madhavi R. Patt et al., *Doctors Who Are Using E-mail with Their Patients: A Qualitative Exploration*, J. MED. INTERNET RES. Apr.-Jun. 2003, http://www.jmir.org/2003/2/e9/ (stating that some physicians believed that e-mail would increase their accessibility to patients); Paul Rosen & C. Kent Kwoh, *Patient-Physician E-mail: An Opportunity to Transform Pediatric Health Care Delivery*, 120 PEDIATRICS 701, 704 (2007) (reporting that it took physicians 57% less time to respond to e-mail than to answer telephone calls); Yi Yvonne Zhou et al., *Patient Access to an Electronic Health Record with Secure Messaging: Impact on Primary Care Utilization*, 13 AM. J. MANAGED CARE 418, 424 (2007) (concluding that patients using electronic messaging had 6.7% to 9.7% fewer outpatient primary care visits than others). *Contra* Steven J. Katz et al., *Effect of a Triage-Based E-mail System on Clinic Resource Use and Patient and Physician Satisfaction in Primary Care*, 18 J. GEN. INTERNAL MED. 736, 742 (2003) (finding that "e-mail volume did not appear to offset phone volume or visit no-show rates").

156. Chen et al., *supra* note 11, 331–32 (reporting that 85% of patients "rated their satisfaction as 8 or 9 on a nine-point scale" and 85% felt that e-mail contact with physicians "enabled them to better manage their health"); Rosen, *supra* note 155, at 705–06 (reporting that families commented that e-mail "is one method of improving communication and providing consumer-driven health care"); Zhou et al., *supra* note 155, at 418 (reporting that 90% of patients with Internet access have a preference for electronic communication with providers).

157. Patt et al., *supra* note 155 (stating that doctors are concerned about e-mails reaching them in a timely fashion); Rosen, *supra* note 155, at 705 (stating that e-mail communication might produce anxiety about increased liability).

frequently enough so that patients are not neglected if the condition about which they are inquiring is serious.[158]

Physicians have been sued successfully for failing to respond to patient communication outside of office visits. In *St. Charles v. Kender*, the court held that an HMO patient had a viable breach of contract claim against a physician who failed to return her phone calls within two days, during which she suffered a miscarriage.[159] Likewise, in *Fletcher v. Ford*, an appellate court affirmed the denial of a doctor's summary judgment motion after he was sued for medical malpractice arising from his failure to return a telephone call that might have saved the life of an infant with meningitis.[160] By extension, plaintiffs might prevail in medical malpractice claims based on clinicians' unresponsiveness to e-mail.

Doctors may be alarmed by a Physician Insurers Association of America report revealing that $71.8 million in indemnity payments were made for 786 telephone-related malpractice claims.[161] A subsequent study of thirty-two telephone-related cases by malpractice insurers confirmed that such cases are costly and that patient injuries can be catastrophic.[162] Representative mistakes included flawed documentation of calls, inappropriate triage because of inadequate information obtained over the phone, and mismanagement of multiple calls made by the same patient.[163] Similar problems and shortcomings could easily arise when clinicians respond to patient e-mails.[164]

---

158. *See* Eric M. Liederman et al., *Patient-Physician Web Messaging*, 20 J. GEN. INTERNAL MED. 52, 52 (2005) (stating that physicians worry about being "overwhelmed by patient e-mails," that liability may arise because of missed diagnoses or delayed treatment, and that patients are dissatisfied with their physicians' response times). This study at the University of California Davis Health System found that 52.6% of "initial responses were sent within 4 business hours; 70.2% within 8 hours; and 85.5% within 16 hours." *Id.* at 54; *see infra* notes 313–16 for recommendations concerning physician-patient electronic communication.

159. 646 N.E.2d 411, 413 (Mass. App. Ct. 1995).

160. 377 S.E.2d 206, 207, 209 (Ga. Ct. App. 1988).

161. Harvey P. Katz et al., *Patient Safety and Telephone Medicine*, 23 J. GEN. INTERNAL MED. 517, 517 (2007).

162. *Id.* at 517.

163. *Id.* at 518–19; *see also* David E. Hildebrandt et al., *Harm Resulting from Inappropriate Telephone Triage in Primary Care*, 19 J. AM. BOARD FAM. MED. 437, 440–41 (2006) (finding that 1% of patients who called their doctors after hours suffered "harm or discomfort"); Barton D. Schmitt, *Telephone Triage Liability: Protecting Your Patients and Your Practice from Harm*, 55 ADVANCES IN PEDIATRICS 29, 31 (2008) (discussing delayed referral to medical care and other errors that occur in the after-hour call process); Bauer v. Mem'l Hosp., 879 N.E.2d 478, 490–91, 505 (Ill. App. Ct. 2007) (affirming the award of damages to plaintiffs for injuries suffered by an infant in part because his mother received inappropriate medical advice over the telephone).

164. *See infra* notes 313–15 for suggested e-mail protocols that could reduce liability risks.

f) Patient Access to PHRs

In addition to having secure messaging ability, patients may have PHRs that enable them to view part or all of their medical records.[165] However, while patients will likely appreciate such unprecedented access to their health data,[166] certain information might cause confusion, resentment, or trauma, and thus have an adverse health impact.

Providers establishing PHRs must decide whether to include the patient's entire problem, medication and allergy lists, laboratory and diagnostic test results, and comprehensive clinical notes.[167] Some commentators are concerned that if providers share candid psychiatric problem lists and complete progress notes, including personal impressions, patients could become less cooperative with or trusting of their doctors.[168] In the alternative, providers could tailor their notes to avoid causing discomfort to PHR readers, but this approach might sacrifice accuracy.[169]

Also, patients who receive bad news through electronically transmitted test results rather than through a conversation with a sensitive clinician could be traumatized, misunderstand their diagnoses, or feel angry or hopeless.[170] Such patients might decide to stop complying with their treatments and suffer clinical setbacks. Plaintiffs with poor outcomes who feel that their doctors were uncommunicative or insensitive in their communication may be more likely than others to sue.[171] Thus, physicians' decisions to post or omit certain information from PHRs could contribute to the likelihood of medical malpractice claims against them.

---

165. *See supra* notes 53–62 and accompanying text.

166. MARKLE FOUNDATION, ATTITUDES OF AMERICANS REGARDING PERSONAL HEALTH RECORDS AND NATIONWIDE ELECTRONIC HEALTH INFORMATION EXCHANGE (2005), http://www.phrconference.org/assets/research_release_101105.pdf (finding that 60% of Americans support the creation of secure PHRs, and only 19% of Americans state they would not use PHRs for any purpose).

167. Halamka, *supra* note 54, at 3–5.

168. *Id.*

169. It should be noted that the HIPAA Privacy Rule allows patients access to their medical records. Specifically, the regulations provide that "an individual has a right of access to inspect and obtain a copy of protected health information about the individual in a designated record set," with some exceptions, such as psychotherapy notes and information compiled for purposes of litigation or administrative proceedings. Furthermore, the Privacy Rule enables individuals to request amendment of PHI that is incorrect. 45 C.F.R. §§ 164.524, 164.526(a) (2008). These provisions, however, would not require doctors to include any specific information in a PHR.

170. Halamka, *supra* note 54, at 4 (reporting that at the authors' institution, all results are released to patients immediately except for HIV results, cytology/pathology results, and results from MRI/CT testing done to follow cancer progression).

171. *See supra* note 107 and accompanying text.

Yet another concern relates to patients' ability to add notes and information to their PHRs.[172] Patients might wrongly assume that they are communicating directly with their doctors by inputting data and expect physicians to review their PHR entries regularly. Doctors would be well-advised to ask patients using PHRs to sign a form that explains the extent to which clinicians will review this submitted data, if at all. Without such a notice, patients who are harmed because their doctors ignored or never saw important details that they noted in their PHRs may file malpractice claims.

g) Product Defects

In some cases, EHR systems themselves or the computing platforms that support them will be flawed.[173] Thus, EHR system use can cause poor outcomes because of product defects rather than user error. In early 2009, the public learned that software glitches in the Veterans Affairs' EHR system exposed veterans to potentially life-threatening drug dosage errors, including excessively prolonged intravenous infusion of the blood-thinner heparin.[174] Other such instances include flawed EHR system software that provided erroneous calculation of intracranial pressure[175] and a case in which ninety-three minutes of data were missing from the automated anesthesia record of a brain tumor patient who woke up from surgery as a quadriplegic.[176]

CPOE systems have been particularly vulnerable to criticism. While some of their weaknesses lead to input errors,[177] they are also susceptible to software defects. These include: (1) incorrect prompts regarding dosages; (2) an absence of warnings that drug orders must be renewed or that certain drug combinations are inappropriate; (3) failure to automatically cancel medication orders when procedures that require the drugs are cancelled or postponed; (4) lack of interoperability and communication among different systems within the same hospital, such as those belonging to the pharmacy

172. *See supra* notes 58–61 and accompanying text.

173. Jonathan K. Gable, *An Overview of the Legal Liabilities Facing Manufacturers of Medical Information Systems*, 5 QUINNIPIAC HEALTH L.J. 127, 129–31 (2001) (describing instances in which improper medical treatment was provided because of computer programming or software errors); Ross Koppel & David Kreda, *Health Care Information Technology Vendors' "Hold Harmless" Clause: Implications for Patients and Clinicians*, 301 J. AM. MED. ASS'N 1276, 1278 (2009) ("[I]n many cases, HIT problems may be caused not by clinicians but by poor software.").

174. Yen, *supra* note 16.

175. Koppel & Kreda, *supra* note 173, at 1276.

176. Michael M. Vigoda & David A. Lubarsky, *Failure To Recognize Loss of Incoming Data in an Anesthesia Record-Keeping System May Have Increased Medical Liability*, 102 ANESTHESIA & ANALGESIA 1798, 1798–99 (2006).

177. *See supra* note 141–42 and accompanying text.

and house staff; and (5) computer crashes and maintenance shutdowns that lead to lost orders.[178]

One review found that information inconsistencies in CPOE systems pose significant risks to patient safety.[179] Information inconsistencies were defined as disparities between data entered through structured templates and information in free-text comment fields.[180] The review examined 55,992 CPOE prescriptions and concluded that 532 of them contained errors, most commonly in dosage, of which twenty percent could have caused moderate to significant harm.[181] Errors were attributable to automated dosage defaults, comments automatically transferred to new prescriptions after modification of existing prescriptions, insufficient training on CPOE systems, and flawed standardized templates.[182]

Both health care organizations and physicians can be held liable for harms associated with use of faulty equipment. Hospitals, clinics, or physicians who purchase low-quality, defective EHR systems or fail to maintain the systems properly could be sued for any resulting harm suffered by patients.[183] Whether or not a decision to adopt a particular product constitutes negligence will depend on professional custom.[184] If providers select an EHR system that is widely recognized as inadequate, and the system causes injury to patients, plaintiffs might be able to establish medical malpractice.[185]

It is also possible that physicians who did not participate in their employer's decision to choose a defective EHR system could be found liable for negligence because of product flaws. While many physicians will not have the technical expertise to detect certain software defects, in some cases they

178. Koppel, *supra* note 112, at 1199–1201.

179. Hardeep Singh et al., *Prescription Errors and Outcomes Related to Inconsistent Information Transmitted Through Computerized Order Entry*, 169 ARCHIVES INTERNAL MED. 982, 989 (2009).

180. *Id.* at 983.

181. *Id.* at 984, 986.

182. *Id.* at 987–88.

183. Lamb v. Candler Gen. Hosp., 413 S.E.2d 720, 721–22 (Ga. 1992) ("It is well recognized that a hospital may be liable in ordinary negligence for furnishing defective equipment for use by physicians and surgeons in treating patients."); Berg v. United States, 806 F.2d 978, 983 (10th Cir. 1986) (upholding a verdict for the plaintiff whose injuries were caused in part by a lack of adequate testing and maintenance of equipment and a lack of adequate training of technicians).

184. *See supra* notes 73–74 and accompanying text.

185. *See* Emory Univ. v. Porter, 120 S.E.2d 668, 670 (Ga. Ct. App. 1961) (stating that a hospital has a "duty of exercising ordinary care to furnish equipment and facilities reasonably suited to the uses intended and such as are in general use under the same, or similar, circumstances in hospitals in the area").

may become aware of system flaws that generate obvious errors. A physician who used an EHR system knowing that it caused particular problems such as dosage errors, who did not demand that her employer ensure that the system is repaired, and who took no precautions, such as reviewing each dosage recommendation to ensure accuracy, might be deemed by a court to be responsible for patient injuries. In *Wickline v. State of California*, a California court of appeals stated in dicta that "the physician who complies without protest with the limitations [of covered hospitalization days] imposed by a third party payer, when his medical judgment dictates otherwise, cannot avoid his ultimate responsibility for his patient's care."[186] Thus, if a court finds that a reasonable physician would not have tolerated her institution's faulty EHR system without protest and without implementing clinical safeguards to avoid patient harm, the individual might be held liable in a medical malpractice case.

Contractual provisions favored by EHR vendors may exacerbate the liability vulnerability of clinicians using EHR systems. Vendors may disclaim implied and express warranties or insert "hold harmless" clauses into their contracts that shield them from liability and shift responsibility for harm to health care providers.[187] Contractual provisions that limit liability can be invalidated as violating public policy if the parties have unequal bargaining power or the provision encourages reckless or negligent behavior.[188] Thus, courts may find "hold harmless" provisions unenforceable if they are convinced that health care providers lack the technical knowledge and sophistication to bargain on equal footing with vendors.[189] Judges may also revoke provisions that are deemed likely to promote carelessness on the part

186. 192 Cal. App. 3d 1630, 1645 (1986). In this case, a doctor sought permission from Medi-Cal to extend his patient's hospital stay by eight days. An extension was granted for only four days, and the doctor released the patient at the end of that period. The patient was later readmitted to the hospital because of complications, and her leg had to be amputated. She sued the state of California, which operated Medi-Cal, but the court of appeals ultimately found that the state was not liable for Wickline's injuries.

187. Koppel & Kreda, *supra* note 173, at 1276; Lisa L. Dahm, *Restatement (Second) of Torts Section 324A: An Innovative Theory of Recovery for Patients Injured Through Use or Misuse of Health Care Information Systems*, 14 J. MARSHALL J. COMPUTER & INFO. L. 73, 78, 92–93 (1995); Gable, *supra* note 173, at 141.

188. Blake D. Morant, *Contracts Limiting Liability: A Paradox with Tacit Solutions*, 69 TULANE L. REV. 715, 734 (1995); Tunkl v. Regents of the Univ. of Cal., 383 P.2d 441, 447 (Cal. 1963) (finding that a hold harmless agreement imposed as a condition of admission to a hospital was invalid because the patient had unequal bargaining power); Emory Univ. v. Porubiansky, 282 S.E.2d 903, 904–06 (Ga. 1981) (holding that a waiver of claims in an informed consent agreement was invalid as a matter of public policy).

189. Koppel & Kreda, *supra* note 173, at 1276 (arguing that there exists a "substantial disparity between buyers and sellers in knowledge and resources").

of manufacturers.[190] In the alternative, states may enact statutes that invalidate particular types of hold harmless clauses.[191] Typically, however, contractual limitations of liability are enforceable.[192]

B. PRIVACY BREACHES

Computerized information is vulnerable to large-scale privacy violations associated with hacking, computer theft, malicious electronic distribution, or accidental disclosure, such as sending a file to the wrong e-mail address.[193] Once data security is breached, the most private information can be dispersed on the Internet to a worldwide audience.[194] Disclosure of psychiatric or sexual histories or other sensitive information can, among other harms, lead to profound embarrassment, ruined careers, or loss of professional and personal opportunities.[195] These, in turn, can generate litigation against those responsible for security breaches.

1. Security Threats and Regulation

Privacy breaches involving EHRs have occurred in the United States with alarming frequency. For example, in 2008, computer files containing health and financial details of more than 2.1 million patients were stolen from a storage company hired by the University of Miami Health System, and information about 6,000 patients of the University of California San Francisco Medical Center was available online for three months.[196] That same year, a laptop belonging to a National Institutes of Health researcher was stolen, compromising private information about nearly 2,500 heart disease patients.[197] According to some estimates, between 250,000 and 500,000 patients suffer medical identity theft each year.[198]

---

190. *Id.* (describing software malfunctions).

191. Carl Giesler, *Managers of Medicine: The Interplay Between MCOs, Quality of Care, and Tort Reform*, 6 TEX. WESLEYAN L. REV. 31, 50 (1999) (reporting that some states enacted statutes that invalidate hold-harmless clauses in contracts between physicians and managed care organizations).

192. Adloo v. H.T. Brown Real Estate, Inc., 686 A.2d 298, 301 (Md. 1995) ("It is well settled in this State, consistent with 'the public policy of freedom of contract,' . . . that exculpatory contractual clauses generally are valid.").

193. Hoffman & Podgurski, *In Sickness, Health and Cyberspace: Protecting the Security of Electronic Private Health Information*, *supra* note 30, at 333.

194. *Id.* at 335.

195. *Id.* at 334–35.

196. American Medical Association, *News in Brief: Miami Patient Data Stolen*, AM. MED. NEWS, May 19, 2008, http://www.ama-assn.org/amednews/2008/05/19/bibf0519.htm.

197. *Safeguarding Private Medical Data*, N.Y. TIMES, March 26, 2008, at A22. In 2006 an Aetna laptop computer containing personal information concerning 38,000 consumers was stolen and a security breach compromised the confidentiality of records from 60,000 patients

To address the threats to patient privacy, the U.S. Department of Health and Human Services (HHS) enacted the Privacy Rule under the Health Insurance Portability and Accountability Act of 1996 (HIPAA).[199] The Privacy Rule requires health care providers to safeguard patient privacy in a variety of ways. For example, with some exceptions, covered entities must obtain a patient's permission before speaking to third parties about the patient's medical condition;[200] must distribute privacy notices containing information concerning use and disclosure of patients' health records;[201] and must allow patients to inspect their health records and request that they be modified or used restrictively.[202] The HIPAA Security Rule, which is part of the Privacy Rule, focuses specifically on data security and the electronic storage and transmission of private health information (PHI).[203] The Security Rule, which became effective on April 20, 2005 for most covered entities,[204] delineates administrative, physical, and technical safeguards to protect the confidentiality, integrity, and availability of electronic protected health information.[205]

Despite the regulatory mandates, many commentators agree that privacy and security threats still abound. A large 2007 study of security vulnerabilities concluded that "commercial EHR systems are vulnerable to exploitation given existing industry development and disclosure practices."[206] A 2008 report issued by the HHS Office of the Inspector General concluded that the

---

who visited Ohio University's health center. *See* Ronald A. Williams, *Statement of Aetna CEO and President Ronald A. Williams on Data Security*, Apr. 6, 2006, http://www.aetna.com/news/2006/pr_20060426.htm; Jennifer Gonzalez, *3rd Computer Breach at OU Within 3 Weeks*, THE PLAIN DEALER, May 12, 2006, at A1.

198. Judith Graham, *Medical Identity Theft Spreads: Purloined Data Often the Crime of Insiders*, CHI. TRIB., Aug. 22, 2008, at 10.

199. The HIPAA Privacy Rule is found at 45 C.F.R. §§ 160.101–164.534 (2008). HIPAA provides statutory authority for these regulations at 42 U.S.C. §§ 1320d–1320d-8 (2006).

200. 45 C.F.R. § 164.510 (2008).

201. 45 C.F.R. § 164.520(a) (2008).

202. 45 C.F.R. §§ 164.520, 164.522 (2008).

203. 45 C.F.R. §§ 164.302–164.318 (2008). Under the Privacy Rule, PHI includes "individually identifiable health information" that is electronically or otherwise transmitted or maintained. 45 C.F.R. § 160.103 (2008).

204. 45 C.F.R. § 164.318 (2008). Small health plans were given an extended adjustment period and were required to comply with the rule by April 20, 2006.

205. For a description and critique of the HIPAA Security Rule, see Hoffman & Podgurski, *In Sickness, Health and Cyberspace: Protecting the Security of Electronic Private Health Information*, *supra* note 30.

206. *eHVRP Study Finds Healthcare Industry Must Do More to Protect Electronic Health Record Systems*, BUS. WIRE, Sept. 17, 2007, *available at* http://www.thefreelibrary.com/_/print/PrintArticle.aspx?id=168732503. The study was conducted over 15 months and surveyed more than 850 provider organizations.

federal government had failed to provide adequate oversight or effective enforcement of the HIPAA Security Rule.[207] Preliminary results of HHS audits of U.S. hospitals revealed "numerous, significant vulnerabilities" in PHI protections that jeopardize its confidentiality.[208]

PHRs may raise particular privacy challenges. Web-based PHRs enable the service provider to obtain and sell health information to marketers and advertisers.[209] Employers who offer PHRs to workers[210] might be tempted to retrieve data and use it for purposes of employment decisions.[211] Those designing PHRs must incorporate safeguards to ensure that patients or their authorized proxies are properly authenticated before accessing their PHRs and that all others are blocked from doing so.[212]

The threats to EHR security have not eluded public notice. When asked, the overwhelming majority of American patients express concern about the privacy of their medical records. A 2005 National Consumer Health Privacy Survey involving 2,000 individuals revealed that sixty-seven percent of respondents were "somewhat" or "very concerned" about PHI confidentiality.[213] Furthermore, thirteen percent of respondents claimed that they had attempted to protect their own privacy by avoiding medical tests or visits to their regular physicians, asking doctors to distort diagnoses, or paying for tests out-of-pocket so that no medical documentation would be sent to insurance companies.[214] That same year, a Markle Foundation survey found that "[a]ttributes of a proposed nationwide health information exchange that focus on security and privacy are rated as the highest priorities among survey respondents."[215] In a 2007 online survey, forty percent of respondents disagreed with the statement that "the benefits of electronic medical records outweigh the privacy risks."[216]

---

207. DANIEL R. LEVINSON, DEP'T OF HEALTH AND HUMAN SERVS., NATIONWIDE REVIEW OF THE CENTERS FOR MEDICARE & MEDICAID SERVICES HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT OF 1996 OVERSIGHT, A-04-07-05064, 3 (2008), *available at* http://www.oig.hhs.gov/oas/reports/region4/40705064.pdf.

208. *Id.* at 3–4.

209. Terry, *supra* note 21, at 237.

210. *See* Halamka et al., *supra* note 54 and accompanying text.

211. DANIEL R. LEVINSON, *supra* note 207, at 3–4.

212. Halamka et al., *supra* note 54, at 5.

213. LYNNE "SAM" BISHOP ET AL., CAL. HEALTHCARE FOUND., NATIONAL CONSUMER HEALTH PRIVACY SURVEY 2005, at 3 (2005).

214. *Id.* at 4.

215. MARKLE FOUNDATION, *supra* note 166, at 2.

216. Robert Steinbrook, *Personally Controlled Online Health Data—The Next Big Thing in Medical Care?*, 358 N. ENGL. J. MED. 1653, 1655 (2008).

2. Potential Litigation

Patients who learn that their medical information has been inappropriately disclosed to third parties may be inclined to sue their physicians. Litigation may be facilitated by the HITECH Act, which includes several provisions designed to enhance the efficacy of the HIPAA Privacy and Security Rules.[217] The law requires that covered entities[218] notify individuals of any security breaches[219] involving their "unsecured" PHI.[220] Thus, if providers comply with this mandate, patients will learn of security breaches that compromise their PHI. In fact, patients might initiate litigation not only when the physician has carelessly or intentionally disclosed PHI, but also when the disclosure occurred because of hacking or an EHR system defect. It will be up to courts to determine whether providers are at fault for such security breaches.[221]

Patients could sue clinicians for privacy breaches under a variety of theories. The tort of invasion of privacy is one possibility. It consists of four elements: (1) public disclosure; (2) of a private fact; (3) that would be objectionable and offensive to a reasonable person; and (4) that is not of

---

217. *See* Reece Hirsch & Rebecca Fayed, *ARRA 2009 and the HITECH Act: The Next Phase of HIPAA Regulation and Enforcement Arrives*, 18 BNA'S HEALTH L. REP. 308 (2009) (detailing the law's privacy-related provisions).

218. Covered entities are health plans, health care clearinghouses, and health care providers who transmit health information electronically for claims, billing or health plan purposes. 45 C.F.R. § 160.103 (2008). The HITECH Act establishes that the Security Rule's requirements also apply to business associates of covered entities. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13401, 123 Stat. 115, 260 (2009) (to be codified at 42 U.S.C. § 17931(a)).

219. The term "breach" is defined as "the unauthorized acquisition, access, use, or disclosure of protected health information which compromises the security or privacy of such information, except where an unauthorized person to whom such information is disclosed would not reasonably have been able to retain such information." American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13400(1)(A), 123 Stat. 115, 258 (2009) (to be codified at 42 U.S.C. § 17921). For exceptions to this definition, see § 13400(1)(B).

220. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13402(a), 123 Stat. 115, 260 (2009) (to be codified at 42 U.S.C. § 17932(a)). Unsecured PHI is to be defined through DHHS guidance, but if the Secretary fails to issue guidance, it will be defined as "PHI that is not secured by a technology standard that renders protected health information unusable, unreadable, or indecipherable to unauthorized individuals and is developed or endorsed by a standards developing organization that is accredited by the American National Standards Institute." American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 13402(h)(1)(B)–(h)(2), 123 Stat. 115, 262–63 (2009) (to be codified at 42 U.S.C. § 17932(h)(1)(B)–(h)(2)).

221. *See supra* note 186 and accompanying text.

legitimate public concern.[222] An alternative tort theory is breach of confidentiality,[223] whose elements are (1) the existence of a doctor-patient relationship, and (2) a physician's or medical entity's disclosure to a third party of confidential information that was gained pursuant to this relationship.[224]

State law can provide plaintiffs with additional causes of action.[225] For example, the California Constitution explicitly establishes that state residents

---

222. *See* Diaz v. Oakland Tribune, 188 Cal. Rptr. 762, 766 (Ct. App. 1983) (reporting that the jury found defendant liable for publicizing the fact that plaintiff had gender-corrective surgery).

223. Peter A. Winn, *Confidentiality in Cyberspace: The HIPAA Privacy Rules and the Common Law*, 33 RUTGERS L.J. 617, 652–58 (2002) (discussing the common law tort theory of breach of confidentiality and its implications).

224. Sorensen v. Barbuto, 143 P.3d 295, 299 (Utah Ct. App. 2006) (discussing claim of breach of confidentiality where physician communicated with party opposing patient while litigation was pending); Biddle v. Warren Gen. Hosp., 715 N.E.2d 518, 523 (Ohio 1999) (establishing that "in Ohio, an independent tort exists for the unauthorized, unprivileged disclosure to a third party of nonpublic medical information that a physician or hospital has learned within a physician-patient relationship"); Kimberly Rathbone, *The Strict Ohio Supreme Court Decisions in* Biddle*: Third Party Law Firm Held Liable for Inducing Disclosure of Medical Information*, 15 J.L. & HEALTH 189, 196–97 (2001).

225. Some states provide aggrieved parties with a general cause of action for privacy breaches. *See* CAL. CIV. CODE §§ 56.35–56.36 (2008); MD. CODE ANN., HEALTH-GEN. § 4-309(f) (2009); MASS. GEN. LAWS ANN. CH. 214 § 1B (2005); MINN. STAT. ANN. § 144.298 (2005 & SUPP. 2009); MONT. CODE. ANN. § 50-16-553 (2007), *amended by* 2009 Mont. Laws 56 (to be codified at MONT. CODE ANN. § 13-15-205); N.H. REV. STAT. ANN. § 151:30 (2009); TENN. CODE ANN. § 68-11-1504 (2009); TEX. HEALTH & SAFETY CODE ANN. § 241.156 (Vernon 2001); TEX. OCC. CODE ANN. § 159.009 (Vernon 2004); WASH. REV. CODE ANN. § 70.02.170 (2009) *amended by* 2009 Wash. Sess. Laws page no. 1493; WYO. STAT. ANN. § 35-2-616 (2009).

Other states provide a more limited cause of action for improper disclosure of specific medical information such as HIV/AIDS test results, genetic testing, and mental health records. Statutes relating to HIV/AIDS are: ARIZ. REV. STAT. ANN. § 36-668 (2009); CAL. HEALTH & SAFETY CODE § 120980 (2009); CONN. GEN. STAT. ANN. § 19a-590 (2003 & Supp. 2009); DEL. CODE ANN. tit. 16, § 1205 (2009); 410 ILL. COMP. STAT. ANN. 305/13 (2005 & Supp. 2009); IOWA CODE § 141A.11 (2005 & Supp. 2009); ME. REV. STAT. ANN. tit. 22, § 825 (2004 & Supp. 2009); ME. REV. STAT. ANN. tit. 5, § 19206 (2009); MICH. COMP. LAWS § 333.5131(8) (2001); MO. REV. STAT. § 191.656(6) (2004 & Supp. 2009); MONT. CODE. ANN. § 50-16-1013 (2007), *amended by* 2009 Mont. Laws 362; N.H. STAT. ANN. § 141-F:10 (2009); N.D. CENT. CODE § 23-07.5-07 (2008); OKLA. STAT. tit. 63, § 1-502.2(H) (2004 & Supp. 2009) *amended by* S.B. 928, 52nd Leg., 1st Reg. Sess. (Okla. 2009); 35 PA. CONS. STAT. § 7610 (2003); TEX. HEALTH & SAFETY CODE ANN. § 81.104 (Vernon 2009); VA. CODE ANN. § 32.1-36.1(c) (2009); WASH. REV. CODE ANN. § 70.24.084 (2002); W. VA. CODE ANN. § 16-3C-5 (2006).

Litigation rights for disclosure of mental health information are provided by: CAL. WELF. & INST. CODE § 5330 (1998 & Supp. 2009); 740 ILL. COMP. STAT. 110/15 (2002 & Supp. 2009); TEX. HEALTH & SAFETY CODE ANN. § 611.005 (Vernon 2009 & Supp. 2009); WASH. REV. CODE ANN. § 71.05.440 (2008); WISC. STAT. ANN. § 51.30 (2008 & Supp.

have a right to privacy,[226] and the California Confidential Medical Information Act (CMIA) generally prohibits health care providers from disclosing their patients' records without their authorization.[227] In *Kina v. United Air Lines Inc.*, a federal district court allowed a plaintiff to proceed with his claim that his state constitutional and statutory rights were violated when his "fitness-for-duty" exam results were disclosed to his employer without his authorization.[228] Similarly, in *Berger v. Sonneland*,[229] the Supreme Court of Washington ruled that a statutory cause of action existed for a physician's unauthorized disclosure of a patient's medical information to her former husband.[230]

Although the HIPAA Privacy Rule does not provide aggrieved individuals with a private cause of action,[231] it might constitute evidence of the appropriate standard of care in negligence actions involving privacy breaches.[232] Furthermore, the HIPAA Privacy Rule authorizes government enforcement action for regulatory violations.[233] Providers may be subject to monetary penalties, with the amount depending on the severity of the offense.[234] Furthermore, the HITECH Act allows state attorneys general to bring civil actions for HIPAA violations in federal court.[235] The combination of federal investigations and litigation by attorneys general may subject providers to vigorous enforcement of the HIPAA Privacy Rule.

2009). Private action for disclosure of genetic information is allowed by: DEL. CODE ANN. tit. 16, § 1227(c) (2009); 410 ILL. COMP. STAT. ANN. 513/40 (2005 & Supp. 2009); MASS. GEN. LAWS ANN. ch. 111, § 706(d) (2003); NEV. REV. STAT. ANN. § 629.201 (2008); N.H. REV. STAT. ANN. § 141-H:6 (2009); N.J. STAT. ANN. § 10:5-49 (2002); N.M. STAT. ANN. § 24-21-6 (2009).

226. CAL. CONST. art. I, § 1.

227. CAL. CIV. CODE § 56.10(a) (2008) ("No provider of health care, health care service plan, or contractor shall disclose medical information regarding a patient of the provider of health care or an enrollee or subscriber of a health care service plan without first obtaining an authorization, except as provided in subdivision (b) or (c).").

228. 2008 WL 5071045, at *8–10 (N.D. Cal. Dec. 1, 2008).

229. 26 P.3d 257, 265 (Wash. 2001) (finding the disclosure to constitute "injuries occurring as a result of health care" under the statute).

230. *Id.* at 259, 269.

231. 45 C.F.R. §§ 160.300–160.552 (2008); Hoffman & Podgurski, *supra* note 30, at 354.

232. Acosta v. Byrum, 638 S.E.2d 246, 253 (N.C. Ct. App. 2006) (explaining that HIPAA was relevant to the extent it provided evidence of the duty of care owed by a physician with respect to the privacy of a patient's medical records).

233. 45 C.F.R. §§ 160.300–160.552 (2008).

234. American Reinvestment and Recovery Act of 2009, Pub. L. No. 111-5, § 13410(d), 123 Stat. 115 (2009) (to be codified at 42 U.S.C. § 1320d-5).

235. *Id.*

C. DISCIPLINARY ACTION BY STATE MEDICAL BOARDS AND CRIMINAL PROSECUTION

The HIPAA Privacy Rule is not the only basis for government intervention with respect to provider misconduct. In egregious cases, health care professionals may also face disciplinary action by state medical boards, criminal prosecution for negligent or reckless treatment of patients, or other penalties.[236]

State Medical Practice Acts empower state medical boards to impose fines, reprimands, censures, probation, suspension, or license restriction or revocation on physicians who engage in misconduct.[237] Doctors who deviate unacceptably from the appropriate standard of care may be disciplined even if no individual patient was placed at risk or suffered tangible harm.[238] A particularly relevant example is *Bogdan v. New York State Board for Professional Medical Conduct*, in which the board imposed a two-year limited probation on an anesthesiologist, in part because of her failure to maintain adequate medical records.[239]

In addition, in extreme cases, physicians can be charged with involuntary manslaughter, negligent homicide, reckless endangerment, reckless homicide, grossly negligent medical care, or other criminal violations.[240] To illustrate, in *People v. Einaugler*, a doctor was convicted of reckless endangerment and willful violation of health laws after he failed to transfer a patient from a nursing-home to a hospital in a timely fashion.[241] In *Commonwealth v. Youngkin*,

---

236. Timothy J. Aspinwall, *Representing Healthcare Professionals in Disciplinary Actions: Containing the Collateral Damage*, 20 No. 3 HEALTH LAWYER 1, 1–6 (2008) (describing a variety of penalties that could be imposed on physicians providing substandard care); Ronald L. Eisenberg & Leonard Berlin, *When Does Malpractice Become Manslaughter?*, 179 AM. J. ROENTGENOLOGY 331, 332 (2002) (noting an increase in the criminal prosecution of physicians for reckless endangerment of patients); Laura J. Spencer, *The Florida "Three Strikes Rule" for Medical Malpractice Claims: Using a Clear and Convincing Evidence Standard to Tighten the Strike Zone for Physician Licensure Revocation*, 28 ST. LOUIS U. PUB. L. REV. 317, 321–24 (2008) (describing disciplinary proceedings by state medical boards).

237. Spencer, *supra* note 236, at 321, 327; James Morrison & Peter Wickersham, *Physicians Disciplined by a State Medical Board*, 279 J. AM. MED. ASS'N 1889, 1890, 1893 (1998) (reporting that in California, approximately 250 physicians are disciplined each year and estimating that 2400 physicians are disciplined each year in the United States).

238. Haw v. State Bd. of Med., 90 P.3d 902, 908 (Idaho 2004); Bogdan v. State Bd. for Prof'l Med. Conduct, 606 N.Y.S.2d 381, 382 (N.Y. App. Div. 1993).

239. *Bogdan*, 606 N.Y.S.2d at 382–83.

240. Paul R. Van Grunsven, *Medical Malpractice or Criminal Mistake? - An Analysis of Past and Current Criminal Prosecutions for Clinical Mistakes and Fatal Errors*, 2 DEPAUL J. HEALTH CARE L. 1, 14–43 (1997) (describing various criminal prosecutions).

241. 618 N.Y.S.2d 414, 414–15 (N.Y. App. Div. 1994).

a physician was convicted of involuntary manslaughter after the death of a seventeen year old patient to whom he prescribed a barbiturate.[242]

Physicians accused of providing substandard care may face other adverse consequences as well. They may lose their medical malpractice insurance, have their medical staff privileges suspended, or see their specialty board certification revoked.[243]

In the future, state board disciplinary proceedings, criminal prosecutions, or other penalties may be initiated because of performance deficiencies that are related to EHR systems. Health care professionals who rely improperly on prior physicians' diagnostic work, fail to review a patient's entire EHR, input data incorrectly, disregard prompts and alerts, or mishandle patient e-mail could face not only private medical malpractice lawsuits, but also governmental intervention.

IV. ADDRESSING LIABILITY RISKS: STRATEGIES AND RECOMMENDATIONS

Litigation and government enforcement actions offer retrospective review of challenged activities and provide post-hoc remedies to aggrieved parties. However, because lives are at stake in the health care setting, it is critical that prospective strategies be available to prevent patient harm before it occurs. We now turn to a variety of initiatives that may be undertaken to optimize EHR systems' effectiveness, maximize their usability for clinicians, and minimize risks to patient safety.

The medical community is at a crossroads. New health information technology has the potential to produce dramatic improvements in health outcomes. However, without safeguards, this technology could impair the performance of health care providers and expose them to unprecedented liability risks. We focus on two strategies to minimize these risks. First, EHR systems must be carefully regulated so that they cannot be marketed without being scrutinized, approved, and subject to ongoing oversight. Second, EHR system experts, clinicians, and the government should develop high-quality clinical practice guidelines and agency guidance concerning EHR systems. Such guidance will educate health care providers about proper EHR system acquisition and use practices and elucidate the standard of care for purposes of litigation.

---

242. 427 A.2d 1356, 1358–59, 1370 (Pa. Super. Ct. 1981).

243. *See* Aspinwall, *supra* note 236, at 5–6; William B. Schwartz & Daniel N. Mendelson, *Physicians Who Have Lost Their Malpractice Insurance: Their Demographic Characteristics and the Surplus-Lines Companies That Insure Them*, 262 J. AM. MED. ASS'N 1335, 1335 (1989).

A. ACHIEVING QUALITY CONTROL

Arguably, innovation in the EHR system industry can only be stimulated if the technology remains unregulated.[244] Government intervention that imposes burdensome requirements could discourage small entrepreneurs from entering the market. However, allowing manufacturers to produce and sell EHR systems whose quality and safety is unregulated could be extremely dangerous for patients and providers.

Without government oversight and quality control, health care providers will risk investing billions of dollars in poorly designed systems that compromise rather than improve health outcomes. Once a practice purchases a system, enters patient records into it, and trains its staff, it is likely to retain it even if it is deficient, rather than incur the high cost of switching systems. Flawed systems that lead to medical errors and poor health outcomes will inevitably increase providers' vulnerability to liability in medical malpractice cases. Similarly, a lack of governmental oversight to ensure that clinicians receive up-to-date, high-quality training concerning EHR systems could contribute to liability exposure.

1. Government Regulations

EHR systems are not currently approved or inspected by any regulatory agency prior to marketing.[245] Rather, a private sector organization called the Certification Commission for Health Information Technology (CCHIT) has developed a voluntary certification process for EHR systems.[246] However, the CCHIT certification process inadequately safeguards the quality and integrity of these products.[247] The short duration of testing and its deficient rigor substantially weaken the certification's utility. All testing occurs during one day, and therefore, inspectors do not observe the system operating over time and in a variety of usage environments.[248] Furthermore, applicants can access testing scenarios and scripts on CCHIT's website prior to testing. Therefore, they are not required to ensure that their systems appropriately handle the variety of user actions that can actually occur in the field.[249]

---

244. *See* Hoffman & Podgurski, *supra* note 13, at 126 (discussing the absence of regulation for EHR systems).

245. *See id.*

246. Certification Commission for Health Information Technology, About the CCHIT, http://www.cchit.org/about (last visited Sept. 24, 2009).

247. *See* Hoffman & Podgurski, *supra* note 13, at 132–34; Blumenthal, *supra* note 4, at 1478 (stating that "[t]ightening the certification process is a critical early challenge" for the Office of the National Coordinator of Health Information Technology).

248. *Id.* (stating that "[t]his inspection takes a full day").

249. CERTIFICATION COMMISSION FOR HEALTH INFORMATION TECHNOLOGY,

The HITECH Act suggests that improved certification criteria must be implemented. Section 3004 calls for the federal adoption of an "initial set of standards, implementation specifications, and certification criteria" by December 31, 2009.[250] Such criteria will presumably go beyond those already used by CCHIT. However, the HITECH Act does not detail how these standards and criteria will be implemented and enforced or what role the government will play in doing so. In fact, the legislation states that adherence to the new requirements will generally be voluntary for private entities.[251] Thus, the Act leaves the important matters of determining the safety and efficacy of these devices ambiguous.

A relaxed approach to EHR system oversight is misguided and dangerous.[252] EHR systems will affect many aspects of patient care and are critical medical tools.[253] Appropriate oversight would protect not only patients, but also clinicians and health care organizations, who would be less likely to use flawed technology that causes patient injuries. While federal regulation would not preclude patients from suing for injuries associated with EHR systems,[254] they may well diminish the likelihood of provider liability by enhancing the quality of the equipment they operate.

PHYSICIAN'S GUIDE TO CCHIT CERTIFICATION 8 (2008), *available at* http://www.cchit.org/sites/all/files/CCHITPhysiciansGuide08.pdf ("The criteria and test scripts are published on the Commission's web site: www.cchit.org.").

250. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3004, 123 Stat. 115, 240 (2009) (to be codified at 42 U.S.C. § 300jj-14).

251. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3001(c)(5)(A), 123 Stat. 115, 232 (2009) (to be codified at 42 U.S.C. § 300jj-11(c)(5)(A)) (discussing the "voluntary certification of health information technology"); Pub. L. No. 111-5, § 3006(a)(1), 123 Stat. 115, 241 (2009) (to be codified at 42 U.S.C. § 300jj-16(a)(1)) (explaining that generally, nothing in the Act shall be construed "to require a private entity to adopt or comply with a standard or implementation specification adopted under [the Act]").

252. In prior work we have argued that EHR systems should be subject to regulatory approval and monitoring processes akin to those applying the highest levels of scrutiny to devices regulated by the Food and Drug Administration. *See* Hoffman & Podgurski, *supra* note 13, at 128–31. The full argument will not be repeated here. We also will not address the important question of which specific agency should be tasked with EHR system oversight. For discussion, see *id.* at 134–40. Rather, we refer to the regulating entity merely as HHS, since the responsible agency will most likely be an arm of this department. The HITECH Act establishes the Office of the National Coordinator for Health Information Technology within HHS. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3001(a), 123 Stat. 115, 230 (2009) (to be codified at 42 U.S.C. § 300jj-11(a)).

253. Hoffman & Podgurski, *supra* note 13, at 128–31.

254. *See* Wyeth v. Levine, 129 S.Ct. 1187, 1204 (2009) (holding that state law failure to warn claims are not preempted by the FDA's approval of a warning label pursuant to federal law).

A regulatory framework that required all EHR systems to be tested extensively and approved before they are marketed, as is the case for drugs and medical devices,[255] could establish design criteria that would maximize EHR system usability and reduce the likelihood of input or chart review errors and other mistakes. Regulations could mandate that EHR vendors employ a "best practices" standard, requiring vendors to make reasonable efforts to identify and employ best practices relating to hazard and risk analysis and mitigation, software development, validation, maintenance, security measures, and system integration and operation. The selected practices should be similar to those commonly used by other industry members, or should be clearly demonstrated to be superior to commonly used measures.[256]

In addition, the regulations should specify requirements for particular features. For the sake of brevity, just two examples of criteria that could impact clinician liability will be provided.[257] First, HHS could articulate standards for CPOE applications and other forms of clinical decision support to optimize their safety and efficacy.[258] Second, it could require vendors to comply with user interface design guidelines for all EHR systems[259] so customers switching to a new EHR product would not require a long training and adjustment period and tend initially to introduce errors into medical records. Such standardization would not necessarily stifle competition, especially if HHS oversight included a mechanism for timely approval of innovative user interface features that conflict with existing guidelines.

Imposing regulatory requirements for design specifications is not unprecedented. The HIPAA Security Rule includes security standards and implementation specifications for security safeguards.[260] Similarly, the HITECH Act contemplates the development of standards, implementation

---

255. *See* 21 C.F.R. § 7.3(f) (2008) (defining the jurisdiction of the Food and Drug Administration); *see generally* 21 C.F.R. §§ 1–1405.670 (2009) (food and drug regulations); *see also* Hoffman & Podgurski, *supra* note 13 at 134–38 (critiquing FDA regulation of devices).

256. Hoffman & Podgurski, *supra* note 13, at 151.

257. For further details, see Hoffman & Podgurski, *supra* note 13, at 150–62.

258. *See* Kuperman, *supra* note 145, at 37 (providing recommendations for CPOE application vendors and drug information knowledge-base vendors). For example, alerts could be differentiated by color, which would indicate the seriousness of the potential harm to patients.

259. This could be done once experts have sufficient experience with EHR systems to determine the design of an optimal user interface.

260. 45 C.F.R. §§ 164.302–.318 (2008). *But see* Hoffman & Podgurski, *supra* note 30, at 344–59 (critiquing the HIPAA Security Rule).

specifications, and certification criteria for EHR systems.[261] We urge that these take the form of detailed regulatory requirements that are mandatory for all EHR system vendors.[262]

Ongoing monitoring is also critical for quality control[263] and can affect clinician liability risks. Currently, some EHR system contracts prohibit users from disclosing product problems to others.[264] Such restrictions increase the risk of harm to patients and should be prohibited by law. Vendors should be required to submit adverse event accounts to HHS, and summary reports of these events should be posted on the agency's website.[265] Adverse events would include all system problems that are associated with a design or operational flaw rather than with user error. Such reports would educate potential purchasers about product defects or usability problems. They may also protect providers who face litigation by proving that a vendor[266] rather than clinician was at fault for an EHR system problem that caused a poor medical outcome.[267]

Finally, state governments could mandate training both with respect to the particular product that a clinician is using and with respect to general EHR system use practices. Comprehensive and effective training is essential to the success of EHR system implementation.[268] As of 2009, sixty-two state medical boards required clinicians to earn continuing medical education

261. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3003(b)(1)(A), 123 Stat. 115, 238 (2009) (to be codified at 42 U.S.C. § 300jj-13(b)(1)(A)).

262. *See* Pub. L. No. 111-5, Title XIII, §§ 3006, 13112, 123 Stat. 115, 241, 243 (2009) (to be codified at 42 U.S.C. §§ 300jj-16, 17902) (requiring compliance only from private entities that enter into contracts with the federal government).

263. Hoffman & Podgurski, *supra* note 13, at 147–50 (discussing the need for ongoing monitoring of EHR systems).

264. Koppel & Kreda, *supra* note 173, at 1278.

265. Hoffman & Podgurski, *supra* note 13, at 148. Postings should delete trade secret information, confidential commercial and financial information, patient information, and information about the identities of the users who reported the adverse events; *see* 21 C.F.R. § 803.9, 814.44(d) (2008) (discussing the Food and Drug Administration's posting of redacted adverse event reports for medical devices).

266. We use the term "vendor" broadly to refer to those who develop or modify EHR system software and to those who sell and install such systems.

267. *See infra* note 340 and accompanying text (discussing the relevance of user problem reports to establishing the standard of care for EHR system use in litigation).

268. Wanda L. Krum & Jack D. Latshaw, *Training*, *in* IMPLEMENTING AN ELECTRONIC HEALTH RECORD SYSTEM 60–66 (James M. Walker et al. eds. 2005) (discussing the importance of training and providing recommendations for development of a successful training program); Kevin Grumbach & James W. Mold, *A Health Care Cooperative Extension Service: Transforming Primary Care and Community Health*, 301 J. AM. MED. ASS'N 2589, 2589 (2009) (noting that many clinicians "have little or no technical assistance to deploy and maintain new practice improvements like EHRs").

(CME) credits for license re-registration.[269] Many states mandate that clinicians study particular subject-matter in CME courses, such as ethics or pain management.[270] Following this precedent, EHR system training should become a uniform requirement for licensing by all state boards. Because CME credits must be approved by the state, and a certain number must be earned every year or two in most states,[271] such oversight would ensure that clinicians receive updated training. The quality of training courses is important as well. The HITECH Act establishes a Health Information Technology Extension Program and Health Information Technology Regional Extension Centers.[272] These federally-sponsored entities could coordinate training courses to ensure that they include suitable content and are of high value.[273] Formal CME training should be supplemented by other forms of support and assistance offered by the Regional Extension Centers.[274]

2. Agency Guidance

Federal regulations can be supplemented by agency guidance that clarifies and explicates regulatory mandates.[275] Because guidance documents are often developed without the public notice and comment period that is required for federal regulations, they generally do not have the force of law. Rather, they provide needed interpretation, instruction, and policy directions for those enforcing the law and those who must comply with it.[276] Guidance

269. AMERICAN MEDICAL ASSOCIATION, STATE MEDICAL LICENSURE REQUIREMENTS AND STATISTICS 2009 (2009), *available at* http://www.ama-assn.org/ama1/pub/upload/mm/40/table16-2009.pdf. This number includes several Doctor of Osteopathic Medicine boards.

270. *Id.*

271. *Id.*

272. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3012(a), (c), 123 Stat. 115, 247 (2009) (to be codified at 42 U.S.C. § 300jj-32(a), (c)). The Health Information Technology Extension Program is to "provide health information technology assistance services to be carried out through" HHS. The Health Information Technology Regional Extension Centers are to "provide technical assistance and disseminate best practices and other information" to facilitate and promote EHR system use.

273. Pub. L. No. 111-5, § 3012(c)(3)(F), 123 Stat. 115, 249 (2009) (to be codified at 42 U.S.C. § 300jj-32(c)(3)(F)) (urging that instruction concerning EHR systems be integrated "into the initial and ongoing training of health professionals").

274. Pub. L. No. 111-5, § 3012(c), 123 Stat. 115, 248 (2009) (to be codified at 42 U.S.C. § 300jj-32(c)); Grumbach & Mold, *supra* note 268, at 2589–90 (emphasizing the importance of "individualized support" and "technical assistance in the application of EHRs").

275. Lars Noah, *The FDA's New Policy on Guidelines: Having Your Cake and Eating It Too*, 47 CATH. U. L. REV. 113, 122 (1997).

276. *Id.* at 125; Paul R. Noe & John D. Graham, *Reflections on Executive Order 13,422: Due Process and Management for Guidance Documents: Good Governance Long Overdue*, 25 YALE J. ON

documents allow agencies to explain complex or ambiguous regulations quickly and provide a flexible and evolving forum for educating and instructing the public.[277] Thus, guidance is essential to successful regulatory programs.[278]

HHS has already begun the process of producing guidance concerning the HITECH Act. It recently issued guidance on health data security, which identified encryption and destruction of private health information prior to product disposal as essential security tools.[279] Furthermore, the HITECH Act establishes the Health Information Technology Research Center[280] within HHS, which would likely play a key role in producing guidance. If regulations governed the design, approval, and monitoring of EHR systems, then HHS guidance could provide detailed instructions concerning issues such as decision support, data display, and adverse event reporting.

B. ESTABLISHING THE STANDARD OF CARE

Government regulations and guidance will also be useful for establishing the standard of care in medical malpractice cases. The key to successfully defending a malpractice lawsuit is establishing that the defendant met or exceeded the applicable standard of care.[281] Typically, expert testimony is the proof mechanism for the standard of care in malpractice litigation.[282] Both

---

REGS. 103, 108 (2008). Some are concerned that agencies use guidance to circumvent the procedural requirements for promulgating regulations and to avoid judicial review, though occasionally courts have found guidance to be ripe for review and required compliance with it. Appalachian Power Co. v. EPA, 208 F.3d 1015, 1020 (D.C. Cir. 2000) (stating that when guidance is issued, "[l]aw is made without notice and comment, without public participation, and without publication in the Federal Register or the Code of Federal Regulations"); Nina A. Mendelson, *Regulatory Beneficiaries and Informal Agency Policymaking*, 92 CORNELL L. REV. 397, 411 (2007); James Hunnicutt, *Another Reason to Reform the Federal Regulatory System: Agencies' Treating Nonlegislative Rules as Binding Law*, 41 B.C. L. REV. 153, 174 (1999).

277. Mendelson, *supra* note 276, at 408.

278. Noe & Graham, *supra* note 276, at 108; Noah, *supra* note 275, at 125.

279. Breach Notification for Unsecured Protected Health Information; Interim Final Rule, 74 Fed. Reg. 42740 (proposed Aug. 24, 2009) (to be codified at 45 C.F.R. pt. 160, 164).

280. American Recovery and Reinvestment Act of 2009, Pub. L. No. 111-5, § 3012, 123 Stat. 115, 247–50 (2009) (to be codified at 42 U.S.C. § 300jj-32). The purpose of the Center is to "provide technical assistance and develop or recognize best practices to support and accelerate" EHR system adoption and use.

281. *See supra* notes 70–74 and accompanying text.

282. William Meadow, *Operationalizing the Standard of Medical Care: Uses and Limitations of Epidemiology to Guide Expert Testimony in Medical Negligence Allegations*, 37 WAKE FOREST L. REV. 675, 676 (2002) (explaining that jurors are informed about the standard of care "through the testimony of medical expert witnesses" who testify "based upon their own experience, knowledge, and training").

plaintiffs and defendants can present experts to testify about liability and to conduct what some have called "the battle of the experts."[283]

Because EHR systems are an emerging technology that is deployed only to a limited extent,[284] identifying professional custom and the standard of care for their use could be particularly challenging. Nevertheless, the infancy of this industry also presents a unique opportunity to establish reliable and clear EHR system guidelines that will optimize their design, promote their responsible use by clinicians, maximize their utility, and facilitate identification of the standard of care by expert witnesses at trial.[285]

The standard of care for EHR system use could be elucidated not only through governmental requirements, but also through clinical practice guidelines developed by professional organizations. In addition, audit trails built into EHR systems could provide powerful evidence of practices employed by the reasonable clinician and facilitate the development of reliable clinical practice guidelines. Each of these data sources will be discussed below.

1. Regulations, Agency Guidance, and Certification as Evidence of Standard of Care

Federal regulations, agency guidance, and certification[286] can serve as limited evidence of the standard of care in negligence cases. Administrative regulations do not provide definitive proof of the standard of care but constitute relevant evidence of it.[287] A defendant who complied with regulatory requirements may be found negligent if a reasonable practitioner would implement additional precautions.[288] Nevertheless, regulatory compliance is admissible in court as exculpatory evidence for defendants.[289]

283. Mello, *supra* note 73, at 684.

284. *See supra* note 6 and accompanying text.

285. *See* Michelle M. Mello, *Using Statistical Evidence to Prove the Malpractice Standard of Care: Bridging Legal, Clinical, and Statistical Thinking*, 37 WAKE FOREST L. REV. 821, 821 (2002) ("Increasingly, there have been calls to supplement expert opinion testimony in medical malpractice cases with more objective empirical evidence of various kinds to establish the legal standard of care.").

286. *See supra* Section IV.A for discussion of federal regulations, guidance, and certification.

287. Distad v. Cubin, 663 P.2d 167, 176 (Wyo. 1981).

288. RESTATEMENT (SECOND) OF TORTS § 288C (1965).

289. Ashley W. Warren, *Compliance with Governmental Regulatory Standards: Is it Enough to Immunize a Defendant from Tort Liability?*, 49 BAYLOR L. REV. 763, 778 (1997); Richard C. Ausness, *The Case for a "Strong" Regulatory Compliance Defense*, 55 MD. L. REV. 1210, 1241 (1996).

Compliance with agency guidance and certification by well-respected bodies, such as the International Standards Organization, has also been found to have probative value in establishing the standard of care in some areas of the law.[290] Thus, HHS guidance would serve not only to enhance the quality of EHR system use, but also to bolster the defense in medical malpractice cases. Although we argue that CCHIT certification should be replaced by a rigorous regulatory process,[291] in the interim, certification by a recognized authority will likely assist defendants in proving that they have met the standard of care to the extent that they adopted an EHR system of appropriate quality.[292]

2. Clinical Practice Guidelines

Clinical practice guidelines (CPGs) can potentially both educate clinicians on how to optimize EHR system use and constitute evidence of the standard of care. Existing CPGs have been subject to harsh criticism in the past.[293] However, the early stages of development of EHR technology may offer a unique opportunity to formulate CPGs that are objective, sound, and reliable.

a) What are CPGs?

CPGs can be defined as "[s]ystematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances."[294] CPGs relating to diagnostic and treatment practices have been developed by professional societies, such as the American Medical Association and other physician specialty boards; federal and state governmental entities, such as the Agency for Healthcare Research

290. Brandt v. Rokeby Realty Co., 97C-10-132-RFS, 2004 Del. Super. LEXIS 297, at *1, 13, 16 (Del. Super. Ct. Sept. 8, 2004) (stating that EPA guidelines may be "helpful" and constitute "evidence of a standard" though they do not establish a standard of care); John Hedley-Whyte & Debra R. Milamed, *Equipment Standards: History, Litigation, and Advice*, 230 ANNALS OF SURGERY 120, 124 (1999) ("Juries and judges are swayed to the side of the defense by the use of equipment that has been certified to the relevant standard."); Janice M. Hogan & Thomas E. Colonna, *Products Liability Implications of Reprocessing and Reuse of Single-Use Medical Devices*, 53 FOOD & DRUG L.J. 385, 396 (1998); Naomi Roht-Arriaza, *Shifting the Point of Regulation: The International Organization for Standardization and Global Lawmaking on Trade and the Environment*, 22 ECOLOGY L.Q. 479, 516–17 (1995).

291. *See supra* notes 245–52 (critiquing CCHIT and discussing potential alternatives).

292. *See supra* notes 183–86.

293. Mello, *supra* note 73, at 708–09; Hal R. Arkes & Cindy A. Schipani, *Medical Malpractice v. The Business Judgment Rule: Differences in Hindsight Bias*, 73 OR. L. REV. 587, 631–32 (1994).

294. BIOMEDICAL INFORMATICS, *supra* note 12, at 924.

and Quality (AHRQ);[295] and health care payers, including health maintenance organizations and health insurers.[296]

Both plaintiffs and defendants have utilized CPGs in litigation.[297] Courts may view CPGs as establishing a presumption of due care, or at least as evidence of a practice that is accepted by a "respectable" minority.[298] Kentucky state law offers health care providers an affirmative defense based on adherence to CPGs.[299] However, some guidelines include disclaimers, stating that they are only advisory in nature or offer broad parameters rather than specific protocols, and such language significantly diminishes their evidentiary value.[300] Furthermore, several commentators are critical of CPGs in general and argue that they should not constitute reliable evidence of the standard of care in medical malpractice actions.

b) A Critique of CPGs

Critics note that the proliferation of CPGs may make it impossible to discern a clear medical custom.[301] A website called National Guideline Clearinghouse features over 2400 CPGs.[302] CPGs vary in quality and may provide inconsistent guidance concerning treatment of the same condition.[303]

---

295. For information about AHRQ, *see* U. S. DEPARTMENT OF HEALTH & HUMAN SERVICES, AGENCY FOR HEALTHCARE RES. AND QUALITY, WHAT IS AHRQ? (2002), *available at* http://www.ahrq.gov/about/whatis.pdf.

296. Mello, *supra* note 73, at 650.

297. Mello, *supra* note 73, at 648, 668 (stating that "empirical evidence indicates that CPGs currently are being used both as exculpatory evidence (by physician defendants) and as inculpatory evidence (by plaintiffs)," though their use is infrequent); Carter L. Williams, *Evidence-Based Medicine in the Law Beyond Clinical Practice Guidelines: What Effect Will EBM Have on the Standard of Care?*, 61 WASH. & LEE L. REV. 479, 498 (2004) (explaining that courts have allowed both plaintiffs and defendants to introduce CPGs as evidence in litigation).

298. FURROW ET AL., *supra* note 76, at 350.

299. *See id.* The Kentucky statute provision reads as follows:

> Any provider of medical services under this chapter who has followed the practice parameters or guidelines developed or adopted pursuant to this subsection shall be presumed to have met the appropriate legal standard of care in medical malpractice cases regardless of any unanticipated complication that may thereafter develop or be discovered.

KY. REV. STAT. ANN. § 342.035(8)(b) (2006). Florida, Maine, and Minnesota enacted similar provisions, but those were subsequently repealed. FLA. STAT. ANN. § 408.02 (2002 & Supp. 2007); ME. REV. STAT. ANN. Tit. 24 §§ 2971–2979 (West 2000); MINN. STAT. § 62J.34(3)(a) (2005).

300. FURROW ET AL., *supra* note 76, at 350.

301. *See* Mello, *supra* note 73, at 653–54; Williams, *supra* note 297, at 491–92 (2004).

302. National Guideline Clearinghouse, http://www.guideline.gov/browse/guideline_index.aspx (last visited July 27, 2009).

303. Williams, *supra* note 297, at 491–92 (asserting that the sheer number of CPGs hinders physicians and that they vary in quality).

Some will be written with particular agendas in mind.[304] For example, health care payers' CPGs may be designed in part to standardize cost-cutting strategies, such as ordering fewer diagnostic tests for particular symptoms or prescribing less-expensive medications.[305] By contrast, professional societies' CPGs may be partially motivated by a desire to safeguard their autonomy and combat the health care payers' competing guidelines.[306]

Even the most well-established CPGs are not uniformly incorporated into practice and have been shown to be followed by only a narrow majority of physicians.[307] Furthermore, CPGs that are not continuously updated may quickly become obsolete as medical knowledge and technology evolves.[308] Moreover, in order to maintain sufficient flexibility to apply to a broad range of patients, medical practices, and circumstances, CPGs are often worded in vague terms.[309] This is because, the more specific the guidelines are, the more likely they are to be inapplicable to particular circumstances.[310] However, their vagueness can diminish their value for clinicians who are seeking detailed guidance.

Finally, litigants may question whether CPGs intend to represent prevailing medical custom, or, instead, ideals that providers should strive to achieve.[311] If they are ideals rather than a reflection of common clinical practice, they may be inappropriate as evidence of what a reasonable practitioner should be expected to do in particular circumstances.[312]

c) The Opportunity Presented by an Emerging Technology

While CPGs for disease diagnosis and treatment are at times controversial, experts may have a unique opportunity to develop helpful and influential CPGs to guide EHR system use. Very few CPGs exist concerning health information technology, and if the tide of CPG proliferation can be

304. *Id.* at 492 (stating that "[p]otential conflicts of interest may . . . create significant credibility problems with CPGs").

305. *See* Mello, *supra* note 73, at 651.

306. *Id.* at 650–51.

307. *Id.* at 680–83 (asserting that a study of 143 guidelines showed a compliance rate of 54.5%); *see also* Mello, *Using Statistical Evidence to Prove the Malpractice Standard of Care: Bridging Legal, Clinical, and Statistical Thinking*, *supra* note 285, at 844 (arguing that compliance level that far exceeds 50% is required to establish custom).

308. *See* Williams, *supra* note 297, at 487; Arkes & Schipani, *supra* note 293, at 632.

309. *See* Mello, *supra* note 73, at 686–87.

310. Arkes & Schipani, *supra* note 293, at 631–32.

311. Mello, *supra* note 73, at 677; B. Michael Dann, *Jurors as Beneficiaries of Proposals to Objectify Proof of the Standard of Care in Medical Malpractice Cases*, 37 WAKE FOREST L. REV. 943, 949 (2002) (stating that CPGs are "more aspirational in nature than purely descriptive of actual practice").

312. Mello, *supra* note 73, at 677.

stemmed early on, many of the traditional shortcomings of CPGs could be avoided.

A literature search revealed only three U.S.-based CPGs regarding electronic communication between physicians and patients. In 1998, the American Medical Informatics Association (AMIA) developed "Guidelines for the Clinical Use of Electronic Mail with Patients."[313] The guidelines include the following recommendations, among others: (1) establish a specific turnaround time for communication; (2) inform patients about privacy matters, such as who might read messages and whether e-mail will be incorporated into the patient's medical record; (3) articulate what transactions are permitted over e-mail and specify that e-mail should not be sent about urgent matters; (4) ask patients to indicate the subject of the e-mail in the subject line (e.g., prescription, appointment, advice) to facilitate routing; (5) instruct patients to include their name and patient number in the message's text; (6) provide automatic replies to acknowledge receipt of e-mail; (7) inform patients through e-mail that their requests were completed; (8) ask patients to acknowledge reading clinicians' responses through autoreply; (9) word messages carefully to avoid insensitivity to patients and other communication problems; and (10) obtain patient informed consent for e-mail use that includes instructions, descriptions of security mechanisms, and indemnity provisions for providers.[314] The American Medical Association and the eRisk Working Group for Healthcare subsequently issued their own CPGs, which offer similar recommendations.[315]

Unfortunately, a study conducted several years after the AMIA guidelines were published revealed that, as is typical with other CPGs, only a minority of practices are adhering to the recommendations.[316] Nevertheless, as providers become more focused on liability associated with EHR system use,

313. Beverley Kane & Daniel Z. Sands, *Guidelines for the Clinical Use of Electronic Mail with Patients*, 5 J. AM. MED. INFORMATICS ASS'N 104 (1998).

314. *Id.* at 106–07.

315. AMERICAN MEDICAL ASSOCIATION, GUIDELINES FOR PHYSICIAN-PATIENT ELECTRONIC COMMUNICATIONS (2003), http://www.imageamerica.com/downloads/AMAGEC.pdf; ERISK WORKING GROUP, ERISK WORKING GROUP FOR HEALTHCARE'S GUIDELINES FOR ONLINE COMMUNICATION (2006), *available at* http://one.aao.org/asset.axd?ID=03e68ca0-e08e-4e3c-a227-16b0d0714872; *see also*, Amy M. Bovi, *Ethical Guidelines for Use of Electronic Mail Between Patients and Physicians*, 3 AM. J. BIOETHICS W43, W46 (2003); CAN. MED. ASS'N, PHYSICIAN GUIDELINES FOR ONLINE COMMUNICATION WITH PATIENTS (2005), http://oscarresourceplone.oscartools.org/it/pd05-03.pdf.

316. Robert G. Brooks & Nir Menachemi, *Physicians' Use of Email With Patients: Factors Influencing Electronic Communication and Adherence to Best Practices*, J. MED. INTERNET RES. e2 (2006) (finding that only 6.7% of doctors participating in a survey adhered to at least half of 13 selected guidelines).

they may be more motivated to adopt recommended safeguards. The AMIA CPGs and others of similar quality are particularly likely to be followed if they contain detailed, unambiguous suggestions that are not contradicted by conflicting guidelines. Thus, newly created EHR system CPGs that are formulated by well-respected authorities and widely adopted by physicians could serve the dual role of providing valuable guidance to clinicians and establishing professional custom for litigation purposes.

d) A Proposed Approach for CPG Development

CPGs could be developed through an open process and careful evaluation that is coordinated by a central organization. This process would be based on the demonstrably successful model used by the Internet Engineering Task Force (IETF)[317] to select the standards that underlie the operation of the Internet.[318] While federal regulations and guidance would address the initial approval and ongoing monitoring of EHR systems, clinical practice guidelines would provide recommendations concerning clinicians' use practices, such as e-mail communication, cutting and pasting, handling of drug alerts, chart review, and other functions.

The IETF is a technical standardization body, whose work is done by approximately 130 working groups.[319] These groups are open to any member of the public with appropriate expertise who is willing to make the necessary time commitment.[320] Working groups endorse documents through "rough consensus" rather than a formal vote, meaning that "a very large majority of those who care must agree."[321]

IETF standards begin as Internet drafts, which can be submitted by anyone and are distributed for public comments through IETF directories.[322] After sufficient discussion and revision, if the working group leaders believe

317. *See* Internet Engineering Task Force, http://www.ietf.org.

318. *See* Internet Engineering Task Force, *IETF Standards Process,* http://www.ietf.org/IETF-Standards-Process.html.

319. Center for Democracy & Technology, The Internet Engineering Task Force, http://www.cdt.org/standards/ietf.shtml. There are three general IETF meetings each year, designed to reinvigorate the working groups, enable them to mix and meet each other, and ensure that work is accomplished. *Id.*

320. *Id.*

321. *Id.*; Andrew L. Russell, *'Rough Consensus and Running Code' and the Internet-OSI Standards War*, 28 IEEE ANNALS HIST. COMPUTING 48 (2006) (quoting David Clark as describing the IETF philosophy as follows: "We reject kings, presidents, and voting. We believe in rough consensus and running code."); ERIC S. RAYMOND, THE ART OF UNIX PROGRAMMING ch. 17 (2003), *available at* http://www.catb.org/~esr/writings/taoup/html/ietf_process.html.

322. Center for Democracy & Technology, *supra* note 319; RAYMOND, *supra* note 321, at ch. 17.

that rough consensus has been achieved, they will enable the draft to become a Request for Comment (RFC).[323] Drafts that do not advance to the RFC stage are deleted after six months.[324] RFCs, in turn, are corrected by authors and other members of the community through field experience, and the RFC editor ultimately marks those that do not survive field testing as "not recommended" or "superseded."[325] Successful RFCs are those that are "stable, peer reviewed, and have attracted significant interest from the Internet community" and preferably have been proven through implementation experience.[326] The IETF steering committee designates successful RFCs as "proposed standards," and these may be elevated to "draft standard" status.[327] Draft standards that enjoy widespread implementation and general acceptance become Internet standards.[328] In 2003, there were 3000 RFCs and only sixty Internet standards.[329]

The IETF process, therefore, is designed to "engage and empower the broader community" rather than to authorize a single committee to develop guidelines.[330] It also emphasizes the importance of demonstrating standards with working implementations because flaws are far less likely to be detected without the reality check of field testing.[331]

The EHR systems community could develop CPGs in a similar fashion. AMIA or some other professional organization, with support from the Health Information Technology Research Center,[332] could serve the function of the IETF, coordinating working groups and shepherding the CPG development process. Anyone with credible credentials should be able to submit a draft CPG concerning EHR use, which would be distributed to the appropriate working group.[333] Drafts would be posted for public comment and move through several levels of review before being elevated to final

323. *Id.*

324. *Id.*

325. RAYMOND, *supra* note 321, at ch. 17.

326. *Id.*

327. *Id.* (explaining that this change occurs if there are "at least two working, complete, independently originated, and interoperable implementations of a Proposed Standard.").

328. *Id.*; *see also* Internet Engineering Task Force, The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force § 8.4, http://www.ietf.org/tao.html. (last visited Oct. 6, 2009) (describing the process by which standards are created).

329. RAYMOND, *supra* note 321, at ch. 17.

330. Russell, *supra* note 321, at 52.

331. *Id.* at 55 (discussing the importance of "running code" and explaining that it means that "multiple actual and interoperable implementations of a proposed standard must exist and be demonstrated before the proposal can be advanced along the standards track").

332. *See supra* note 280 and accompanying text.

333. Internet Engineering Task Force, *supra* note 328, at §§ 8.1–8.3 (discussing Internet drafts).

CPGs that are endorsed by the authoritative coordinating organization. The review process should require proponents to prove that the CPG was successfully implemented in the clinical setting. For example, those supporting a CPG concerning e-mail use would need to prove that their recommended e-mail handling procedure was satisfactory to a large cohort of patients and clinicians and did not result in an unacceptable number of adverse events. This relatively elaborate development method, however, could only succeed if a standardized EHR user interface existed[334] so that different CPGs would not need to be developed for each separate EHR product. A similar process could be used to establish user interface design guidelines as well as to refine CPGs in light of later experience with them or to modify them in response to technological innovations.

The key differences between the proposed approach and current CPG formulation are the existence of a central CPG coordinating organization, a uniform process for their approval, and an emphasis on field evaluation. Coordination by a single professional organization and approval through a careful, multi-step process, including field testing, would ensure that only the best proposed guidelines become final CPGs. It would prevent EHR system users from being flooded with CPGs that are contradictory, of varying quality, and unreliable. CPGs that are ultimately endorsed should not be met with resistance from the medical and EHR communities because the CPG development process would be inclusive and open to any qualified professional who wishes to propose a CPG or provide public comments. Furthermore, since CPGs would address use practices and not the approval, marketing, or certification of EHR systems, parties with competing financial interests should be able to cooperate in developing CPGs. If a process similar to that of the IETF were established for CPG development, it would be reasonable for courts to allow proof of compliance with final CPGs to establish a rebuttable presumption that the defendant met the standard of care in a medical malpractice case.

3. Audit Trails, User Problem Reports, and the Collection of Data about EHR System Use

The actual EHR systems and reports of user problems should yield significant information about how clinicians typically use the technology.[335]

334. *See supra* note 258 and accompanying text.

335. *See* Hoffman & Podgurski, *supra* note 13, at 154–55 (discussing audit trails and capture/replay capabilities); *supra* notes 263–66 and accompanying text (discussing regulatory requirements for adverse event reporting).

This data will be invaluable for both establishing the standard of care in litigation and developing CPGs.

EHR systems should feature audit trails, which are "generalized recording[s] of 'who did what to whom, when, and in what sequence,' " used to "satisfy system integrity, recoverability, auditing, and security requirements."[336] Effective audit trails would detail all interactions between systems and their users and between different systems. They would be similar in principle to flight data recorders that the Federal Aviation Administration requires for many airplanes.[337] Audit trails are intended to promote system validation and problem diagnosis and resolution. Consequently, these trails should include all system input and output that could affect clinical actions or could reflect the reliability, safety, usability, and security of the system. Audit trails, therefore, would enable litigants and researchers to collect significant information about how EHR systems are operating and being used.[338]

Litigants and courts would need to recognize the limitations of audit trails. These tools provide a one-dimensional view of complex and multi-dimensional processes.[339] They do not capture verbal communication between clinicians and patients, gestures, hand-written notes or instructions given to patients, or other human interactions. Nevertheless, audit trails would provide an unprecedented amount of information about patients' treatment histories.

Federal regulations requiring vendors to submit reports of significant user problems and mandating that summary reports be publicly available would also be useful for establishing the standard of care in medical malpractice cases.[340] Careful analysis of adverse event reports may reveal usability problems or common misunderstandings of a system's interface or displays. Such evidence may assist defendants in proving that a reasonable clinician would not have acted differently same in the circumstances.

---

336. Lawrence A. Bjork, Jr., *Generalized Audit Trail Requirements and Concepts for Data Base Applications*, 14 IBM SYSTEMS J. 229, 229 (1975).

337. 14 C.F.R. § 121.343 (2008).

338. *See* McLean, *supra* note 119, at 77–81 (discussing the use of EHR metadata in medical malpractice litigation).

339. *See* Jorge Aranda & Gina Venolia, *The Secret Life of Bugs: Going Past the Errors and Omissions in Software Repositories*, PROC. OF THE 2009 IEEE 31ST INT'L CONF. ON SOFTWARE ENG'G 307 (2009), *available at* http://portal.acm.org/citation.cfm?id=1555001.1555045&coll=&dl=GUIDE&type=series&idx=SERIES402&part=series&WantType=Proceedings&title=ICSE# ("The histories of even simple bugs are strongly dependent on social, organizational, and technical knowledge that cannot be solely extracted through the automated analysis of software repositories.").

340. *See supra* notes 263–66 and accompanying text (discussing regulatory requirements for adverse event reporting).

Collecting audit trail and adverse event data would be consistent with calls for a change in the way the standard of care is determined for litigation purposes. Some commentators have suggested that the standard of care should be empiricized and ascertained through physician surveys or epidemiologic studies of physician practices.[341] In the words of one author who is a judge, "statistical approaches provide a useful objective check, or yardstick, to use in judging the more subjective opinion evidence introduced by the parties."[342]

Audit trails will constitute a valuable tool for obtaining clear and unbiased evidence concerning commonly used medical practices. Empirical methods for obtaining proof of the standard of care are traditionally cumbersome and may lead to inconclusive results. Ordinarily, records must be pulled, organized, and abstracted by highly trained and highly paid specialists.[343] Some database evidence may also be criticized as representing a patient population that is too small to be statistically meaningful or including too few cases that are factually equivalent to the plaintiff's.[344] By contrast, interoperable EHR systems with audit trails would allow appropriately authorized personnel[345] to access large volumes of data and analyze it through carefully constructed electronic searches. Experts would then base their testimony on abundant records and be able to verify similarity of circumstances through well-crafted queries.

In addition, audit trails and user problem reports could supply information that would be used to formulate CPGs concerning EHR system operation. Researchers who obtain institutional review board approval and informed consent from EHR system users[346] could search audit trails to determine how clinicians are operating EHR systems and which practices

341. William Meadow, *Operationalizing the Standard of Medical Care: Uses and Limitations of Epidemiology to Guide Expert Testimony in Medical Negligence Allegations*, 37 WAKE FOREST L. REV. 675 (2002) (discussing the adoption of a data-based standard of care and explaining that the consequence of doing so "would be to shift the locus of power away from what might be considered an adversarial formulation of standard medical care towards a more rational, scientific view"); Cramm et al., *supra* note 74, at 726 (recommending the employment of physician surveys to determine customary care); Dann, *supra* note 311, at 950–51 (arguing that empirical proof sources will be helpful for jurors).

342. Dann, *supra* note 311, at 951.

343. Mello, *supra* note 295, at 849.

344. *Id.* at 848–49.

345. *See* FED. R. CIV. P. 34(c), 45(a)(1)(C) (establishing that nonparties can be compelled to produce electronic documents through a subpoena).

346. Mello, *supra* note 285, at 849. For a discussion of institutional review boards and informed consent, see Sharona Hoffman, *Regulating Clinical Research: Informed Consent, Privacy, and IRBs*, 31 CAP. UNIV. L. REV. 71, 76–80 (2003).

should be recommended. Summary reports of user problems that are posted on an HHS website would reveal similarly useful information. Thus, CPGs could incorporate the actual experience of large numbers of health care professionals to ensure that the guidelines are clinically relevant and represent best practices that a reasonable clinician could be expected to employ.

The data captured in audit trails and user problem reports could, therefore, influence and bolster expert testimony in two ways.[347] First, the reports would provide independent evidence of the standard of care by showing how practicing clinicians are operating EHR systems. Second, the information could and should be used to develop CPGs, which could in turn be introduced as evidence of professional custom. CPGs would thus be based on practices that are in reality commonly used by health care providers. There is no better proof of professional custom than actual records of what is being done in the field.

The medical profession should not allow the standard of care for EHR system use to be set through isolated medical malpractice decisions that are rarely published and emerge only after years of litigation.[348] Too much is at stake for patients and clinicians. Instead, modern technology could allow the standard of care to be elucidated in a more expedited fashion. Researchers and experts submitting proposed CPGs or CPG revisions would rely on audit trails and user problem reports to facilitate field evaluation. Furthermore, electronic communication will allow swift distribution of final guidelines to every practitioner in the country.

V. CONCLUSION

The highly-touted technology of EHR systems raises serious liability concerns for health care providers at the same time that it excites hope of dramatic improvements in health care outcomes. This Article intends to alert clinicians to the hazards of EHR system use, which cannot be ignored.

Nevertheless, several strategies and techniques can improve both the technology and the practices of those who use EHR systems and thereby diminish the risks of liability. For example, an informed consent process could educate patients about the risks of e-mail, including privacy concerns

347. Mello, *supra* note 285, at 852–53 (asserting that expert testimony would remain indispensable if empirical evidence was used in medical malpractice litigation).

348. *See generally* Peter Siegelman & John J. Donohue III, *Studying the Iceberg from Its Tip: A Comparison of Published and Unpublished Employment Discrimination Cases*, 24 LAW & SOC'Y REV. 1133, 1145–47 (1990) (discussing the process that generates published opinions).

and potential response delays.[349] Likewise, providers utilizing PHRs could ask patients to sign notifications regarding what information will be included in the PHR and to what extent clinicians will review patient input into PHRs.[350] E-mails should be screened by triage nurses, and patients should be advised never to use e-mail for urgent matters such as chest pain.[351] To address concerns about the review of voluminous EHRs in interoperable networks,[352] physicians could assign nurses to read through the records and provide them with summary reports of the patient's medical history, though admittedly, the nurses themselves might miss critical details. In the future, technology may facilitate document summarization, thus alleviating some of the concern about information overload.[353] Technology could also improve screen displays and the effectiveness of drug alerts,[354] and mandatory adverse event reporting could provide invaluable and occasionally life-saving information to purchasers and users of EHR systems.[355] Even the potential feelings of alienation experienced by patients whose doctors lavish attention on computers rather than on them[356] could be partially obviated by strategic choices. For example, doctors could strategically place computers in examination rooms to allow patients to view the screen. This would also allow them to discuss their computer activities so that patients feel that electronic chart review and other EHR work includes them and enhances their care.

The first step to improving EHR systems and reducing clinicians' risk of liability exposure is federal regulation that establishes approval and monitoring processes and EHR system standards and implementation specifications.[357] Federal regulation is essential to ensuring the safety and

349. Bovi, *supra* note 315, at W46; Alissa R. Spielberg, *On Call and Online: Sociohistorical, Legal, and Ethical Implications of E-mail for the Patient-Physician Relationship*, 280 JAMA 1353, 1356–57 (1998).

350. *See supra* note 172 and accompanying text.

351. Spielberg, *supra* note 349, at 1356–57.

352. *See supra* notes 108–11 and accompanying text.

353. Stergos Afantenos et al., *Summarization from medical documents: a survey*, 33 ARTIFICIAL INTELLIGENCE IN MED. 157, 161–73 (2005) (discussing summarization techniques and the challenges that must be overcome); Karen Sparck Jones, *Automatic summarizing: The state of the art*, 43 INFO. PROCESSING AND MGMT. 1449, 1454–58, 1476 (2007) (discussing advances in automatic summarization and its current limitations); Michael Stacey & Carolyn McGregor, *Temporal abstraction in intelligent clinical data analysis: A survey*, 39 ARTIFICIAL INTELLIGENCE IN MED. 1, 18–20 (2007) (analyzing the limitations of temporal abstractions and how it could be improved in the future).

354. *See supra* notes 145–50 and accompanying text.

355. *See supra* notes 263–65 and accompanying text.

356. *See supra* notes 101–06 and accompanying text.

357. *See supra* Section IV.A.1.

integrity of EHR systems, and health care providers should enthusiastically support such regulation. In addition, state regulation should obligate clinicians to undergo EHR system training as part of their CME requirements for license re-registration.

Regulation should be supplemented by agency guidance and CPGs,[358] which will serve the dual role of educating clinicians about proper EHR system use and elucidating the standard of care for litigation purposes. The opportunity to develop authoritative and efficacious guidance is especially ripe given that EHR systems are still in the early stages of development. Thus far, there has been no proliferation of competing CPGs generated by groups with conflicting agendas and varying levels of expertise, and CPGs that are developed responsibly could help optimize the safety and usefulness of EHR systems.[359] To that end, this Article has proposed that a central professional organization coordinate a uniform, multi-step CPG development process.[360] In addition, adverse event reports and the technology built into the systems—audit trails and electronic search features—could provide copious evidence of best practices and could also facilitate CPG formulation.[361] Reliable CPGs and published empirical evidence garnered from EHR systems could elucidate the standard of care for various aspects of EHR system use, providing instruction for clinicians and some degree of predictability in litigation.

EHR systems cannot remain unregulated and largely unscrutinized. Only with appropriate interventions will they become a blessing rather than a curse for health care professionals and patients.

358. *See supra* Sections III.A.2 & IV.B.2.
359. *See supra* Section IV.B.2.c).
360. *See supra* Section IV.B.2.d).
361. *See supra* Section IV.B.3.

THE TROUBLE WITH TROLLS: INNOVATION, RENT-SEEKING, AND PATENT LAW REFORM

Robert P. Merges[†]

ABSTRACT

This Article analyzes the secondary market for patent rights. It defines a patent troll as a participant in this market that does not contribute to the social goal the patent system was meant to serve: technological innovation. The legitimate secondary market, in which patent rights are bought and sold in ways that compensate real innovators (and also often involve the transfer of information and/or technology, in addition to the legal right), is distinguished from the more questionable market for the settlement of lawsuits involving weak, outdated or irrelevant patents. The presence of willing buyers and willing sellers does not necessarily imply that social welfare is being served; at times, the legal system must shut down markets when the things being exchanged have no social value—as in the case of blackmail. The Article reviews the prospects for corrective policies to reign in some activities in the current patent system. Political economy considerations make Congress a long shot to fix the problem, which leaves the courts, and in particular the Federal Circuit. Recent caselaw on damages is presented as a case study of a desirable Federal Circuit course correction involving the secondary market for patents. Economically rational valuation techniques applied to the question of appropriate damages for patent infringement can help to undermine the incentives to litigate, and hence the market for, patents on minor features that can be used strategically to demand large damage awards under some readings of damages doctrine.

---

† Wilson Sonsini Goodrich & Rosati Professor of Law and Technology, U.C. Berkeley School of Law; Co-Director, Berkeley Center for Law and Technology. The author thanks Joe Siino, Alex Cohen, Steve Horowitz and Lisa Mcfall, all of Ovidian LLC of Berkeley, California, for help understanding the secondary market for patents; the editors of the Berkeley Technology Law Journal for editing help; and participants in the Searle Center Research Symposium on Property Rights Economics and Innovation, Northwestern Law School, Nov. 13, 2008, at which I presented an early draft of this Article.

TABLE OF CONTENTS

I. INTRODUCTION

The growth of economic activity surrounding information-based assets has, as theory would predict, led to a strengthening of property rights over those assets.[1] But now, the strengthening of property rights over information assets has also led to a binge of rent-seeking that has put significant pressure on the innovative industries that were the intended beneficiaries of those rights.[2] These glaring problems with the current patent system show how

1. *See, e.g.*, Robert P. Merges, *One Hundred Years of Solicitude: Intellectual Property Law 1900-2000*, 88 CALIF. L. REV. 2187 (2000) (summarizing economic theory tracing a connection between asset value and property strength).

2. On the phenomenon of patent trolls, see Jessica Holzer, *Supreme Court Buries Patent Trolls*, FORBES, May 16, 2006, http://www.forbes.com/2006/05/15/ebay-scotus-patent-ruling-cx_jh_0516scotus.html. For critiques of the activities of trolls, see generally *Patent*

property rights institutions can lose traction with the underlying economic situation they govern. In response, property rights must be constantly and continuously updated, so as to maintain the underlying relationship between increased asset values and the appropriate specification of property rights they occasion. This updating, however, is not a straightforward process; it implicates complex details of political economy, including the optimal division of labor between legislatures and courts, and all of the messy particulars of legislative influence and Congressional action. Among these, one that is quite important is the question of what role courts should play when economic conditions indicate a need for adjustments in property rights specifications, but different industry groups have mutual and reciprocal veto power over legislative enactments. I argue that in the case of damages measures in recent patent reform legislation, we have reached just such an impasse. And I come down on the side of judicial action in the face of the current legislative stalemate.

To some extent, the patent system has already embarked in this direction. The most important indication of this is the Supreme Court's 2006 opinion in *eBay, Inc. v. MercExchange, L.L.C.*[3] That case, which I will refer to often, rejected the "automatic injunction" rule of the Federal Circuit (the unified federal appeals court for patent cases), and replaced it with a flexible test based squarely in the traditions of equitable remedies.[4] The concurrence by Justice Kennedy (joined by three other Justices) contains the crucial rationale for this move.[5] He explained that the threat of an injunction was being used by some plaintiffs in patent cases to extract disproportionate settlements from manufacturers of complex, multi-component products:

> An industry has developed in which firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees. . . . For these firms, an injunction, and the potentially serious sanctions arising from its violation, can be employed as a bargaining tool to charge exorbitant fees to companies that seek to buy licenses to practice the patent. . . . When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply

---

Trolls: Fact or Fiction?: Hearing Before the Subcomm. on Courts, the Internet & Intellectual Property of the Comm. on the Judiciary H.R., 109th Cong. 2 (2006). For an argument that the troll phenomenon is good, not bad, see James F. McDonough III, *The Myth of the Patent Troll: An Alternative View of the Function of Patent Dealers in an Idea Economy*, 56 EMORY L.J. 189 (2006).

3. 547 U.S. 388 (2006).

4. *Id.*

5. *See id.* at 395–97 (Kennedy, J., concurring) (joined by Justices Stevens, Souter, and Breyer).

> for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest.[6]

This is precisely the sort of institutional adjustment I am arguing for in this Article. The Federal Circuit's injunction standard was part of a sweeping strengthening of patent protection which made sense given the increasing importance of intangible assets in overall economic activity. But rent-seeking ensued in the wake of this sweeping change. Micro-adjustments were therefore in order. I believe *eBay* provides an excellent prototype for adjustments of this type in the patent context. When adjustments are made, we see a demonstration of how real-world institutions can adapt property rights to shifting economic conditions. Of course, this is an ongoing process. In the Conclusion, I argue that an adjustment of this sort is now necessary for the doctrines related to damages, and specifically for the need to more rationally apportion damages in patent cases.

II. THE TROUBLE WITH TROLLS

A. DEFINING THE PROBLEM

For some, *eBay* raised a troubling question: Is there really such a thing as a patent troll? Listening to some commentators, one would believe that this label is highly misleading. Some believe the troll label is a meaningless epithet, applied only to a plaintiff in a patent lawsuit with whom one has a legal conflict. Other perfectly legitimate innovators have even argued that they should be classified as patent trolls, as a way of arguing against the troll category altogether.[7] This is a fundamentally misguided effort. I clarify the situation by comparing it to personal injury lawsuits in tort law, and by crafting a careful definition of a true patent troll.

In the early days of tort reform, and even today, trial lawyers often mocked the caricature of the greedy personal injury bar. To hear the trial bar tell it, all plaintiffs in personal injury suits are seeking the same basic remedy: to be made whole from a legitimate injury. For this group, the entire enterprise of "tort reform" is merely an effort to taint a respectable and indeed

---

6. *Id.* at 396–97.

7. *See Patent Trolls: Fact or Fiction?: Hearing Before the Subcomm. on Courts, the Internet & Intellectual Property of the Comm. on the Judiciary H.R.*, 109th Cong. 11 (2006) (statement of Dean Kamen, President, DEKA Research & Development Corporation) ("I only recently found out after reading the definition of a troll that I am one."). He was wrong about this; any reasonable definition of a troll would exclude an innovator of his stature. He was misled by those who claimed that troll status is dependent solely on whether a patentee manufactures and sells his or her own inventions.

honorable area of the law. In the same way, those who attack the very idea of patent trolls argue that this label is employed exclusively by disgruntled defendants whose real objection is to the application of patent law in a particular case.

1. Defining "Patent Troll"

The entire debate has been fed by a lack of clarity in defining the term "patent troll." Partly as a result of the arguments leading up to the Supreme Court's decision in *eBay*,[8] the patent troll label has become associated with the idea of a patentee that does not manufacture a consumer product. Many who attack the troll label make the perfectly plausible point that patent law does not and should not favor patentees who happen to be in the business of manufacturing.[9] In this they are entirely correct. Yet it is nonetheless true that the troll label signifies an important, negative trend in patent law. The true distinction of the troll label concerns the difference between patentees who make real contributions to innovation and those who do not.[10] As we will see below, the troll episode is hardly unique in the annals of patent law; there is a long history of using patents as pure instruments of rent-seeking.[11] The fact is that a number of legal games have emerged through which patents can be employed strictly for unproductive ends. Patentees in this position make little or no contribution to actual innovation. The details of their tactics need not be reviewed here. Suffice it to say that in many industries, the profusion of patent troll litigation threatens the very legitimacy of the entire patent enter-

8. *eBay*, 547 U.S. 388.

9. *See* Robert P. Merges, *Introductory Note to Brief Amicus Curiae in eBay v. MercExchange*, 21 BERKELEY TECH. L.J. 997 (2006) (describing a practical test to determine which patent-case plaintiffs ought not to receive an injunction; the test focuses on the patentee's contribution to research and innovation, and not simply manufacturing).

10. Famed Silicon Valley entrepreneur Judy Estrin, in her recent book, discusses trolls in these terms:

> The country's patent system was created to promote progress by protecting inventors' intellectual property, but nearly everyone now agrees that it is in need of reform. Beginning in the late 1990s, the money spent annually on patent litigation by publicly traded companies exceeded the profits they earned from the patents they have. Significant changes in the existing system will have to be made to mitigate the tensions between different industries, as well as a new breed of "patent trolls" that have made a business out of buying patents on spec, rather than using them to further innovation.

JUDY ESTRIN, CLOSING THE INNOVATION GAP: REIGNITING THE SPARK OF CREATIVITY IN A GLOBAL ECONOMY 170 (2008). Ms. Estrin has founded seven high-tech startups, and is former Chief Technology Officer at Cisco Systems, Inc. Author's Biographical Information available at http://www.theinnovationgap.com/judy-estrin-bio.

11. *See infra* Section II.B.

prise. There is no doubt among most actual innovators that the patent troll label is very real.

One attack on the troll label centers on the idea of market making.[12] The argument here is simple: "trolls" are just middlemen. Their form of arbitrage involves buying patents from those poorly positioned to exploit them, and licensing them to or asserting them against primarily large enterprises, which are in fact making use of the patented technology. There is something accurate here, but something misleading as well. The accurate part is that sometimes valuable technology and good ideas (i.e., innovations) are held by one entity, but could be of use to another such as a large enterprise. If the technology or information is covered by a patent, and if the market maker brings the innovation to the attention of the large enterprise, all is well. Commerce as usual and no complaints. Some who have been accused of "trollery" no doubt fit this classification and ought to be exonerated.

But there is also a problem with the argument that all trolls are just market makers and hence beneficial to economic activity. Not all arbitrage exchange is in fact efficient and socially desirable. For example, someone who engages in blackmail can be seen as an agent of arbitrage. The blackmailer acquires information and brings it to the attention of someone who values it highly (or, more accurately, highly values its nondisclosure). There follows a voluntary exchange after which the parties are, by some measure, both better off. Yet this is not a market making exercise that is efficient. Of course, for the analogy to work, it must be true that patent trolls are selling information with no social value, like the blackmailer. I believe that in some cases at least it is easy to defend this proposition. I defer for later a discussion and analysis of why it is legitimate to shut down a market that contributes nothing to social welfare.[13] At this point, again, my contention is merely taxonomic: There is such a thing as a patent troll—someone who engages in inefficient, socially wasteful patent transactions. I will discuss why that matters later, when I also make some suggestions about which troll-related activities need to be reigned in immediately and which may demonstrate some social value in the long run.[14]

2. Rents: Innovation vs. Litigation

My argument in this Article depends on the idea that the fundamental purpose of patent law is to encourage true innovation. It also depends on the idea that there is a difference between a reward for true innovation and a le-

12. *See infra* Section II.D.
13. *See infra* Section II.D.
14. *See infra* Part III.

gal instrument which permits rent-seeking activities. Only if there is a gap between what is truly innovative and what is permissibly patented and asserted is there space for the concept of a patent troll.

The first proposition—that patents are about innovation—is easy to establish. Recall that our Anglo-American tradition of patent law begins in many ways with the British statute of monopolies.[15] The well-known history behind this statute illustrates that policymakers have long recognized that only patents for true inventions are worthwhile from a social welfare perspective. It is important to remember that in Britain, patents are carved out as an exception to a blanket prohibition on monopolies. Only insofar as a patent represents a true innovation does it qualify for this exception. This history is well known, and forms part of the backdrop for our American system of patent law as well.[16] At the doctrinal level, this fundamental purpose of patent law is built into the fabric of all patent requirements. For example, the nonobviousness requirement is in place to prevent a trivial advance from receiving patent protection.[17] This may be stated in the converse: a patent for a trivial advance would confer illegitimate economic power on its holder, and so is disallowed.

Another example of an ex ante innovation-screening doctrine is the requirement of utility.[18] This has been described as a legal rule that tries to optimize the timing of a property rights award. Building on the seminal work of David Haddock,[19] students of the utility requirement have shown that it is designed to prevent rent-seeking on the part of those who would obtain a patent before a new technology has been adequately described or understood. The obvious rationale for this requirement is that it prevents the dissipation of legitimate rents by requiring those who obtain a patent to show real technological progress. The award of a patent at too early a stage in the innovation process would clearly lead to excessive expenditures of resources in an attempt to draft an early and broad patent instrument. The utility requirement in patent law prevents these wasteful expenditures by requiring that an innovator achieve actual technical milestones prior to receiving a patent. Investment and effort are therefore directed toward the socially useful goal of developing the technology, rather than simply racing to the patent office.

15. ROBERT P. MERGES & JANE C. GINSBURG, FOUNDATIONS OF INTELLECTUAL PROPERTY 13–15 (2004).

16. DOUGLASS C. NORTH & ROBERT PAUL THOMAS, THE RISE OF THE WESTERN WORLD: A NEW ECONOMIC HISTORY 146–48 (1973).

17. *See* 35 U.S.C. § 103 (2006).

18. *See* 35 U.S.C. § 101 (2006).

19. David D. Haddock, *First Possession Versus Optimal Timing: Limiting the Dissipation of Economic Value*, 64 WASH. U. L. Q. 775 (1986).

This is a perfect example of a patent doctrine which prevents rent-seeking at the ex ante stage.

Doctrines directed at restricting the activities of patent trolls—such as the discretionary injunction rule of *eBay*—simply implement this logic at the ex post stage of the patenting process. Many of the tactics of patent trolls take advantage of the fact that a minor innovation which deserves patent protection ex ante may, through changed circumstances, devolve into a legal instrument with powerful rent-seeking potential in the ex post period. It is in these changed circumstances that patent trolls typically operate.

To clarify my point here, I need to say a few words about this ex ante-ex post distinction, and why an economically rational party could not protect himself against the ex post risk. In some ways, the distinction I am talking about is similar to a frequent topic in the economics of contract law. This literature emphasizes the important transition that occurs at the end of bargaining when a contract is signed. Oliver Williamson describes this as the "fundamental transformation."[20] The risk of opportunism accompanying this transition is something that rational contracting parties must always take account of. Williamson and others spend a good deal of effort describing legal and extralegal precautions that can be taken to protect against the ex post risk of opportunism occurring after this transformation.[21] In the same vein, scholars in the "new property rights" tradition write frequently about mechanisms to protect against this same kind of opportunism.[22] In this literature, contracting parties protect themselves by allocating property rights so as to create an effective fallback position for a party who is at risk of opportunism. In all these cases, rational contracting parties can take steps to protect themselves against the risk of ex post opportunism.[23]

Now consider the situation with patent trolls. Here, the ex ante time frame corresponds to the period before a company makes sunk cost investments in any given technology. The ex post time frame is the time after these investments have been made. The patent troll strategy is to take advantage of

20. Oliver E. Williamson, *The Logic of Economic Organization*, *in* THE NATURE OF THE FIRM: ORIGINS, EVOLUTION, AND DEVELOPMENT 90, 98–100 (Oliver E. Williamson & Sidney G. Winter eds., 1991).

21. *See, e.g.*, *id.*

22. *See, e.g.*, OLIVER HART, FIRMS, CONTRACTS, AND FINANCIAL STRUCTURE (1995).

23. For an application of these ideas to the IP context, see Ashish Arora & Robert P. Merges, *Specialized Supply Firms, Property Rights, and Firm Boundaries*, 13 INDUS. & CORP. CHANGE 451 (2004).

"lock-in" that occurs as a result of these investments.[24] Typically, the troll waits until a technology is fully entrenched before scouting around for patents to acquire or asserting the patents it holds. Again, there is nothing intrinsically wrong with this strategy unless the patents at issue do not represent a true innovation. This is, of course, much the same strategy as that pursued by an opportunistic contracting party.

There is no way for an economic actor to protect himself against this strategy in the patent troll context. The key difference between contracting and the patent situation is that in the latter case, information is not only highly asymmetric, but it is virtually impossible to effectively insure against the relevant risk. In particular, there is no way for an economic actor to effectively learn about or anticipate the vast majority of potential patent troll activity. This is so for several reasons. First, patents may be kept secret during the entirety of prosecution,[25] so a clever patentee can suppress the issuance of a patent until a technology matures. Under current law, a troll pursuing this strategy will forego foreign patent rights. This may still be an effective strategy because patent trolls are often willing to sacrifice some coverage in exchange for the advantage of surprise. In addition, while it may be difficult for a contracting party to fully estimate the risk of opportunism, the costs for an innovator facing a patent troll strategy are much, much higher. There are literally millions of patents in force at any time. In a complex field such as commercial software or semiconductors, there are potentially tens of thousands of relevant patents that might be interpreted so as to cover one or more components of a complex product. Because of uncertainty in the process of patent claim construction, it is essentially impossible to screen all the patents that one might infringe. As a consequence, it is much harder to protect against the ex post risk in the patent context. This is why special doctrines and rules to guard against patent troll activity are necessary; self-help is simply impossible in a broad number of cases.[26]

---

24. For a description of a similar phenomenon in the standard-setting context, as well as a suggestion for preventing it, see Robert P. Merges & Jeffrey M. Kuhn, *An Estoppel Doctrine for Patented Standards*, 97 CALIF. L. REV. 1 (2009).

25. Patent applications that will also be filed overseas are published eighteen months after U.S. filing, but those that are only filed in the U.S. will remain unpublished unless the applicant elects otherwise. *See* 35 U.S.C. § 122 (2006).

26. One might argue that the "patent protection racket" industry that has emerged provides insurance against this risk. I would argue in response that this form of "insurance" is of questionable social value if the only risk insured against is rent-seeking litigation. *See infra* Part III. It should be noted, however, that this is true only when these "insurance" companies are simply selling freedom from lawsuits under questionable patents. To the extent that these companies help create an "exit option" for small inventors and companies that have tried and failed to introduce innovative products on the market, or use the proceeds

B. HISTORICAL EXAMPLES OF PATENT-BASED RENT-SEEKING

1. Early History

There is a long tradition of rent-seeking based on the acquisition of patents. Several episodes in the history of patent law are well documented in this respect. The first extensive episode of rent-seeking in the history of patent law came about after the 1793 patent act was passed. Rent-seeking was possible under this statute because patents were registered by the patent office, instead of being examined. This essentially threw all problems of patent validity into the courts. The cost of litigation was such that nuisance suits proliferated, as any economist would predict. The solution was to reinstitute patent examination as part of the Patent Act of 1836.[27] During the middle years of the nineteenth century, a controversy erupted over the misuse of the patent re-issuance procedure. As with the 1836 Patent Act, the solution here was also legislative: major reforms changed the standards for granting a patent re-issue, eliminating many opportunities for rent-seeking.[28]

At the same time, the U.S. Supreme Court was confronted with a growing number of patent cases, many arising out of the easily manipulated registration system of the 1793 Act. The Court had very few doctrinal tools for weeding out low-quality patents. Basically, only the two statutory elements of utility and novelty were required for a patent to be valid. Against this background, the Supreme Court decided *Hotchkiss v. Greenwood* in 1851, creating the "invention" test.[29] This was a wholly new standard that made it substantially more difficult for an inventor to obtain a patent. While the case did not specifically mention the flood of patents as a reason for stating the new requirement, it is widely acknowledged now that this was a factor in its thinking.[30]

Two other episodes from the nineteenth century are also worth mentioning. First, during the 1860s and 70s, a number of entrepreneurial business people acquired patents of dubious utility which covered widely used agricul-

---

from their activities to fund productive activities such as future-oriented R&D, things may be a bit more complex. *See infra* Part IV.

27. *See generally* EDWARD C. WALTERSCHEID, TO PROMOTE THE PROGRESS OF USEFUL ARTS: AMERICAN PATENT LAW AND ADMINISTRATION, 1787–1836 (1998) (describing lead-up to 1836 Patent Act).

28. Steven Lubar, *The Transformation of Antebellum Patent Law*, 32 TECH. & CULTURE 932, 944 (1991).

29. 52 U.S. 248, 267 (1851). The "invention" test was the historical precursor of today's "nonobviousness" requirement.

30. *See, e.g.*, Edmund W. Kitch, *Graham v. John Deere Co.: New Standards for Patents*, 1966 SUP. CT. REV. 293 (1966) (describing general trends leading up to the Supreme Court's decision in *Hotchkiss*, 52 U.S. 248).

tural techniques. These economic actors—who came to be known as "patent sharks"—created an enormous upheaval in the agricultural sector, leading to a populist outcry against the entire patent system.[31]

According to a recent account of the patent shark episode, when the Patent Office decided to permit patents on minor ornamental design features in the late nineteenth century, patent applications spiked sharply upward. The volume of applications, together with the lowering of standards for patents on designs, made it easy for patentees to acquire design patents on modest new designs for familiar farm tools, including "crowbars, spades, plows, scrapers," and others.[32] This spate of poor quality patents on farm implements created a business opportunity, which entrepreneurs quickly seized on. As with today's trolls, most of the resulting litigation "came [not from inventors or their companies, but from] third parties that specialized in litigation and bought up the dormant patents."[33] Importantly, there is no evidence that the creation of a secondary market for simple agricultural implement patents led to significantly greater innovation in that field, which had already undergone rapid modernization and which was characterized by a wave of large-scale mechanization that far exceeded the scope of these simple design patents.

Second, a similar episode took place in the railroad industry in the late nineteenth century.[34] At the time, this industry was characterized primarily by internal research and development teams. Formal research and development, and use of the patent system, was relatively unknown in the early years of the railroad industry.[35] Outside inventors often developed and submitted new technologies to large incumbent railroad lines. In some cases, these technologies were in fact innovative and patents facilitated new entry into the industry. The Westinghouse Company, which developed the innovative triple

---

31. *See* Gerard N. Magliocca, *Blackberries and Barnyards: Patent Trolls and the Perils of Innovation*, 82 NOTRE DAME L. REV. 1809, 1811 (2007) (describing the rise of patent sharks).

32. *Id.* at 1821 (quoting HECTOR T. FENTON, THE LAW OF PATENTS FOR DESIGNS 224, 259 (1889)). *See also* Gerard N. Magliocca, *Ornamental Design and Incremental Innovation*, 86 MARQ. L. REV. 845, 874–79 (2003) (describing the ill-fated attempt between the 1860s and 1880s to classify farm implements as items of industrial design, and hence qualified to receive utility patents from the Patent Office).

33. Magliocca, *supra* note 31, at 1823.

34. Robert P. Merges, *The Uninvited Guest: Patents on Wall Street*, 88 FED. RES. BANK ATLANTA ECON. REV. 1, 7–8 (2003) (describing the disruptive effect of patents in the nineteenth century on the railroad industry).

35. STEVEN W. USSELMAN, REGULATING RAILROAD INNOVATION: BUSINESS, TECHNOLOGY, AND POLITICS IN AMERICA, 1840–1920, at 117 (2002) (accounting of the patent battles that assailed the railroad industry in the late nineteenth century, and the two-pronged response—legislative and judicial—that ultimately succeeded).

valve air brake under the direction of George Westinghouse, is perhaps the most famous example.[36] However, in many other cases, patents were developed and acquired that made essentially no contribution to the technological development of the industry. The railroad industry responded to this development with a dual track approach: legislation was introduced to prevent the most egregious practices, and many cases were pursued through the courts and ultimately to the Supreme Court. In the end, a significant Supreme Court case ended one of the most destructive practices of the railroad industry patentees.[37] In this case the Court rejected a theory of patent damages—based on a controversial measure of "cost savings" that juries often used to jack damages up far beyond any reasonable measure—that had proven quite lucrative to the outside patentees.[38]

The number of patents awarded for various aspects of railway technology grew steadily throughout the nineteenth century.[39] A modest number of "outside inventions" were adopted by the railroads during this period. But the patent system really burst into prominence when courts began awarding huge damage awards to the holders of patents who had sued the railroads.[40] In the wake of several much-discussed infringement suits, patent matters rose to the highest levels of discussion within the railroad companies. According to the leading historian of this era,

> The mounting array of patents constituted an expanding minefield of potential lawsuits and financial liabilities.
>
> During the decade following the Civil War, railroads [which had traditionally exchanged information freely] and the patent system raced forward on a collision course. . . . With the number of patents proliferating . . . railroads [were] exposed to new liabilities of unprecedented scale.[41]

---

36. *Id.* at 130–31.

37. Ry. Co. v. Sayles, 97 U.S. 554 (1878).

38. *Id.* at 555–56.

39. JACOB SCHMOOKLER, PATENTS, INVENTION, AND ECONOMIC CHANGE 140–155 (Zvi Griliches & Leonid Hurwicz eds, 1972).

40. *See generally Sayles*, 97 U.S. at 555–56 (1878) (summarizing district court proceedings from 1865 through 1875); *In re* Cawood Patent, 94 U.S. 695 (1877) (concerning patent for "swedge-block" used to repair and straighten worn railway rails).

41. USSELMAN, *supra* note 35 at 101. This led one industry member to write that "Patents . . . will be the death of me!" *Id.* at 117 (quoting D.L. Harris, President, Connecticut River Railroad, Dec. 23, 1868). *See generally* Steven W. Usselman, *Patents Purloined: Railroads, Inventors, and the Diffusion of Innovation in 19th-Century America*, 32 TECH. & CULTURE 1047 (1991) (describing the coming of patents to the railroad industry).

The Supreme Court caught wind of this discontent, and corrected course in the late nineteenth century. Though not drawn from the railroad industry, an 1883 Supreme Court case condemned patent-based rent-seeking in no uncertain terms, and captured the spirit of Court-led patent reform during this era:

> The design of the patent laws is to reward those who make some substantial discovery or invention, which adds to our knowledge and makes a step in advance in the useful arts. Such inventors are worthy of all favor. It was never the object of those laws to grant a monopoly for every trifling device, every shadow of a shade of an idea, which would naturally and spontaneously occur to any skilled mechanic or operator in the ordinary progress of manufactures. Such an indiscriminate creation of exclusive privileges tends rather to obstruct than to stimulate invention. It creates a class of speculative schemers who make it their business to watch the advancing wave of improvement, and gather its foam in the form of patented monopolies, which enable them to lay a heavy tax upon the industry of the country, without contributing anything to the real advancement of the art. It embarrasses the honest pursuit of business with fears and apprehensions of concealed liens and unknown liabilities to lawsuits and vexatious accountings for profits made in good faith.[42]

Despite these nineteenth century reforms, the early turn of the century automobile industry also suffered its period of patent extortion. It took the form of a patent issued to patent lawyer George Selden.[43] The Selden patent on an automobile design had as its key claim the use of a light, gasoline powered internal combustion engine. The claim was quite general, failing to specify many important details about the engine. The Patent Office allowed that claim, and district courts upheld it twice, despite arguments that the broad idea was obvious, and that the engine referred to in the claim was of a particular kind not encompassing all the engines that were claimed to infringe. Eventually, the Second Circuit drastically narrowed the claim, stating that it covered only the particular kind of gasoline engine used by Selden.[44]

Many in the industry—in particular, Henry Ford—hated the Selden patent and all that it stood for.[45] Although the Selden patent was eventually nar-

---

42. Atl. Works v. Brady, 107 U.S. 192, 200 (1883).

43. Road Engine, U.S. Patent No. 549,160 (filed May 8, 1879) (issued Nov. 5, 1895).

44. For the relevant history, see generally Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 889–90 (1990).

45. JAMES J. FLINK, AMERICA ADOPTS THE AUTOMOBILE, 1895–1910, at 323–25 (1970).

rowed,[46] and thus made irrelevant,[47] this did not occur until late in the patent's life. For many years automobile manufacturers paid royalties begrudgingly. But did the presence of the Selden patent actually hinder technological progress in the industry? That is perhaps a bit more speculative. Law suits based on it surely did absorb considerable time and attention of people like Henry Ford, whose production methods revolutionized the industry. Perhaps more importantly, smaller firms may have been put off by the threat of suit. At this early stage in the history of the technology, those firms that left the industry or chose not to enter may well have taken valuable improvements with them. In any event, the Selden episode has often been held up as a prime example of rent-seeking through patent assertion.

2. Recent History: "Patent-Oriented" Strategies in the Early Biotech Industry

In the 1980s, the name of the game in the biotechnology industry was to isolate and sequence important naturally-occurring genes that produced useful proteins. Erythropoetin (Epo) was one such protein. A then-small biotechnology company called Amgen was the first to isolate the Epo gene, clone it, and express Epo in clinically effective quantities.[48]

A small rival named Genetics Institute (GI), though behind in the race to sequence the Epo gene, conceived of a strategy to overtake Amgen. GI filed a patent on "isolated and purified" Epo, derived by non-genetic engineering techniques.[49] When the patent issued, GI sued Amgen. Amgen counterclaimed on the strength of its own patent to the gene sequence and associated protein.[50]

GI actually had a tenable claim, based on conventional patent law. Technically speaking, the fact that the isolated protein was derived without genetic engineering techniques was irrelevant; the only relevant question was whether the Amgen protein fell within the specified purity ranges claimed by GI, and it appeared that it did.

However, the courts—like most observers—understood full well that Amgen was the scientific pioneer, not GI. There was a general perception

46. Columbia Motor Car Co. v. C.A. Duerr & Co., 184 F. 893, 908–09 (2d Cir. 1911).

47. *See* Merges & Nelson, *supra* note 44 (describing lawsuit late in the life of the patent that substantially narrowed the patent and thus permitted competitors to operate without a license).

48. Michael Rosen, *The Birthplace of Biotech: San Francisco, Boston, Geneva, or Chicago?*, WTN NEWS, Aug. 25, 2004, http://wistechnology.com/articles/1118/ (explaining the early history of the biotech industry).

49. Amgen, Inc. v. Chugai Pharm. Co., 927 F.2d 1200, 1203–04 (Fed. Cir. 1991).

50. *Id.* at 1204.

that GI was attempting to use a clever legal strategy to jump ahead of Amgen.

The Court of Appeals for the Federal Circuit found in favor of Amgen by invalidating the GI patent on the ground that GI had not enabled the wide purity range it had claimed.[51] While the ruling was technical in nature, it seems implausible that it was not influenced by the underlying facts and broad equities of the case. Ultimately, the clever patent strategy lost out to the true scientific innovation. Amgen profited mightily, as Epo grew into a $2.5 billion per year pharmaceutical product.[52]

A similar episode involved claims to short gene sequences or "ESTs." A clever patent strategy emerged in which firms filed patents on short snippets of genes whose function and relevance were as yet unknown. The idea was simple: obtain enough patents like this, and some were sure to cover portions of genes that turn out later to have important medical uses. When those genes were identified and cloned, and therapies based on them were developed, the owners of these patents would profit handsomely.

The objection to this strategy was that these patents would give their owners a reward highly disproportional to their actual intrinsic value. Patents such as this would only become lucrative when later researchers revealed the full gene of which they are a part, and discovered the medical significance of the gene. These EST patents were valuable only as holdup rights. This led several commentators to argue that EST patents ought to fail the utility requirement in patent law[53]—an argument that the Federal Circuit later accepted.[54]

C. TECHNOLOGY MARKETS AND RENT-SEEKING

Because many industry players defend today's patent trolls on the grounds that they are merely (beneficial) "market makers," it is a good idea to pause here for a moment to see what can be learned from the story of the "patent sharks." In pioneering work on the nineteenth century "market for technology," Naomi Lamoreaux and Kenneth Sokoloff discovered a dense network of independent inventors, patent lawyers, and corporate buyers that

51. *Id.* at 1217.

52. *See* FundingUniverse.com, Amgen Inc. Company History, http://www.fundinguniverse.com/company-histories/Amgen-Inc-Company-History.html (last visited, Oct. 18, 2009) (tracking growth of Amgen's EPO sales, sold under its trade name Neupogen, from $53 million before the GI lawsuit to over $1 billion per year by the mid 1990s).

53. Rebecca S. Eisenberg & Robert P. Merges, OPINION LETTER AS TO THE PATENTABILITY OF CERTAIN INVENTIONS ASSOCIATED WITH THE IDENTIFICATION OF PARTIAL CDNA SEQUENCES, 23 AIPLA Q.J. 1 (1995).

54. *In re* Fisher, 421 F.3d 1365 (Fed. Cir. 2005).

helped create an active market for technology in the era before large, vertically integrated corporate research and development (R&D) establishments.[55] It is tempting to fit what came to be known as "sharks" into this framework, dismiss the inflamed rhetoric of the time as excessive and short-lived, and conclude that the system for the most part worked just fine.

I think this would be a mistake. It ignores the real dislocation felt by an entire class of economic actors—small farmers—and the resulting damage to the image and integrity of the patent system. It would also direct our attention past an important issue: the precise mechanism by which this rent-seeking threat was pushed back. After all, as things developed, it was important that the patent system did not succumb to a populist movement that would have weakened or eliminated it. Otherwise it would not have been in place to stimulate and participate in the revolutionary technological developments at the turn of the twentieth century.

Gerard Magliocca is correct that the change in standards for design patents led to the rent-seeking episode of the "patent sharks." But he is wrong about two related issues.[56] First, as my research on the nineteenth century railroad industry shows,[57] the agricultural-industry "patent sharks" were not, as he claims, the only nineteenth century analogue to today's patent trolls. Other rent seekers were operating at the same time as the agricultural sharks. And second, the elimination of an entire category of patents is not the only effective way to end a rent-seeking episode. The "surgical" intervention of the Supreme Court in railroad industry patent litigation during this same era shows that less drastic legal changes can be effective.[58] This is crucial to remember. As Magliocca himself recognizes, there are potentially significant costs to his preferred policy fix: negative impacts on an entire segment of industry when its incentive for R&D is reduced by the elimination of patents over an entire category of technology.

D. BUT WHY WOULD WE INTERFERE WITH THE "MARKET FOR PATENT RIGHTS"?

One obstacle to confronting the troll problem is that trolls and their defenders have constructed a superficial defense for their activities. The de-

55. Naomi R. Lamoreaux & Kenneth L. Sokoloff, *Long-Term Change in the Organization of Inventive Activity*, 93 PROC. NAT'L ACAD. SCI. U.S. AM. 12686, 12686–92 (1996); Naomi R. Lamoreaux & Kenneth L. Sokoloff, *Inventors, Firms, and the Market for Technology: U.S. Manufacturing in the Late Nineteenth and Early Twentieth Centuries*, *in* LEARNING BY DOING IN FIRMS, ORGANIZATIONS, AND NATIONS 19 (Naomi Lamoreaux et al. eds., 1998).

56. *See* Magliocca, *supra* note 32.

57. *See supra* Section II.B.

58. *Id.*

fense is based on the idea that trolls are performing a valuable market-making function. In their telling, they identify undervalued patents and invest time and effort marketing those patents to other firms. It sounds appealing, a simple case of arbitrage.[59] In this story, the enemies of the trolls are firms that have simply missed the boat on this valuable new market. Now those enemies are taking aim at a viable, functioning market for undervalued patents. Ultimately, the trolls argue, their enemies cannot be in the right, because the enemies' solutions to the troll problem all hinge on shutting down this emergent, well-functioning market.[60] The basic logic is that, now that trolls have pioneered a market for a new class of assets, these enemies want to obliterate it, and return to the days when ideas could be obtained for free.

The basic premise behind this defense is surely correct. There is no reason at all not to encourage and support a well-functioning market for patentable inventions. And, given the well-known advantages that accrue from specialization, there is no legitimate reason to discriminate legally between a firm that embeds its innovation in manufactured products, and one that sells its innovations in disembodied form—a pure idea shop.[61]

But this conventional account of the advantages of specialization may not account for at least part of the contemporary patent troll industry. Many patent assertion companies do not perform research and development as those terms are commonly understood. They do not participate in the growth of knowledge and technology. True trolls do not really innovate at all. They are opportunistic litigation mills, not research firms. They cloak themselves in the legitimacy of patents, exploiting the widespread perception that where there is a patent there must be innovation. Sadly, this is not always true.[62]

---

59. *See, e.g.*, James F. McDonough III, *The Myth of the Patent Troll: An Alternative View of the Function of Patent Dealers in an Idea Economy*, 56 EMORY L.J. 189 (2006) (defending trolls as efficient market-makers).

60. *Id.* at 190.

> [McDonough's] Comment argues that, contrary to popular belief, patent trolls actually benefit society. These trolls act as a market intermediary in the patent market. Patent trolls provide liquidity, market clearing, and increased efficiency to the patent markets—the same benefits securities dealers supply capital markets. Ultimately, . . . the emergence of patent trolls is simply a stage in the natural evolution of the patent market.

Id.

61. Indeed, I have provided a spirited theoretical defense for just such firms, highlighting the role that patents can play in making them economically viable as standalone firms. *See* Arora & Merges, *supra* note 23; Robert P. Merges, *A Transactional View of Property Rights*, 20 BERKELEY TECH. L.J. 1477 (2005).

62. *See, e.g.*, John R. Allison & Mark A. Lemley, *Empirical Evidence on the Validity of Litigated Patents*, 26 AIPLA Q.J. 185, 205 (1998) (noting in a study of 300 litigated patents, 46% were found invalid).

And this stark fact explains succinctly why the market for true troll activity is not worth defending. It is a market for a product that has no social value at all. In this, the trolls are reminiscent of another famous agent of "arbitrage," the rent-seeking personal injury lawyer.[63] The market for concocted, unfounded litigation is not one that society ought to encourage and the ongoing tort reform movement is directed specifically at shutting it down.[64]

The analogy to spurious personal injury settlements or nuisance suits brings home the key point: The market for patents unconnected to innovation is not a market that the legal system ought to encourage or even tolerate. In this sense, tort litigation is an excellent analogy.[65] But to address the broader point—that solving the troll problem will involve shutting down a functioning market—it might help to look to another, equally apt example: the case of blackmail.[66]

As a legal matter, blackmail has fascinated scholars for a long time. It raises some famously knotty problems of individual versus social harm. But for an economist, the puzzling aspect of blackmail is that it involves a voluntary and seemingly Pareto-satisfying exchange. The blackmailer has information the blackmailee wants; they agree to a price; and the deal is done. From the point of view of libertarian theory, if not pure market exchange, what's not to like?

After some discussion of these issues, the answer came clear enough to Ronald Coase when he wrote about blackmail in 1984.[67] He emphasized the social wastefulness of blackmail transactions: "Blackmail involves the expenditure of resources in the collection of information which, on payment of blackmail, will be suppressed. It would be better if this information were not collected and the resources were used to produce something of value."[68]

Even if no resources were expended to acquire the information—if it dropped fortuitously into the blackmailer's hands, for instance—Coase em-

63. See the discussion regarding personal injury lawyers and tort reform, *supra* Section II.A.

64. Again, this assumes that the original inventor receives either nothing from the troll or very little, and hence that payments to the troll do little or nothing to stimulate or reward real invention and innovation. When a substantial portion of troll income does pass to real innovators, the story changes, and trolls may be more defensible. *See infra* Part III.B.2.

65. For a sophisticated proposal based on the troll-tort suit analogy, see Ranganath Sudarshan, *Nuisance-Value Patent Suits: An Economic Model and Proposal*, 25 SANTA CLARA COMPUTER & HIGH TECH. L.J. 159 (2009).

66. See discussion of blackmail, *supra* Section II.A.1.

67. Ronald H. Coase, *The 1987 McCorckle Lecture: Blackmail*, 74 VA. L. REV. 655 (1988).

68. *Id.* at 674.

phasizes that the transaction would still be wasteful.[69] In fact, he goes further than that. In a statement that is strong medicine indeed for a dyed-in-the-wool economist, he condemns blackmail as something more than inefficient. He says it is wrong.[70] And for Coase, this justifies the classification of blackmail as not only a private offense, but a crime.

Blackmail is part of a broader pattern in which the legal system sorts out which voluntary transactions ought to be enforced. Where the underlying purpose of the exchange is legitimate or productive, there is no question that enforcement of bilateral exchange relations ought to be a matter of course. But where there is no social welfare gain possible from the exchange, and especially where enforcement encourages wasteful expenditures (again from the perspective of social welfare), there is good reason not to promote voluntary exchange.

Judge Richard Posner has at times echoed this same concern. In discussing the criminal law, for example, he has talked about why the law does not encourage sterile, purely redistributive "exchange."[71] In a similar vein, in a case on trade secret law, Judge Posner addressed the requirement that a trade secret owner take "reasonable precautions" to prevent a given piece of information from becoming widely known. He explained this element of a trade secret cause of action in terms of two related but distinct theories of trade secret law—both of which reflected an understanding of the importance of segregating out productive from unproductive interactions:

---

69. *Id.* ("While it is true that in such a case no resources were used to collect the information, resources would certainly be employed in the blackmailing transaction.").

70. *Id.* at 675–76. Coase explains that

> [t]he blackmailer's actions generate fear and anxiety—blackmailing involves more than the employment of resources which leave the value of production unchanged—it causes real harm which reduces the value of production . . . . The victim, once he succumbs to the blackmailer, remains in his grip for an indefinite period. It is moral murder. . . . [I]t is only certain threats in certain situations which cause harm on balance and in which the harm is sufficiently great as to make it desirable that those making them should be prosecuted and punished. I think it is clear what is wrong with blackmail. The problem is to know how to deal with it.

Id.

71. *See, e.g.*, Posting of Richard Posner to The Becker-Posner Blog, Crime and Corruption—Posner's Comment, http://www.becker-posner-blog.com/archives/2007/05/ (May 6, 2007 19:55).

> The basic economic objection to crime is that a crime is a costly but sterile transaction. It redistributes wealth, which doesn't increase the size of the social pie; and therefore the costs involved in crime—the time and other inputs of the criminal, and the defensive measures taken by potential victims—are a deadweight loss to society.

Id.

> It should be apparent that the two different conceptions of trade secret protection are better described as different emphases. The first emphasizes the desirability of deterring efforts that have as their sole purpose and effect the redistribution of wealth from one firm to another. The second emphasizes the desirability of encouraging inventive activity by protecting its fruits from efforts at appropriation that are, indeed, sterile wealth-redistributive—not productive—activities. The approaches differ, if at all, only in that the second does not limit the class of improper means to those that fit a preexisting pigeonhole in the law of tort or contract or fiduciary duty—and it is by no means clear that the first approach assumes a closed class of wrongful acts, either.[72]

This emphasis on the importance of sorting out productive from unproductive transactions goes back far beyond Coase and Posner, though in former times the language of efficiency was more thoroughly intertwined with concepts of virtue and morality. It is a consistent theme in the writings of Adam Smith, for example. He always tempered his belief in the importance of self-interest with discussion of ethical virtues, such as justice and prudence. As the economist Deirdre McCloskey has noted, these features of Smith's thought actually form a crucial underpinning for well-functioning capitalist economies.[73] This aspect of Smith's thought is perhaps best captured in a little ditty he included in The Theory of Moral Sentiments: "So Vice is beneficial found/when it's by Justice lopt and bound."[74] Others have noted the same theme, emphasizing the importance to Smith of institutional—including legal—rules and frameworks that channel self-interest and promote collectively beneficial exchange and commerce. The philosopher William Campbell wrote,

> Smith never glorifies selfishness, greed, and an unbridled pursuit of personal gain, either in the Moral Sentiments or in the Wealth of Nations. It is the purpose of Smith's moral, legal and economic thought to devise the appropriate institutional framework within

72. Rockwell Graphic Sys., Inc. v. DEV Indus., Inc., 925 F.2d 174, 178 (7th Cir. 1991).

73. DEIRDRE N. MCCLOSKEY, THE BOURGEOIS VIRTUES: ETHICS FOR AN AGE OF COMMERCE 407–15 (2006).

74. ADAM SMITH, THE THEORY OF MORAL SENTIMENTS 357 (1759). This theme is also apparent in WILLIAM J. BAUMOL, ROBERT E. LITAN & CARL J. SCHRAMM, GOOD CAPITALISM, BAD CAPITALISM AND THE ECONOMICS OF GROWTH AND PROSPERITY 252 (2007) (discussing ways to "reduce the incentives for enterprising class action[] [lawsuits] that, in effect, blackmail defendants with deep pockets").

> which self-interest can be expressed without inflicting harm on other individuals.[75]

From this traditional perspective in the history of economic thought, it is quite apparent that we should not be blinded by fears of shutting down or regulating an existing market. The market for patents unrelated to innovation adds nothing to overall social welfare. Rent seekers who employ patents are often said to engage in a form of extortion.[76] When a charge like this is true, conventional wisdom suggests only one efficient (and proper) course of action: shutting the socially wasteful market down.

E. SUMMARY: HISTORY LESSONS

In all these cases, rent-seeking is made possible by the nature of patent law and its relationship to technological inventions. It is an inherently difficult and complex task to divide up a stream of technological innovation into discrete property bundles. It is difficult to describe particular increments of technological advance in clear and precise language. As a result, the costs of establishing and enforcing property rights in this area are inherently high. Patent examiners, administrative law judges within the patent system, and federal judges generally are of course not experts in any particular technology. This reality, coupled with the inherent complexity of the enterprise, means that there are numerous opportunities to creatively define and apply patent claims. In practice, clever lawyering can often produce a patent claim that covers more technological ground than is truly warranted by the underlying invention.

Of course, numerous patent doctrines exist to police this activity. But the history of patent law shows that these doctrines do not always do an adequate job of preventing rent-seeking. At certain times, and for various reasons, the patent system is overwhelmed with rent-seeking activities. During these times, the normally effective doctrines of patent law do not serve their appointed function. This leads to extensive rent-seeking episodes such as the ones I have just described. In my opinion, the current wave of patent trolls shows that we may very well be undergoing another of these episodes right now.

From the perspective of property rights theory, this can be explained quite simply. These episodes show that measurement costs at times increase

---

75. William F. Campbell, *Adam Smith's Theory of Justice, Prudence, and Beneficence*, 57 AM. ECON. REV. 571, 572 (1967).

76. *See, e.g.*, USSELMAN, *supra* note 35, at 111 (Owners of patents on train brakes "extort money from railroad companies under the pretense of a patent which they know must be invalid" (quoting Expert Report of John Cochrane, Baltimore & Ohio Railroad, 1860)).

so as to put pressure on the overall functioning of the property rights system.[77] That is, holding the value of underlying assets constant, an increase in the cost of measuring and enforcing property rights (which is one way to characterize the combination of new technologies and pressures on the patent system which accompany these rent-seeking episodes) can be expected to lead to a change in the specification of property rights. But here we encounter a practical problem with the theory. Property rights regimes are not so fine grained that they can self-adjust to micro-level changes such as this. Indeed, there is ample theory to demonstrate that we would not want them to. So for example, while it might be optimal to eliminate patents for certain technologies when the measurement costs associated with them have undergone a rapid increase, it is practically impossible to do so. For example, there would be all kinds of difficulties in carving out railroad technology from other industrial technology. In addition, problems like this are often short lived. Once the patent system adjusts to the new technology, it might make sense to reinstitute property rights. But again, property institutions cannot be calibrated so finally or changed so frequently. Stability of expectations is important too.

What this means practically is that internal adjustments must often be made that carry out, as far as possible, the optimal recalibrations suggested by the theory. In the historical examples described earlier, there is good evidence that just such recalibrations in fact took place.[78] And I argue in this Article that, as we find ourselves in a similar situation today with patent trolls, we need to look for ways to effect similar recalibrations.

III. WHAT ABOUT PRIVATE SECTOR COURSE CORRECTION?

One might accept that the specification of property rights has deviated in some way from the optimal, yet still refrain from advocating any self-conscious course correction or affirmative policy response. Perhaps the property rights system will self-correct. Firms and individuals may have some techniques for mitigating the effects of inefficient property rights specifications. If so, there may be no need for a public policy response.

---

77. *See, e.g.*, OLIVER E. WILLIAMSON, THE ECONOMIC INSTITUTIONS OF CAPITALISM 29 (1985) (discussing the "measurement branch" of transaction cost economics); Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. (PAPERS & PROC.) 347 (1967) (discussing the importance of measurement costs).

78. For further discussion of recalibration in IP law, see Merges, *supra* note 1; Robert P. Merges, *Intellectual Property Rights and the New Institutional Economics*, 53 VAND. L. REV. 1857 (2000).

A. FIRST NORMS, THEN RIGHTS

I have described a version of private self-correction in my account of "private intellectual property systems" that emulate the functioning of a full-bore, publicly specified property regime.[79] For example, Hollywood writers who submit scripts to movie studios developed a "script registry" under the auspices of the Writer's Guild that acted much like a private "copyright office" for uncopyrightable script ideas. A more recent example is described in a paper by Dotan Oliar and Christopher Sprigman that documents widely understood norms prohibiting "joke stealing" by comedians operating at the higher levels of the standup comedy industry.[80] These norms protect investment in creation of comedy material, despite the absence of formal IP rights.[81] In some cases, norms like these may eventually find their way into formal legal rules. In the meantime, they are good examples of a purely private (i.e., non-governmental) response to a deficiency in the formal specification of property rights. Indeed, from a purely functional standpoint, norms like this constitute a new property rights specification; the distinction between formal and informal makes little difference.[82]

How about the opposite case? Is there any evidence of systemic self-correction when there is "too much" formal, legally-specified IP? The answer is once again yes, though this is a more recent phenomenon and the theory surrounding it is thus necessarily more speculative.

The earliest literature on private action to mitigate excess property entitlements centered on institutions to lower transaction costs.[83] Here the em-

---

79. *See* Robert P. Merges, *Contracting into Liability Rules: Intellectual Property Rights and Collective Rights Organizations*, 84 CALIF. L. REV. 1293, 1361–62 (1996) (describing five examples of the phenomenon); Robert P. Merges, *From Medieval Guilds to Open Source Software: Informal Norms, Appropriability Institutions, and Innovation* (Nov. 13, 2004) (unpublished essay presented at Conference on the Legal History of Intellectual Property, on file with University of Wisconsin Law School) (describing how informal norms of nondisclosure to the guild interacted with the sharing of some information among guild members).

80. Dotan Oliar & Christopher Sprigman, *There's No Free Laugh (Anymore): The Emergence of Intellectual Property Norms and the Transformation of Stand-up Comedy*, 94 VA. L. REV. 1787 (2009).

81. *Id.* at 1802–03 (noting that copyright law protects only "expression" and not "ideas," making it easy to take the gist of a joke or routine without copying the precise way it is expressed).

82. It may be desirable, even so, for emergent norms to be enacted into formal law. This can both cement them into place and make them more widespread and durable. *See, e.g.*, Robert P. Merges, *A New Dynamism in the Public Domain*, 71 U. CHI. L. REV. 183 (2004) (proposing to codify into copyright and patent law a robust waiver or "dedication to the public" mechanism along the lines of the contractual Creative Commons licenses now popular in the online setting).

83. Merges, *Contracting into Liability Rules*, *supra* note 79.

phasis was on the ability of private actors to create institutions that smoothed the way for high volume IP exchange. The point of the theory was to show that property rights sometimes induce investments in transactional mechanisms and, therefore, that explicit policy interventions were not always necessary to lower transaction costs.

Later, the theme of self-correction through private action was made more explicit. Private investments to prevent rivals from obtaining property rights were observed, and it was proposed that these investments could be expected to increase as the value of property rights (and hence the economic leverage in the hands of rivals who hold them) increased.[84]

A more sophisticated approach to self-correction was described in a recent paper by Jonathan Barnett.[85] Barnett is interested in studying private sector responses to existing property regimes.[86] He describes industries in which some, typically large, firms have a steady demand for "outside" inventions. He proposes that these industries can effectively respond to the threat of overly strong property specifications—but only if coordination costs among firms are low. Under these circumstances, firms can develop mutual nonenforcement norms, collective transactional mechanisms, lobbying efforts, and outright dedication of some inventions to the public domain, all as a way to offset the inefficiently strong property rights they are confronted with.[87] But he theorizes that private responses will not be effective where coordination costs among firms are high. In this case, Barnett says firms will find themselves in what he calls a "property trap," where innovators (large and

84. Merges, *supra* note 82 (arguing that private investments to offset competitors' (arguably excessive) property rights help to mitigate the "overpropertization" trend).

85. Jonathan M. Barnett, *Property as Process: How Innovation Markets Select Innovation Regimes* (Univ. of S. Cal. Law & Econ. Working Paper Series, Paper No. 86, 2008), *available at* http://law.bepress.com/usclwps/lewps/art86.

86. *Id.* at 5 (stating that his article's goal is "to identify the conditions under which privately-interested innovator populations will (and will not) have the incentives and capacity to undertake socially-interested actions that avoid or substantially remedy any excessive propertization outcome").

87. *Id.* at 7.

> Building in part on established lessons from the public-choice literature, [Barnett] argues that markets are likely to resist and correct overpropertization—that is, the property trap is likely to be broken—where two conditions are satisfied: (i) adversely-affected innovators tend to enjoy low coordination costs, which is likely to be the case where innovators are few in number (or act through a collective organization) and occupy a dominant market position, and (ii) adversely-affected innovators are neither clearly net users nor clearly net producers of the relevant pool of intellectual goods . . . .

Id.

small firms) defect from a pre-existing "sharing" equilibrium by racing aggressively to acquire more and more property rights.[88]

B. MECHANISMS OF REFORM: POLITICAL ECONOMY CONSIDERATIONS

Scholarship since Harold Demsetz's 1967 article[89] has emphasized the need to augment the bottom-up view of property evolution.[90] A 2005 article by Katrina Wyman captures the basic thrust of the newer literature:

> While directly affected parties must agree to rearrange rights through market transactions, many directly affected parties may not be consulted personally when rights are rearranged through political processes, let alone given a veto over the decision to change. Since the political process does not require unanimity to proceed, it is important, in determining the probability of change, to analyze the expected distribution of the benefits and costs of private property among the influential interest groups who are likely to be consulted.[91]

---

88. *Id.*; Raustiala and Sprigman, in a related vein, show the adaptation of the fashion industry to a low level of IP protection. The authors argue that the fashion industry has settled on a low-IP protection "equilibrium" that permits a form of insurance; when firms miss out on an important fashion trend, they can copy other firms' designs for the mutually tolerated "knockoff market." *See* Kal Raustiala & Christopher Sprigman, *The Piracy Paradox: Innovation and Intellectual Property in Fashion Design*, 92 VA. L. REV. 1687, 1698–1717 (2006).

89. Demsetz, *supra* note 77 (explaining that growth of economic activity concerning economic assets leads to a strengthening of property rights over those assets, and that generally property rights specifications adjust to changing economic conditions).

90. *See, e.g.*, Saul Levmore, *Property's Uneasy Path and Expanding Future*, 70 U. CHI. L. REV. 181, 184–86 (2003) (distinguishing between efficiency (Demsetzian) and interest group theories of property rights); Saul Levmore, *Two Stories About the Evolution of Property Rights*, 31 J. LEGAL STUD. 421, 429–33 (2002) (describing competing economic efficiency and interest group theories of the evolution of property rights).

91. Katrina Miriam Wyman, *From Fur to Fish: Reconsidering the Evolution of Private Property*, 80 N.Y.U. L. REV. 117, 122 (2005). There is an interesting middle ground between what has been called the "naïve theory" (which does not take political economy into account at all) and an explicitly political theory of property right change. This might be described as the "property rights possibility frontier," and it is suggested in some comments about the demand for property rights by economist Lee Alston:

> There seems to be some confusion in the literature over which way causation runs between property rights and value. The confusion is cleared up if we remember the following. It is true that a resource becomes more valuable the greater the rights one has over the resource, and in this sense value (or actual rent) is a function of property rights. But it is not actual rent, but rather potential rent, that drives the demand for property rights. Potential rent is a function of the inherent rental stream (e.g., world price of the resource) and some benchmark set of possible property rights that are culturally and institutionally specific to a time and place.

William Landes and Richard Posner, in their book on the economic structure of intellectual property law, describe the "asymmetry between the private value of intellectual property rights and the private value of the . . . public domain."[92] This asymmetry drives the public choice aspects of their analysis of the demand for IP protection through legislation. They make the common sense, but important, point: The private value of specific IP extensions can be very high, which serves as a strong motivation for firms to lobby heavily for stronger IP protection statutes.[93]

This public choice story is often deployed to explain the "overexpansion" of IP protection during the past fifteen or twenty years.[94] In the article quoted earlier, Wyman goes on to state an important point:

> [R]ecognizing the significance of political decisionmaking rules underscores the need to examine these rules closely in any particular context as variations in them may affect the success of rearranging rights. In particular, the more the collective-choice rules tend toward mandating the unanimity of the affected parties to alter rights in the market, the more difficult it may be to rearrange rights politically.[95]

Although IP legislation does not demand congressional unanimity, voting rules and procedures in this domain follow the general pattern in Congress. This means that it is much easier to veto proposed legislation than to get a particular bill passed. Due in part to the increasing value of intellectual property, and the increasing investment in IP lobbying that has resulted (as public choice theory would of course predict), there are now many more "veto players" in the IP legislation arena than there were, say, twenty years ago.[96] Recent efforts to pass "patent reform" legislation are only the latest evidence of this trend.

In particular, the recent battles over patent reform in Congress show that there is a major divergence between the interests of the biomedical industries

---

Lee Alston, *Toward an Understanding of Property Rights*, *in* EMPIRICAL STUDIES IN INSTITUTIONAL CHANGE 31, 32 (1996).

92. WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 409 (2003).

93. *Id.* at 407–09.

94. This is a common theme in the works of Lawrence Lessig. *See, e.g.*, LAWRENCE LESSIG, CODE VERSION 2.0 (2006). For a similar perspective, see also YOCHAI BENKLER, THE WEALTH OF NETWORKS: HOW SOCIAL PRODUCTION TRANSFORMS MARKETS AND FREEDOM (2006).

95. Wyman, *supra* note 91, at 124.

96. For background on veto players, see generally GEORGE TSEBELIS, VETO PLAYERS: HOW POLITICAL INSTITUTIONS WORK (2002).

(pharmaceuticals, biotechnology, medical devices) and information technology companies (semiconductors, software, and the like).[97]

The upshot is that rent-seeking will have to be curtailed in the courts. As Polk Wagner stated it recently:

> At the same time that the patent system is plainly becoming more economically important, more utilized, more costly, and more complex, the emergence of the technology industry as a major player—and one with divergent interests from the traditional players—seems likely to have a deeply politicizing effect. As the patent law becomes more politicized and the stakes rise, the opportunities for substantial reform of the system narrow. This is in large part because the structure of the U.S. political system is well designed to slow the pace of change of controversial legislation, especially such legislation that has a ratio of economic importance to public visibility. This fact does not, of course, mean that there will be less legislative activity surrounding the patent system; indeed, with higher public visibility, more controversy, and more lobbying dollars likely to be spent, legislative activities, hearings, proposed legislation, and the like should only increase. But these activities, I suggest, will fall short of real, substantive patent reform.[98]

---

97. This split, and its stalling effect on patent reform legislation, is described in a Congressional Research Service study from 2006. WENDY H. SCHACHT, CONG. RESEARCH SERV., CRS REPORT NO. RL33367, PATENT REFORM: ISSUES IN THE BIOMEDICAL AND SOFTWARE INDUSTRIES (2006). For background on the formation of patent reform lobbying groups centered in rival industries, see generally Candace Lombardi, *Tech Firms to Lobby for Patent Litigation Reform*, ZDNET NEWS, May 11, 2006, http://news.zdnet.com/2100-9595_22-148032.html; *New Coalition Seeks to Protect American Innovation*, IP FRONTLINE, Mar. 23, 2007, http://www.ipfrontline.com/depts/article.asp?id=14571&deptid=8 (illustrating the formation of the 21st Century Coalition for Patent Reform, an organization of pharmaceutical companies, some universities, and companies from other industries).

98. R. Polk Wagner, *The Supreme Court and the Future of Patent Reform*, 55 FED. LAW. 35, 35 (2008). The real action, according to Wagner, will be in the courts and even inside the PTO:

> [Major trends today include] a growth in patent-related activity, and the emergence of the technology industry (on the West Coast) as a major player in the political economy of the patent system. It is these "plate tectonics," . . . that both explain the recent interest in the patent system as well as suggest important features of its future. . . . [A]s the paths for change narrow, meaningful patent reform will increasingly fall to the courts. This case-by-case, litigation-driven change has, . . . important consequences. . . . This, in turn, suggests that a re-evaluation of patent reform options is required, and that, in particular, the understudied role of the U.S. Patent and Trademark Office (PTO) should be revisited.

Id. at 35.

Real change—real reforms to rein in rent-seeking—will have to come from the courts.

1. Policing the eBay *Line*

What the Court recognized in *eBay* was that it must police the line between rent-seeking and innovation.[99] This opinion recognized the important threat that non-innovating patent owners posed to the health of the innovation system. And it announced a considered approach to maintaining the overall viability of the macro-environment for innovation.

As I have been arguing, the fault line between innovation and rent-seeking defines a major policy issue in the IP field. In my view, the Court in its eBay opinion tried to establish some basic parameters for drawing this line. The purpose of the line is to separate socially productive innovation from socially wasteful rent-seeking. This is easy enough to see at the conceptual level; the difficulties all come when we try to apply this principle in individual cases. I will discuss here just two examples of these difficulties, though many more are sure to arise, starting with troll-related activities and then turning to university research.

Rooting out pure rent-seeking might seem easy, but it is not always so. It is tempting to simply target specific companies or entities—law firms (meaning contingency fee patent firms) that acquire patents and then assert them against numerous defendants, for example;[100] or perhaps large-scale "patent aggregators" that acquire many patents and then sell "litigation insurance" to many companies, in exchange for a promise not to assert those patents against companies willing to pay the "premium."[101] It may be relevant that a specific company is a repeat offender in the rent-seeking game. But typically, it is not specific entities but rather specific tactics or practices that are most relevant. Intellectual Ventures, for example, has engaged in an effort to finance forward-looking "pure" R&D; patents arising from this sort of effort may wind up being a far cry from the acquisition of a patent in bankruptcy, or a patent bought on the cheap and later asserted against numerous defen-

99. *See supra* Part I.

100. *Cf.* Raymond P. Niro, *Who Is Really Undermining the Patent System—"Patent Trolls" or Congress?*, 6 J. MARSHALL REV. INTELL. PROP. L. 185 (2007) (defending patent acquisition and assertion from one often accused of being a troll). Niro cites to and argues against an opposing article, Brenda Sandburg, *You May Not Have a Choice. Trolling for Dollars*, THE RECORDER, July 30, 2001, at 1 (describing of patent troll tactics). *See* Niro, *supra*, at 186.

101. The most prominent is Intellectual Ventures, Inc. *See generally* Nicholas Varchaver, *Who's afraid of Nathan Myhrvold?*, FORTUNE, July 10, 2006, at 110.

dants.[102] Trolling, to put it simply, is a matter of behavior rather than status. One can act as a troll, but it will usually not be true that one simply is a troll. The "troll line," in other words, must be policed case-by-case and fact-by-fact.

Now let us consider university research. Mark Lemley recently wrote an article whose title is self-explanatory: "Are Universities Patent Trolls?"[103] Lemley notes the growth of patenting by universities, which held sixteen times as many patents in 2004 as in 1980, and the concomitant sprouting of university technology transfer offices (which are 100 times more numerous now than in 1980).[104] This is not of course bad in itself; federal policy has been aimed at just this result since the Bayh-Dole Act of 1980.[105] But what is troubling is that the universities are increasingly seeking to maximize not technology transfer per se, but short-term licensing revenues. This has led to what Lemley describes as "a growing frustration on the part of industry with the role of universities as patent owners. Time and again, when [talking] to people in a variety of industries, their view is that universities are the new patent trolls."[106] As he points out, this is not a good development. Lemley offers a variety of policy recommendations to offset it, but the stark fact remains: universities,[107] at least some of them, have crossed the line between innovators and rent-seekers. This is not good for society, and ultimately, not good for the universities themselves.

Even so, an overreaction might be just as bad as no reaction at all. That's because universities continue to generate important, horizon-stretching tech-

---

102. Intellectual Ventures, Who We Are, http://www.intellectualventures.com/about.aspx (last visited Nov. 20, 2009).

103. Mark A. Lemley, *Are Universities Patent Trolls?*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 611 (2008).

104. *Id.* at 614.

105. Bayh-Dole Act, Pub. L. No. 96-517, 94 Stat. 3019 (codified as amended at 35 U.S.C. §§ 200–12 (1980)). *See generally* James D. Clements, *Improving Bayh-Dole: A Case for Inventor Ownership of Federally-Sponsored Research Patents*, 49 IDEA 469 (2009) (arguing that the Bayh-Dole Act disincentivizes university patenting).

106. Lemley, *supra* note 103 at 615.

107. More accurately, technology transfer offices within universities. There is a growing "agency problem" in this area; university scientists, and the law professors who study IP law, usually counsel restraint and a long-term orientation as the focal points of university licensing policy. But technology transfer offices are profit centers, and they are evaluated on the basis of net short-term financial contributions to the university. So it is not surprising to see a congressional patent reform hearing on legislation to curtail rent-seeking where a policy expert from a university argues in favor of the measure, but a technology transfer officer from the same university argues against it.

nologies.[108] Clearly the right response is not to unilaterally curtail university licensing. It is instead to redraw the fault line, to more effectively rule out rent-seeking and thus more thoroughly encourage the real innovation we are after.

2. A Case Study: Policing the Troll Line through Damages Doctrine

In the end what matters most is that property rights be appropriately monitored and maintained. Like a traditional stone wall demarcating a physical boundary, property rights must be patrolled and policed. Where there are signs of decrepitude, some agent must step in to fix the fallen structure, to replace the fallen rocks. Otherwise the property line loses its meaning and ceases to perform its correct function. This in turn creates a threat to the integrity of the boundary.

We have seen that this process is already underway when it comes to patent institutions. *EBay* is the best current example. As I have explained, however, the patent troll phenomenon is robust and adaptable. More action is needed to shut down the avenues of rent-seeking activity. A current, pressing example is the problem of damages in patent cases.

The problem here is driven by the same logic noted by the Supreme Court in *eBay*. Under current damages rules, patents over small components can often be effectively leveraged into disproportionately large monetary awards—creating rents that are then sought out by patent trolls. Congressional testimony over proposed reforms in this area summarized the reasons why this is possible:

> Unfortunately, current law does not do a good job of ensuring that a patentee receives a royalty in proportion to the true role of the patented invention. As an example, in many cases damages' experts will rely on the traditional principle that, as a "rule of thumb," licensors should receive a quarter to a third of the profit made on a product. However, if there are five patents relevant to a complex

108. To take one example among many: Harvard University recently licensed a series of patents on "black silicon" technology, which is a technique for transforming silicon into a much more effective light sensor and power generator. Silicon treated using the Harvard process becomes much more receptive to photons (i.e. light). So transformed silicon has potential applications in medical imaging (where light is absorbed to make an image), digital cameras, and solar power (where silicon-based photovoltaic cells are used to absorb light and transmit electrons to generate electricity). Dylan McGrath, *Harvard Spinout Licenses 'Black Silicon' Patents*, EE TIMES, Oct. 13, 2008, *available at* http://www.eetimes.com/showArticle.jhtml?articleID=211200183.

The licensee in this case, a small company called SiOnyx, is in the process of developing the technology for a number of applications. *Id.* This active participation in research and development is what sets this company apart from a patent troll.

> product, much less thousands, all the profit and then some would go to patent licensors applying this "rule of thumb." The party that actually created and sold the product would be forced to lose money on its products sales, under this common royalty analysis. Yet, this type of testimony is often permitted because of years of authority and longstanding licensing practices from a bygone era.
>
> Another factor is that the legal form of patent claims can be manipulated to inflate damage demands and awards. A patentee can draft a patent claim to cover a large and expensive product even where the invention relates only to a minor and inexpensive component. For example, if one were to invent a new type of windshield wiper, patent law permits the patent to be granted on a standard car with the improved windshield wiper. Under common interpretations of patent law, the royalty percentage is then based on the price of the entire car, not just the improved windshield wipers. This, not surprisingly, inflates unduly the plaintiffs' demands.
>
> Put simply, in the real world, a host of factors impede attempts to put a patent in context so one can effectively explain to a jury this concept of proportionality. For example, judges often do not want a trial to involve what other patents may cover a product beyond those that are allegedly infringed because it is complex enough for the jury to determine whether the asserted patent or patents cover the product. In addition, a juror is subjected to so much focus on the asserted patent and the accused feature in the trial process that efforts to put into perspective the limited role of the patented technology are difficult.[109]

The solution here, stated broadly, parallels the new injunction rule announced in *eBay*: shut down the opportunities for rent-seeking.[110] What that means practically is that we need a simple test for damages in patent cases that measures a patentee's compensation strictly with reference to the actual economic value of the patented invention relative to the overall product produced and sold by the defendant. The test should inquire into the difference between the actual profit to the infringer, made with the patented invention incorporated into the infringer's product, and what the infringer's profit would have been if its product had instead included the next best (unpatented) alternative technology.[111] This would conform the damages test with

---

109. *Patent Trolls: Fact or Fiction?: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary*, 109th Cong. 10 (2006) (statement of Edward R. Reines, Patent Litigation Partner, Weil, Gotshal & Manges LLP).

110. *See supra* Part I.

111. This test follows closely one announced by Judge Frank Easterbrook, sitting as a trial judge in a patent case. Grain Processing Corp. v. Am. Maize-Prods. Co., 979 F. Supp. 1233 (N.D. Ind. 1997) (Easterbrook, J., sitting by designation), *aff'd*, 185 F.3d 1341 (Fed. Cir.

general compensation principles in patent law, and simultaneously reduce the opportunities for rent-seeking via excessive damage awards.[112]

IV. CONCLUSION

Patent trolls threaten the integrity of the innovation system in the U.S. today. We must not be blinded to the threat by the rote invocation of market-oriented mantras. All the evidence points to a major incidence of rent-seeking, mixed in with the emergence of a perhaps valuable market for independent ideas and inventions. If we are to preserve the traditional justification of patents as an important part of our innovation system, and if we are to uphold the social value of real innovation versus legal gamesmanship and "paper rent-seeking," there is only one course to take: We must act to delineate troll activity more precisely, and when it is present to shut it down, for now, primarily through the courts; in the future, through whatever means present themselves. By carefully distinguishing artificial rents from true innovation, and shutting off or reducing rents when we find them, we can put the trolls out of business while preserving and perhaps nurturing a valuable market for patented innovations. The idea is simple: to make sure patent law is serving its intended purpose, by encouraging real, socially-useful innovations.

1999). See the write-up of these issues in JOHN W. SCHLICHER, PATENT LAW: LEGAL AND ECONOMIC PRINCIPLES § 13:138 (2d ed. 2008):

> This test [for deciding whether the entire market value of the product is attributable to the patented invention] would properly measure the value of an invention only if it asks, "What are the profits available to the infringer from selling a product with the patented feature or component, and what would be the profits from selling a product with the next-best noninfringing substitute feature or component." The difference measures the value of the invention and may be the entire profits or only part of them. In determining lost profits, the courts have recognized that the value of a particular invention is this difference, as the court of appeals made clear in Grain Processing.

112. The Federal Circuit recently took a step in this direction. *See* Lucent, Inc. v. Gateway, Inc., 580 F.3d 1301 (Fed. Cir. 2009) (vacating jury's damage award; remanding for reconsideration on the basis of more realistic evidence of royalty rates from truly comparable licensing agreements).

REFORMING THE REPRODUCTION RIGHT: THE CASE FOR PERSONAL USE COPIES

Ashley M. Pavel[†]

ABSTRACT

The new realities of the digital age have rendered the 1976 Copyright Act inadequate for protecting reasonable personal copying and have created incentives for copyright holders to implement objectionable strategies to protect their rights.

The Note explains that the only current shield to litigation for consumers is the fair use defense, which is inadequate due to the difficulty in proving that a personal copy is transformative. High costs of litigating, coupled with potentially ruinous penalties for losing, leaves little incentive for consumers not to settle even when the personal copy is clearly a fair use.

The Note then explains that the Copyright Act also fails to protect copyright holders due to its focus on "copying" as the proxy for infringement. This is ineffective to prevent filesharing as it is hard to prove that "copying" has occurred, and it forces the holder to invade consumers' privacy by using programs that track their activities. This also incentivizes holders to litigate out of existence developing technologies that aid consumers in making personal copies in direct contravention of the constitutional purpose of copyright.

The Note concludes that to better protect the rights of copyright holders in the digital age, legislation should be enacted that changes the proxy for infringement from "copying," to communicating works to the public, and that grants the copyright holder the exclusive right to authorize such communication. Furthermore, legislation that demarcates private use as non-

† Associate, O'Melveny & Myers LLP. I would like to thank Pamela Samuelson and the participants of the 2008 copyright reform seminar at the University of California, Berkeley, School of Law (Boalt Hall) for their comments on an earlier draft of this note. The views expressed in this note are those of the author, and do not reflect those of O'Melveny & Myers LLP.

infringement will ensure that private use copies for productive use and sharing between friends and family is protected.

TABLE OF CONTENTS

I. INTRODUCTION

Many of the greatest innovations in consumer electronics, from the personal computer to TiVo to the iPod, derive much of their marketability from the presumption that personal use copies do not violate the Copyright Act of 1976. These technologies implicate varying levels of intent to copy

without permission from the copyright owner. At one extreme is the user whose enjoyment of a piece of copyrighted software necessitates the creation of an iterative copy in the computer's Random Access Memory (RAM)—a process of which many computer users are completely unaware. At the other end is the so-called "pirate" who consciously stocks his iPod with hundreds of copyrighted songs he illegally downloaded. Common sense holds that the illegal download is copyright infringement but the RAM copy is not.

Few today would argue with this proposition, although it is far from clear where this consensus finds support in the 1976 Act. This shortcoming is particularly evident when a technology is new, such as when a federal court found RAM copies to be copyright infringement in 1993,[1] or when content industries are threatened by a technological advancement that facilitates or improves personal use copying, such as the Videocassette Recorder (VCR) or Digital Audio Tape.[2]

Most of this difficulty arises because the fair use doctrine[3] is ill-suited to evaluate non-transformative personal use copies. Consequently, judicial application of the fair use factors to personal use technologies is difficult for innovators and users to predict ex ante. An Office of Technology Assessment study found that one's perceived familiarity with copyright law did not correlate with copying habits.[4] This result is unsurprising given the unpredictability of a judicial fair use determination and the Byzantine array of statutory sui generis regulations for specific technologies and uses within the Copyright Act.

Despite this legal uncertainty, public opinion is clear. The same Office of Technology Assessment survey found that most members of the public, whether they engage in the practice or not, believe that personal use copying is acceptable as long as the copies are not sold.[5] With the proliferation of peer-to-peer (P2P) file sharing and illegal BitTorrent downloading sites such as the Pirate Bay, the public may now consider both sale and widespread unauthorized distribution to be unacceptable. The core belief, however, that

1. *See* MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993).

2. *See* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984); Audio Home Recording Act of 1992, Pub. L. No. 102-563, 106 Stat. 4237 (codified at 17 U.S.C. §§ 1001–1010 (2006)).

3. *See* 17 U.S.C. § 107 (2006).

4. OFFICE OF TECHNOLOGY ASSESSMENT, COPYRIGHT AND HOME COPYING: TECHNOLOGY CHALLENGES THE LAW 163 (1989).

5. *Id.* at 3.

strictly private personal uses of a purchased copy are "none of the copyright owner's business" still exists.[6]

To remedy the 1976 Act's uncertainty and disconnect from popular norms, this Note proposes that Congress adopt a general personal use exemption to infringement liability for the courts to interpret through the standard common law process. This Note further proposes that, because the copy itself is not the locus of injury in the digital era, Congress should offer the copyright industry a distribution right more suited to digital technology in exchange for its acceptance of the user's right to make personal copies.

Part II presents an overview and critique of the current state of the reproduction right. Part III argues that the 1976 Act regime does not adequately accommodate personal use copies. Finally, Part IV proposes a reform to the Copyright Act that would clearly allow personal use copies and include a more effective distribution right.

II. CURRENT STATE OF THE REPRODUCTION RIGHT

The 1976 Act was enacted pursuant to Article I, Section 8, Clause 8 of the U.S. Constitution, which delegates to Congress the power to grant authors the exclusive right to their "Writings" for "limited times" in order "to promote the Progress of Science and useful Arts."[7] Copyright fulfills this constitutional purpose by motivating authors' creative efforts and ultimately enhancing public access to creative works. This congressionally-granted monopoly is not exclusively or even primarily intended to foster a private commercial benefit to individual authors or copyright industries.[8] Rather, copyright is an incentive given to authors as a means of enhancing public access to creative works and promoting progress in the arts and sciences.[9] The Supreme Court summarized copyright's purpose:

> The sole interest of the United States and the primary object in conferring the [copyright] monopoly lie in the general benefits derived by the public from the labors of the authors. It is said that

6. COMMITTEE ON INTELLECTUAL PROPERTY RIGHTS IN THE EMERGING INFORMATION INFRASTRUCTURE, NATIONAL RESEARCH COUNCIL: THE DIGITAL DILEMMA: INTELLECTUAL PROPERTY IN THE INFORMATION AGE 129, 134 (2000) [hereinafter THE DIGITAL DILEMMA] ("Many members of the general public appear to believe that all or virtually all private, noncommercial copying of copyrighted works is lawful.").

7. U.S. CONST. art. I, § 8, cl. 8. Note, what constitutes "limited" is left to Congress to decide. Eldred v. Ashcroft, 537 U.S. 186, 187 (2002).

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

9. *Id.*

> reward to the author or artist serves to induce release to the public of the products of his creative genius.[10]

Thus, when examining the aggressive enforcement practices of the content industries, it is important to remember that the goal of a copyright regime is to facilitate a rich and expansive creative commons, not to protect or benefit any particular commercial interest.[11]

This Section presents an overview of the current status of the reproduction right. Section II.A presents a basic explanation of the reproduction right granted by the 1976 Act. Section II.B explains the fair use exemption.

A. REPRODUCTION RIGHT UNDER THE 1976 COPYRIGHT ACT

Under the 1976 Act, the copyright holder has the exclusive right "to reproduce the copyrighted work in copies or phonorecords."[12] The 1976 Act defines "copies" as

> material objects . . . in which a work is fixed by any method now known or later developed, and from which the work can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device. The term "copies" includes the material object, other than a phonorecord, in which the work is first fixed.[13]

A work is "fixed" when "its embodiment in a copy . . . is sufficiently permanent or stable to permit it to be perceived, reproduced, or otherwise communicated for a period of more than transitory duration."[14]

Within this statutory framework, the act of copying is "essential to, and constitutes the very essence of all copyright infringement."[15] For a copy to be infringing, it must be (1) embodied in a material object, or "tangible"; (2) "fixed" such that it may be perceived for more than a "transitory duration"; and (3) "intelligible"—meaning that it must be perceivable directly or with the aid of a machine.[16] Notably, in this system, where reproduction is

---

10. *Id.* at 429 (quoting United States v. Paramount Pictures, Inc., 334 U.S. 131, 158 (1948)) (internal quotation marks omitted).

11. *See id.* at 427.

12. 17 U.S.C. § 106(1) (2006).

13. 17 U.S.C. § 101 (2006).

14. *Id.*

15. 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 8.02[A] (2009).

16. 17 U.S.C. §§ 101, 106(1) (2006); 2 NIMMER & NIMMER, *supra* note 15 § 8.02[B][1]-[B][3] (2009) (categorizing and articulating the three requirements for violation of the reproduction right embodied in Title 17).

the crux of infringement, distribution need not take place to give rise to a suit for infringement.[17]

Consequently, personal use copies, even if never shared or even consciously made, technically constitute copyright infringement.[18] These uses, regardless of the normative consensus as to their legitimacy, or the social utility they generate, are only excused if they can pass a fairly strict fair use test.[19]

B. THE FAIR USE EXCEPTION

Notwithstanding the expansive rights enumerated in the Copyright Act, § 107 exempts from infringement liability certain uses that Congress has deemed socially valuable.[20] The fair use doctrine is a safety valve that allows courts to avoid rigid application of the 1976 Act when doing so would "stifle the very creativity which that law is designed to foster."[21]

Section 107 states that copies made "for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research" are fair uses.[22] This list serves as a non-exhaustive guideline.[23] The factors courts consider to determine whether a particular case is a fair use are:

1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
2) the nature of the copyrighted work;

17. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv., Co., 499 U.S. 340, 361 (1991) ("To establish infringement, two elements must be proven: (1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.").

18. *See* John Tehranian, *Infringement Nation: Copyright Reform and the Law/Norm Gap*, 2007 UTAH L. REV. 537, 543–48 (2007) (analyzing the breadth of commonplace actions that infringe copyright).

19. 17 U.S.C. § 106(1) (2006) (providing an exclusive right of reproduction); 17 U.S.C. § 107 (2006) (stating the fair use test); *see also* Fred von Lohmann, *Fair Use as Innovation Policy*, 23 BERKELEY TECH. L.J. 829, 830 (2008). *But see* 17 U.S.C. § 108 (2006) (exempting libraries); 17 U.S.C. § 110 (2006) (exempting certain performances); 17 U.S.C. § 111 (2006) (exempting secondary transmissions).

20. 17 U.S.C. § 107 (2006).

21. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 577 (1994); *see also* Meeropol v. Nizer, 560 F.2d 1061, 1068 (2d Cir. 1977) ("The doctrine offers a means of balancing the exclusive right of a copyright holder with the public's interest in dissemination of information . . . .").

22. 17 U.S.C. § 107.

23. *See id.* ("In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall *include* . . . .") (emphasis added).

> 3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> 4) the effect of the use upon the potential market for or value of the copyrighted work.[24]

Courts do not consider these factors in isolation, but weigh all of the factors together to determine, on balance, if a particular act constitutes fair use.[25] Furthermore, the four factors allow courts to conduct flexible case-by-case analysis, rather than adhere to a bright line rule.[26]

The fair use analysis is reasonably well adapted to handle "transformative" uses such as parody, even where a copyrighted work is appropriated for profit.[27] The fair use doctrine, however, is less suited to non-transformative uses with little to no commercial impact.[28]

Although noncommercial uses are presumptively fair,[29] it is far from certain that "consumptive" private use copying can survive the four-factor fair use analysis.[30] The ongoing debate as to whether fair use is a defense to copyright infringement or an affirmative user's right[31] is indicative of this problem.

III. THE PROBLEM WITH THE 1976 ACT

As it stands today, the reproduction right under the 1976 Act also fails both copyright consumers and copyright producers. Personal use copies fall through the cracks of the 1976 Act regime in three principal ways. The 1976 Act fails consumers because a fair use defense is an almost pathetic shield against even unjustified copyright infringement claims. The 1976 Act fails copyright producers because, where personal use copies are involved, the

---

24. *Id.*

25. *Campbell*, 510 U.S. at 578.

26. *Id.* at 577.

27. *See, e.g.*, *id.* at 594 (finding rap group's parody to be fair use despite its commercial nature and substantial appropriation of copyright holder's song). *But see* Dr. Seuss Enters. v. Penguin Books USA, Inc., 109 F.3d 1394, 1400 (9th Cir. 1997) (finding that, unlike parody, satire has a "diminished" claim to fairness in borrowing from a copyrighted work).

28. *See* 17 U.S.C. § 107 (2006).

29. Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 448 (1984) (finding that recording copyrighted video content for "time shifting" purposes was noncommercial and presumptively fair).

30. THE DIGITAL DILEMMA, *supra* note 6, at 134.

31. *See* THE DIGITAL DILEMMA, *supra* note 6, at 134 (discussing the debate). *Compare Campbell*, 510 U.S. at 590 ("[F]air use is an affirmative defense . . . .") *with* Bateman v. Mnemonics, Inc., 79 F.3d 1532, 1542 n.22 (11th Cir. 1996) ("Although the traditional approach is to view 'fair use' as an affirmative defense, this writer, speaking only for himself, is of the opinion that it is better viewed as a right granted by the Copyright Act of 1976.").

Act's focus on copying as the proxy for injury leads to ineffective and expensive litigation strategies. Finally, the current regime stifles innovation because it allows and encourages content owners to sue out of existence small ventures that are developing new technologies. This, in turn, harms consumers because it limits competition in new consumption technologies to the few large technology companies who have leverage with the content industries and who can afford to litigate expensive fair use claims. The following Sections address these problems by examining the evolution of the courts' treatment of personal use copies under the fair use doctrine and by arguing that fair use is not an adequate framework to evaluate consumptive copying for private use.

A. FAIR USE: THE NOT-SO-SAFE HARBOR

Mounting a fair use defense is always a risky bet for a copyright defendant. A fair use defense is expensive,[32] unpredictable,[33] and subject to the economic savvy of the implicated content industries.[34] All of these problems are compounded where personal use copies are involved because defendants are typically individuals with modest means and their copying is not transformative. Moreover, with the advent of new technologies, commoditization of personal use copies may now be possible, tipping the fourth fair use factor, market harm, away from the private use defendant for the first time.

1. The Procedural Mechanics of Fair Use Encourage Meritless Litigation

The fact-specific nature of the fair use inquiry coupled with the uncertainty of its outcome makes fair use an arduous and cost-intensive defense.[35] Adding to the cost is the fact that the defendant bears the burden of proving that a fair use "limitation" on the exclusive rights of the copyright owner applies to the particular circumstances.[36]

The unpredictability of a judicial fair use determination[37] is aptly illustrated by the fact that every fair use case to reach the Supreme Court was

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

33. *See infra* Section III.A.2 (examining this unpredictability through the narrow lens of the transformative test).

34. *See infra* Section III.A.3.

35. *See, e.g.*, David Nimmer, *A Modest Proposal to Streamline Fair Use Determinations*, 24 CARDOZO ARTS & ENT. L.J. 11, 16 (2006) ("[N]obody can know what fair use is until the full process of litigation has run its course.").

36. *See Campbell*, 510 U.S. at 590 ("fair use is an affirmative defense"); 17 U.S.C. § 107 (2006).

37. For a more detailed examination of the unpredictability of the fair use doctrine, *see* Section III.A.2.

overturned at each level of review.[38] Section 107's failure to produce clear and consistent results has led to outright contempt from preeminent copyright scholars. For example, David Nimmer wrote that "[b]asically, had Congress legislated a dartboard rather than the particular four fair use factors embodied in the Copyright Act, it appears that the upshot would be the same."[39]

Furthermore, the disproportionate statutory damage awards authorized by the Copyright Act make the costs of a fair use loss enormous, particularly for an individual defendant. A copyright holder can forego proving actual damages and elect to collect statutory damages.[40] The damages range from $750 to $30,000 per work infringed, "as the court considers just."[41] Furthermore, if a court finds willful infringement, it can increase damages up to $150,000 per work.[42] If a court finds that a defendant was unaware that his acts were infringing, the court may only reduce statutory damages to $200 per work.[43] These statutory damages awards are particularly egregious in light of recent file sharing litigation. Under this framework, a file sharing defendant would have to pay, at minimum, $200 per song or TV show obtained from a P2P server, whereas the content industry's actual damage is limited to their share of the profits of the approximately $0.69 to $1.29 the defendant would have paid for a digital copy from a vendor like iTunes.[44]

When facing an uncertain defense strategy and a financially ruinous penalty for losing, most copyright defendants have little incentive to invoke the fair use doctrine, even if, in the abstract, it seems as if the defendant's conduct was clearly fair use.[45] Overly-aggressive copyright litigation has run rampant because individuals frequently lack the incentive or the means to

38. *See* Nimmer, *supra* note 35, at 16; *Campbell*, 510 U.S. 569 (unanimous finding of fair use); Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539 (1985) (split opinion); Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984) (split opinion).

39. David Nimmer, *"Fairest of them All" and other Fairy Tales of Fair Use*, 66 LAW & CONTEMP. PROBS. 263, 280 (2003).

40. 17 U.S.C. § 504(c)(1) (2006).

41. *Id.*

42. 17 U.S.C. § 504(c)(2) (2006).

43. *Id.*

44. *See* 17 U.S.C. § 504(c)(2) (2006) ("the court in its discretion may reduce the award of statutory damages to a sum of not less than $200"); Greg Sandoval, *Will Consumers Determine iTunes Prices?*, CNET NEWS.COM, Apr. 7, 2009, http://news.cnet.com/will-consumers-determine-itunes-prices/.

45. *See* Michael W. Carroll, *Fixing Fair Use*, 85 N.C. L. REV. 1087, 1106 (2007) ("[M]ost defendants lack incentives to defend novel fair use interpretations.").

defend themselves rather than settle.[46] One particularly salient example is a recording industry representative's suggestion that students drop out of college to pay their copyright infringement settlements.[47]

Although P2P file sharers are not the most sympathetic defendants, this skewed incentive structure also emboldens the copyright holders to use copyright as a sword to attempt to sue scathing critics into silence[48] and what they perceive to be threatening technological innovation[49] out of existence.[50] Content industries have used the courts to try to eradicate such technologies as the DVR,[51] digital music players,[52] and P2P software.[53] Although content industries are threatened by these technologies, they promote progress in the arts and sciences by making copyrighted content more accessible and useful to the consumer and thus growing demand for the works.[54] Far from being a safety valve for the freedom of expression, the sheer magnitude of the fair use doctrine's flaws allows content producers to use their copyrights to deliberately circumvent copyright's constitutional purpose. Unfortunately, it appears that "fair use in America simply means the right to hire a lawyer."[55]

2. *The Fair Use Focus on Transformativeness Unreasonably Disfavors Personal Use Copying*

In addition to its unpredictability, the fair use analysis's increasing focus on whether a particular work is "transformative" disadvantages personal use

46. *See, e.g.*, ELECTRONIC FRONTIER FOUNDATION, RIAA V. THE PEOPLE: FOUR YEARS LATER (2007), http://w2.eff.org/IP/P2P/riaa_at_four.pdf [hereinafter FOUR YEARS LATER].

47. Cassi Hunt, *Run Over by the RIAA: Don't Tap the Glass*, THE TECH, Apr. 4, 2006, at 9, *available at* http://tech.mit.edu/V126/N15/RIAA1506.html ("[An RIAA representative] even had the audacity to say, 'In fact, the RIAA has been known to suggest that students drop out of college or go to community college in order to be able to afford settlements.' ").

48. *See, e.g.*, Doe v. Gellar, 533 F. Supp. 2d 996 (N.D. Cal. 2008) (using a DMCA claim to try to remove a YouTube video debunking Uri Geller's "psychic" abilities); Savage v. Council on American-Islamic Relations, 2008 U.S. Dist. LEXIS 60545 (N.D. Cal. 2008) (using a copyright infringement claim to attempt to stop the defendant from reposting Savage's remarks concerning the defendant on its website along with its response).

49. *See, e.g.*, FOUR YEARS LATER, *supra* note 46.

50. *See, e.g.*, *id.*; Benny Evangelista, *Reining in Tech; Learning from the Napster Case, the Entertainment Industry is Trying to Block New Technology Before it Takes Off*, S.F. CHRON., Aug. 30, 2004, at C1; see also infra Section III.A.2.b) (discussing how studios sued RePlayTV out of existence).

51. Paramount Pictures Corp. v. RePlayTV, 298 F. Supp. 2d 921 (C.D. Cal. 2004).

52. Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., Inc., 180 F.3d 1072 (9th Cir. 1999).

53. MGM Studios, Inc. v. Grokster, Ltd., 545 U.S. 913 (2005).

54. *See* U.S. CONST. art. I, § 8, cl. 8.

55. LAWRENCE LESSIG, FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY 187 (2004).

copying.[56] The effect of this is that personal use copies, which should usually be fair uses because of their similarity to the exceptions granted to other instances of iterative copying that also have a negligible market impact,[57] are often held to be acts of copyright infringement.[58] This Section documents courts' focus on "transformativeness" as nearly determinative of the fair use question, and then demonstrates how this approach causes difficulty for consumptive personal uses.

a) The Fair Use Test's Focus on "Transformativeness"

A finding that a particular use of a copyrighted work is "transformative," although not absolutely necessary, substantially shifts the analysis in favor of a finding of fair use.[59] For example, in *Campbell v. Acuff-Rose Music*, the Court found that 2 Live Crew's parody rap song of Roy Orbison's "Oh, Pretty Woman" was "transformative" of the original and therefore a fair use, despite the parody's commerciality and potential harm to the market for Orbison's original work.[60] The Court emphasized that transformativeness is "at the heart" of the fair use analysis,[61] and that the permissibility of iterative copies for classroom use was a "statutory exemption" to this rule.[62]

The Court's heavy-handed emphasis on transformativeness as nearly determinative of the fair use question helps to explain some lower courts' almost nonsensical fair use determinations.[63] In what is perhaps the seminal example of how fair use can be a grossly inept safety valve for new technologies, the Ninth Circuit ruled that unauthorized RAM copies of a

56. *See, e.g.*, Rebecca Tushnet, *Copy this Essay: How Fair Use Doctrine Harms Free Speech and How Copying Serves It*, 114 YALE L.J. 535, 555–56 (2004).

57. *See, e.g.*, 17 U.S.C. § 107 (2006) (exempting copies made for teaching, scholarship, or research); 17 U.S.C. § 108 (2006) (exempting libraries).

58. *See, e.g.*, Am. Geophysical Union v. Texaco Inc., 60 F.3d 913 (2d Cir. 1994) (finding that internal research copies of copyrighted journals were not fair use because whole articles were copied as part of a commercial enterprise).

59. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994). The Court stated:

> Although . . . transformative use is not absolutely necessary for a finding of fair use, the goal of copyright, to promote science and the arts, is generally furthered by the creation of transformative works. Such works thus lie at the heart of the fair use doctrine's guarantee of breathing space within the confines of copyright, and the more transformative the new work, the less will be the significance of other factors . . . that may weigh against a finding of fair use.

Id. (internal citations omitted).

60. *Id.* at 569.

61. *Id.* at 579.

62. *Id.* at 579 n.11.

63. *See* Tushnet, *supra* note 56, at 555.

software program constituted copyright infringement.[64] This decision has since been superseded by statute.[65] Nonetheless, video game giant Blizzard Entertainment prevailed in a copyright infringement suit against the makers of a "bot," an add-on program that allowed a user's computer to automatically progress through Blizzard's video game.[66] Because automating game-play violates Blizzard's End User License Agreement and Terms of Service, Blizzard argued that RAM copies made by users of the program were unauthorized and therefore copyright infringement.[67]

In this climate where powerful and litigious copyright interests aggressively pursue personal use copies, courts find it increasingly necessary to squeeze even ill-fitting consumptive uses into the "transformative" category in order to declare the use fair.[68] For example, in *Perfect 10, Inc. v. Amazon.com, Inc.* the Ninth Circuit held that thumbnail previews of the plaintiff's copyrighted photos were "transformative" just because the exact copies were smaller and of a lower resolution to support its fair use determination.[69] The court went on to state that "a search engine may be more transformative than a parody because a search engine provides an entirely new use for the original work, while a parody typically has the same entertainment purpose as the original work."[70]

This "new use" conception is a stretch from the transformativeness standard that the Supreme Court set out in *Campbell*, when it held that transformative works do not "merely supersede the objects of the original," but instead "add[] something new, with a further purpose or different

64. MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511, 518 (9th Cir. 1993) (finding that RAM copies, made by a third party while using a licensed, copyrighted software program, were outside the scope of the license, and therefore constituted infringement).

65. Computer Maintenance Competition Assurance Act of 1997, Pub. L. No. 105-304, § 301, 112 Stat. 2860 (1998) (codified at 17 U.S.C. § 117 (2006)) (ensuring that independent computer maintenance service providers could conduct business without being hampered by need to license software on customer machines or risk infringement liability).

66. MDY Indus., LLC v. Blizzard Entm't, Inc., 2009 U.S. Dist. Lexis 24151 (D. Ariz. Mar. 10, 2009) (granting permanent injunction for tortuous interference with contract, contributory and vicarious copyright infringement, and violation of the DMCA, but staying the copyright and DMCA injunction pending appeal); Ben Kuchera, *Blizzard Attempt to Kill WoW Bot Bad News for Copyright Law*, ARS TECHNICA, May 7, 2008, http://arstechnica.com/news.ars/post/20080507-blizzard-attempt-to-kill-wow-bot-bad-news-for-copyright-law.html.

67. *Id.*

68. *See* Tushnet, *supra* note 56, at 556 ("Nonetheless, courts apparently believe that a finding of transformation is necessary for fair use, and they therefore strain to find transformation where they conclude that a defendant ought to prevail.").

69. 508 F.3d 1146, 1165 (9th Cir. 2007).

70. *Id.*

character, altering the first with new expression, meaning, or message."[71] It seems entirely plausible that a large sector of Perfect 10's market, particularly the market for Perfect 10's sized-down images for cell phone downloads, was completely superseded by the low-resolution thumbnail copies provided by search engines. The court's insistence on calling the defendant's wholly iterative scaled down copies "transformative" likely reflects the court's hesitance to shut down a powerful and popular search tool, rather than the belief that thumbnail images are "transformative" in the traditional sense. Moreover, the idea that a use need only have a new function or purpose to be "transformative," and thus constitute a fair use, does not explain contrary determinations regarding technologies like digital music lockers that enable new personal uses such as "space shifting."[72]

b) Many Consumptive Personal Use Copies Should Be "Fair" but Are Not "Transformative"

The fair use doctrine's shift toward a focus on transformativeness has rendered the doctrine particularly ineffective when courts evaluate personal use copies. This Section first examines fair use and research copies, and then examines fair use and consumer electronics.

i) Personal Use Copies for Research

Private copies made for research purposes are explicitly enumerated in § 107's preamble as an example of fair use.[73] This fits squarely within copyright's purpose of promoting progress in the arts and science because research, regardless of its setting, presumably advances knowledge.[74]

Nevertheless, in what is perhaps one of the most puzzling cases dealing with personal use copies, a court found that making personal use copies of a scholarly article obtained under paid license was copyright infringement.[75] In *Texaco*, one of the defendant's employees made photocopies of articles from the *Journal of Catalysis*—to which Texaco had three subscriptions—that the employee felt would facilitate his research, though he did not use all of them immediately.[76] The Copyright Act specifically enumerates copies made for

71. Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 579 (1994).

72. *Compare* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984) (finding that recording television broadcasts to "time shift," or watch at a later time, was fair use) *with* UMG Recordings, Inc. v. MP3.com, Inc., 92 F. Supp. 2d 349 (S.D.N.Y. 2000) (finding that copying CDs into a digital source locker to make them available from other locations, a practice later characterized as "space shifting," was not fair use).

73. 17 U.S.C. § 107 (2006).

74. *See* U.S. CONST. art. I, § 8, cl. 8.

75. Am. Geophysical Union v. Texaco Inc., 60 F.3d 913 (2d Cir. 1994).

76. *Id.* at 915–16.

"purposes such as . . . research" as an example of a fair use.[77] The *Texaco* court nonetheless found that the employee's copying was not fair use.[78] The court held that Texaco's copies were not "for purposes of research" because they were not used in the employee's published research, but were merely an "intermediate step" that might aid the employee's research.[79] This separation of the personal use copy from its use casts serious doubt as to whether a court would approve any personal use research copy, § 107 notwithstanding. Additionally, the court found that the district court did not over-emphasize the transformative test because,

> [t]o the extent that the secondary use involves merely an untransformed duplication, the value generated by the secondary use is little or nothing more than the value that inheres in the original. Rather than making some contribution of new intellectual value and thereby fostering the advancement of the arts and sciences, an untransformed copy is likely to be used simply for the same intrinsic purpose as the original, thereby *providing limited justification for a finding of fair use*.[80]

Finally, the court found that Texaco's copies were not transformative because they did not add anything new under *Campbell* and because all transformation incident to the photocopy was limited to the "*material object* embodying the intangible article that is the copyrighted individual work."[81] Therefore, the court held that Texaco was liable for copyright infringement.[82] Interestingly, under this standard, where research as part of a commercial enterprise is virtually presumptively unfair, attorneys, in their routine practice, would be considered rampant copyright infringers.[83]

Furthermore, *Texaco* casts doubt on whether even research universities, which often profit from their scientific advances through patent licensing, could make fair use copies for research purposes. In light of *Texaco*, scholars have expressed doubt that the fair use exception for research copies would even protect academics as the sweeping *Texaco* holding leaves little room to distinguish academic from commercial research.[84] In fact, *Texaco* casts doubt

77. 17 U.S.C. § 107 (2006).
78. *Am. Geophysical Union*, 60 F.3d 913.
79. *Id.* at 920 n.7.
80. *Id.* at 923 (emphasis added).
81. *Id.* (emphasis in original).
82. *Id.* at 931.
83. *See generally* Steven D. Smit, *"Make a Copy for the File . . .": Copyright Infringement by Attorneys*, 46 BAYLOR L. REV. 1 (1994).
84. Maureen Ryan, *Fair Use and Academic Expression: Rhetoric, Reality, and Restriction on Academic Freedom*, 8 CORNELL J. L. & PUB. POL'Y 541, 566–67 (1999).

as to whether any personal use copies made for research purposes would not be infringing, notwithstanding § 107.

By way of example, imagine a graduate student working on her dissertation. During the course of her research, she may make a photocopy of an article from her supervising professor's personal collection.[85] Although this graduate student would be a sympathetic defendant, her actions are infringing under *Texaco*. First, because *Texaco* separates the physical act of copying from its contextual purpose, the graduate student's copies would also be an "intermediate step" and thus would not be for research purposes under the statute. Second, the graduate student's photocopies would not be transformative because any content she adds in the course of writing her dissertation would not be considered. The photocopy, not the dissertation, is the act of infringement. Thus, if copyright owners strictly enforced the *Texaco* rule, virtually no research copies would constitute fair use, although this result is plainly contrary to congressional intent.[86]

More generally, stripping acts of copying from their larger contexts and conflating transformativeness with fair use are both impractical and inconsistent with the purposes of copyright law. In *Texaco*, the disputed copies were clearly not transformative in the way that a parody would be transformative—but they should have been fair use nonetheless. The copies were not distributed outside a small intimate circle, and they were made for purposes of research, which undoubtedly promotes progress in the sciences and arts.[87]

ii) Personal Use Copies as Content for Consumer Electronics

Historically, the copyright consumer has felt that her purchased copy was "hers" and that she was entitled to make personal use copies to maximize the portability and accessibility of her purchase.[88] With advances in technology, customers increasingly demand the ability to make personal use copies to "time shift" and "space shift" their media.[89] These personal use copies

85. This example was deliberately crafted so as to avoid the library exception under 17 U.S.C. § 108 (2006).

86. *See* Tehranian, *supra* note 18, at 544 n.33. Concededly, if the *Texaco* analysis was vigorously enforced, the conflict with section 107 would become glaringly apparent, and the courts would reverse the rule. The fact that copyright owners are unable or unwilling to push the law to its limits does not excuse bad law.

87. *See* U.S. CONST. art. I, § 8, cl. 8.

88. *See, e.g.*, OFFICE OF TECHNOLOGY ASSESSMENT, *supra* note 4, at 163.

89. *See, e.g.*, Brian Stelter, *Serving Up Television Without the TV Set*, N.Y. TIMES, Mar. 10, 2008, at C1 (outlining new digital distribution trends in television content); *see also*, Wikipedia, Total iPod Sales Chart, http://en.wikipedia.org/wiki/File:Ipod_sales.svg (last

should be permitted and encouraged under copyright law because they provide start-up capital that drives technological innovation in consumer electronics.[90] For example, without the belief that a consumer could lawfully copy his CD collection into his new iPod, or could legally record his favorite TV show transmissions into his TiVo, the early market for these devices would probably have been much more limited.[91]

The content industries, however, tend to be hostile to the idea of providing start-up capital to the consumer electronics industry. First, large content industry companies frequently flex their market power to try to squash new personal use copy-enabling technologies before they ever reach the consumer.[92] If this approach fails, content companies frequently sue technology companies under a theory of secondary liability for facilitating the end-user's alleged copyright infringement.[93] Despite the Supreme Court's ruling that personal use copies for purposes of "time shifting" as enabled by the VTR were fair use,[94] many distributors of newer technologies allowing analogous uses have been found to be infringing[95] or sued out of existence.[96] This Section will examine the fair use doctrine's inability to protect personal use copies within the context of the television and motion picture industry.[97]

Until very recently, the motion picture and television industries have been overwhelmingly hostile to technologies that enable consumers to make personal use copies of their content.[98] In *Sony v. Universal City Studios*—a

visited Nov. 8, 2009) (summarizing proliferation of the iPod, which is used to time and space shift media).

90. *See generally* von Lohmann, *supra* note 19, at 836–37 (arguing that many successful technological innovations such as the iPod and TiVo relied on personal use copies as startup capital).

91. *See id.*

92. *Reining in Tech*, *supra* note 50.

93. *See, e.g.*, Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984) (VTRs); Recording Indus. Ass'n of Am. v. Diamond Multimedia Sys., Inc., 180 F.3d 1072 (9th Cir. 1999) (digital music players).

94. *Sony*, 464 U.S. 417.

95. *See, e.g.*, UMG Recordings, Inc. v. MP3.com, Inc., 92 F. Supp. 2d 349 (S.D.N.Y. 2000).

96. *See* Paramount Pictures Corp. v. RePlayTV, 298 F. Supp. 2d 921 (C.D. Cal. 2004) (explaining in subsequent motion hearing that RePlayTV creator SONICblue filed for bankruptcy, and that the plaintiffs settled with the purchaser of SONICblue's assets).

97. The music industry, with its prominent use of Digital Rights Management and the infamous file sharing lawsuits brought by the Recording Industry Association of America (RIAA) also provides an example of particularly voracious copyright enforcement against personal use copies. This paper uses television as an example because *Sony* provides a particularly shocking backdrop for fair use's failure.

98. *See, e.g.*, *Sony*, 464 U.S. 417 (VTRs); Complaint, Paramount Pictures Corp. v. RePlayTV, Inc., 298 F. Supp. 2d 921 (C.D. Cal 2004) (No. CV 01-9358 FMC(Ex)); *cf.* Stelter,

foundational case for both the treatment of personal use copies and secondary liability under the 1976 Act—two copyright industry giants, Universal City Studios and Walt Disney Productions, sued Sony, the manufacturer of Betamax VTRs, for copyright infringement.[99] The studios claimed that consumers who used the Betamax to make copies of broadcast television programs were engaging in copyright infringement, and that Sony, as the manufacturer of the device facilitating this infringement, was secondarily liable.[100]

The Court, however, found that the consumers' personal copying was largely for purposes of "time shifting" broadcast television programs for viewing at a later date.[101] The Court found this "time shifting" to be fair use, and thus concluded that Sony was not secondarily liable.[102]

Furthermore, the Court explicitly disclaimed the idea, which lower courts nevertheless later embraced,[103] that a use needs to be transformative or "productive" to be fair:

> Congress has plainly instructed us that fair use analysis calls for a sensitive balancing of interests. The distinction between "productive" and "unproductive" uses may be helpful in calibrating the balance, but it cannot be wholly determinative.[104]

Unfortunately, *Sony* has been relatively unhelpful for subsequent technological innovators who do not develop their products with the blessing of content industries.[105] For example, SonicBlue developed RePlayTV, a digital age version of the VCR, that made digital copies of

---

supra note 89 (describing recent trends by studios to allow time and space shifting of their content through online services such as Hulu and licensing personal use copies through Apple's iTunes store).

99. *Sony*, 464 U.S. at 420.

100. *Id.*

101. *Id.* at 455.

102. *Id.* The staple article of commerce doctrine as a defense to secondary liability is outside the scope of this Note.

103. *See, e.g.*, Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir. 2007) (finding that scaled-down iterative copies of photos were transformative, and thus fair use); Am. Geophysical Union v. Texaco Inc., 60 F.3d 913 (2d Cir. 1994) (holding that research copies were not fair use, largely because they were not transformative); *see infra* Section III.A.2.a).

104. *Sony*, 464 U.S. at 455.

105. *Compare Reining in Tech*, *supra* note 50 (reporting that the industry sued the makers of RePlayTV, a DVR which allowed for easy commercial skipping, into bankruptcy) *with* Sandeep Junnarkar, *TiVo Casts NBC Exec as President*, CNET NEWS.COM, Apr. 30, 2003, http://news.cnet.com/TiVo-casts-NBC-exec-as-president/2100-1041_3-998937.html (describing TiVo's hiring of an executive vice president at NBC to develop partnerships between the DVR service and television networks).

television programming and allowed users to automatically skip commercials.[106] Even though roughly twenty years earlier the Supreme Court clearly held that making personal use copies of broadcast television programming to time shift was fair use, the television industry went into attack mode. Turner Broadcasting executive Jamie Kellner stated that, although the industry begrudgingly tolerates viewers leaving the room during commercials, using technology to avoid watching them was theft.[107] This bold declaration led *The San Jose Mercury News* to ask sarcastically whether going to the bathroom during a commercial break was copyright infringement.[108] But perhaps most importantly, the VTR, a recording device the Supreme Court specifically approved just two decades earlier, also allows users to skip commercials with a remote control.[109] The seemingly clear *Sony* precedent notwithstanding, the incredibly fact-specific fair use defense proved too expensive to litigate for the small Silicon Valley startup, and the industry successfully sued the innovator into bankruptcy.[110]

The industry then partnered with TiVo to create a nearly identical product with heavy industry involvement.[111] A TiVo digital video recorder (DVR) does not incorporate an auto-skip feature, and among other industry-friendly features, displays advertisements when a viewer fast forwards through a commercial.[112]

There is no real harm to consumers in the DVR example, as consumers can make time-shifting copies as easily with TiVo as they could have with RePlayTV. What is troubling as a matter of policy is that conceivably, if the industry had not been interested in making a suitable alternative, the technology never would have been distributed at all.[113] Additionally, this

106. Paramount Pictures Corp. v. RePlayTV, 298 F. Supp. 2d 921, 923 (C.D. Cal. 2004).

107. Benny Evangelista, *Hot Button Issue; TV Moguls Are Threatened by DVRs that Zip Past the Ads*, S.F. CHRON., May 27, 2002, at E1 ("It's theft . . . Your contract with the network when you get the show is you're going to watch the spots . . . I guess there's a certain amount of tolerance for going to the bathroom.").

108. Editorial, *Watch Commercials, Hollywood Screams; Litigation was Used to Harass Silicon Valley Company*, SAN JOSE MERCURY NEWS, Mar. 28, 2003, at 8C.

109. *Hot Button Issue*, *supra* note 107 ("The idea that someone would not be allowed to fast forward or skip commercials is a pretty outlandish concept to a country that has gotten used to 20 years of VCR ownership.") (comments of Fred von Lohmann).

110. *RePlayTV*, 298 F. Supp. 2d 921; *Watch Commercials*, *supra* note 108.

111. *See, e.g.*, Junnarkar, *supra* note 105.

112. Richard Shim, *TiVo Tests Pop-up-style Ads*, CNET NEWS.COM, Mar. 28, 2005, http://news.cnet.com/TiVo-tests-pop-up-style-ads/2100-1041_3-5644197.html?tag=mncol.

113. Von Lohmann, *supra* note 19, at 841. As the *Sony* litigation makes clear, the content industries are unable to discern innovations which enhance the value of their content from those that devalue it ex ante. Today, the home video and DVD markets, enabled by Sony's

dynamic harms the competitive market for consumer electronics. Because of the prospect of dubious yet expensive copyright claims, small companies who have no clout with content industries or the money to afford the fair use fight are completely boxed out of the market. In the case of RePlayTV, the small startup was squeezed out of the market for a technology it invented.

This ability of the copyright industry to use infringement lawsuits to block innovation is in direct contravention of the constitutional purpose of copyright.[114] The RePlayTV saga powerfully demonstrates that the fair use doctrine cannot adequately protect socially valuable personal use copies, even when Supreme Court authority clearly states that a particular use is fair.

If technological innovation affecting the copyright industries is to thrive, the copyright regime needs to be reformed to make clear that personal use copies do not infringe, such that a defendant can have a chance of having lawsuits dismissed on summary judgment rather than be forced into an expensive trial while gambling on fair use.[115]

3. Content Industries Can Indirectly Delineate the Border Between Fair Use and Infringement

The fair use doctrine is also particularly ill-suited to personal use copying because, in large part, the content industries can manipulate ex ante the fourth fair use factor, market impact, thereby intentionally shrinking the scope of fair use. The content industries reap an unjust benefit from fair use's unpredictability through a perverse cycle that James Gibson dubs "Copyright's Feedback Loop."[116] The cycle works as follows: first, it is almost impossible to determine ex ante whether a court would find that a particular use is fair or needs to be licensed;[117] second, the severe penalties a user would incur for wrongly deciding that a particular use is fair creates an overwhelming incentive to secure a license, even in cases where it should not be needed;[118] third, copyright owners are then able to show that what was previously seen as a "fair use" generates a significant licensing revenue stream, which tips the fourth fair use factor in its favor.[119] In cases where

---

allegedly infringing technology, account for a greater share of studio revenue than the box office.

114. *See* U.S. CONST. art. I, § 8, cl. 8.

115. *See infra* Part IV; *see also* von Lohmann, *supra* note 19, at 859.

116. James Gibson, *Risk Aversion and Rights Accretion in Intellection Property Law*, 116 YALE L.J. 882, 887–907 (2007).

117. *Id.* at 884.

118. *Id.* ("Better safe than sued.").

119. *Id.*; *see* 17 U.S.C. § 107 (2006).

iterative copies have been made, the fourth fair use factor is "undoubtedly the single most important element of fair use."[120] Therefore,

> the practice of licensing within gray areas eventually makes those areas less gray, as the licensing itself becomes the proof that the entitlement covers the use. Over time, public privilege recedes, and the reach of copyright expands . . . .[121]

If unchecked, the copyright feedback loop's effect on personal use copies could have several unfortunate consequences. With the advance of technology, personal use copies are no longer beyond the reach of copyright owners. For example, until January of 2009, songs purchased through Apple's popular iTunes service contained Digital Rights Management (DRM) encryption that allowed them to be played on only five user-authorized machines and to be burned onto storage media only seven times.[122] By creating a licensing scheme that is priced to a specific number of personal use copies, Apple and the music labels have extended the beginnings of the copyright feedback loop into the realm of personal use copies. Therefore, absent a strong legislative statement that personal use copies should be beyond the reach of copyright liability, it is only a matter of time before the fair use feedback loop consumes personal use copying, and extends the prying eyes of copyright enforcers into the privacy of the user's home.[123]

B. THE REPRODUCTION RIGHT'S FLAWS AS AN ENFORCEMENT MECHANISM FOR COPYRIGHT HOLDERS

The content industries should equally favor an expansive reform of the reproduction right. The current copyright regime as an enforcement mechanism is ill-suited to the realities of digital technology. Copyright

120. Harper & Row Publishers, Inc. v. Nation Enters., 471 U.S. 539, 566 (1985); *cf.* Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 590–92 (1994) (stating that in the case of transformative parody, the fourth factor takes diminished importance).

121. Gibson, *supra* note 116, at 884.

122. iTunes Store Terms of Service §§ 9(b), 10(b), http://www.apple.com/legal/itunes/us/service.html (last visited Nov. 8, 2009). Apple now offers some selections in DRM-free formats for a higher price—but only from those labels that have agreed to the change. *See, e.g.*, Greg Sandoval, *Sources: Apple to Expand DRM-free Music, Pricing*, CNET NEWS.COM, Jan. 5, 2009, http://news.cnet.com/8301-1023_3-10131761-93.html (noting that EMI had authorized DRM-free tracks, but that EMI accounted for only 10% of the iTunes library); Greg Sandoval, *Upgrading to a DRM-free iTunes Library Will Cost You*, CNET NEWS.COM, Jan. 6, 2009, http://news.cnet.com/8301-13579_3-10132759-37.html?tag=mncol;txt (explaining that stripping DRM from already owned tracks could be done for a fee, but that not all tracks were available DRM-free because of licensing issues).

123. For an examination of the privacy implications of digital copyright enforcement, *see generally* Megan L. Richardson, *Downloading Music off the Internet: Copyright and Privacy in Conflict?*, 13 J. L. & INFO. SCI. 90 (2002), *available at* http://ssrn.com/abstract=597362.

enforcement under the 1976 Act focuses on the act of "copying" as the infringement.[124] The purpose of the Copyright Act is not to prevent copying of original works of authorship but to create an incentive structure to encourage authors to enhance the public domain with their creative works.[125] The prevention of unauthorized copying in and of itself is thus not the goal, but a means to an end.[126] Policing physical copies once served as an efficient means to that end, because copies were a prerequisite for distribution, and personal use copies, which did not lead to distribution, were virtually undetectable.[127] These realities do not hold in the digital environment.

In the digital age, economic harm to the copyright holder no longer correlates to the number of unauthorized reproductions. While the courts are divided as to whether making copyrighted works available on a file sharing server violates copyright in the absence of proof of downloading,[128] common sense informs us that uploading a single copy to Kazaa represents a greater economic harm to the copyright owner than a home user burning her iTunes purchase onto fourteen different mix-CDs.[129] Faced with the reality of P2P file sharing, copyright owners are in need of an enforcement mechanism that addresses the actual harms that copyright law is meant to protect.

Most importantly, however, is that use in the digital age often *requires* copying.[130] For example, a computer must make a copy of a copyrighted software program in its RAM for the user to enjoy the program she purchased.[131] This enforcement problem has created an environment where copyright holders can prevent scientists from making lab copies[132] but cannot collect damages from those who make copies of copyrighted music available to anyone with an Internet connection if they cannot prove actual copies

124. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv., Co., 499 U.S. 340, 361 (1991) ("To establish infringement, two elements must be proven: (1) ownership of a valid copyright, and (2) copying of constituent elements of the work that are original.").

125. *See* U.S. CONST. art. I, § 8, cl. 8; THE DIGITAL DILEMMA, *supra* note 6, at 140.

126. THE DIGITAL DILEMMA, *supra* note 6, at 140.

127. *Id.* at 142.

128. *Compare* Elektra Entm't Group v. Barker, 551 F. Supp. 2d 234 (S.D.N.Y. 2008) (making a file available on a P2P distribution network is copyright infringement) *with* London-Sire Records, Inc. v. Doe, 542 F. Supp. 2d 153 (D. Mass. 2008) (making a file available on a P2P distribution network is not copyright infringement).

129. *See supra* Section III.A.3 (demonstrating the privacy implications of new DRM technologies).

130. THE DIGITAL DILEMMA, *supra* note 6, at 142.

131. *Id.* Note that early cases held that such copies constituted copyright infringement. *See, e.g.*, MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993).

132. Am. Geophysical Union v. Texaco Inc., 60 F.3d 913 (2d Cir. 1994).

were made.[133] From a constitutional standpoint, copying scientific research to allow further breakthroughs actually spurs science forward and promotes progress, whereas mass unauthorized distribution of copyrighted music destroys its market value and vastly undercuts the record company's incentive to promote and distribute new music. Therefore, though current copyright law holds that making copies available on the Internet is acceptable, while making private research copies is not, the intellectual property clause of the U.S. Constitution mandates the exact opposite result.[134]

Given these realities, copyright holders are in equal need of comprehensive reform. The content industries' financial interests would be satisfied by a new system that protected the making of private research copies and prohibited the distribution of files over the Internet. As such, content industries should favor the comprehensive reform proposed in the next Section.

IV. THE REFORM

The current copyright regime is ill-suited to the needs of both copyright consumers and copyright producers. Section IV.A argues that, in light of the problems discussed in Part III, both users and copyright holders would benefit from comprehensive copyright reform.

To create a copyright enforcement mechanism more attuned to digital realities, Section IV.B proposes a commercial appropriation right that would identify unauthorized distribution, not copying, as the locus of economic injury. To promote and protect the social benefits of personal use copies, Section IV.C proposes that a specific personal use exemption, similar to that codified in Swiss Law, be incorporated into Title 17.

A. THE NEED FOR REFORM

As stated above, the purpose of copyright protection in U.S. law is to expand the public domain by promoting progress in the arts and sciences.[135] An intellectual property regime that does not achieve its constitutional goal should be amended. The 1976 Act's focus on exclusive reproduction rights as a means of controlling economic incentives to create served its constitutional purpose before the advent of digital storage media, but it has

133. *London-Sire Records*, 542 F. Supp. 2d 153 (making a file available on a P2P distribution network is not copyright infringement).

134. *See* U.S. CONST. art. I, § 8, cl. 8.

135. *See id.*

since become antiquated.[136] With the advent of digital storage, the Internet, and file sharing technologies, a single digital copy offered to the public can be infinitely more damaging to the copyright incentive structure than a dozen physical personal use copies.[137] The current regime under the 1976 Act focuses too much on the physical copy, while denying sufficient recovery for unauthorized digital communications.[138] It cannot effectively preserve artists' incentives to create, and it encourages invasion of privacy as a means of copyright enforcement. For example, if a user were making files available on public P2P file sharing networks, a copyright holder would have to prove that downloads had actually taken place.[139] Such a burden of proof, however, would encourage record companies to monitor the private activities of users.[140]

In this sense, the 1976 Act has also become inadequate for users. Before digital technologies, a robust personal use exception to copyright liability was not particularly important. Personal use copies in the "analog" era were virtually undetectable by copyright owners and thus unenforceable through direct infringement lawsuits.[141] In the digital age, however, content owners can and do track personal use copies.[142] The content owners' perceived right to police personal use copies could have disastrous consequences for users' privacy.[143]

Although Apple's previous DRM restrictions on personal use copies, discussed above, seem relatively innocuous, other content providers have introduced more nefarious tracking. For example, in 2005, Sony BMG, quietly introduced two controversial software programs, MediaMax and Extended Copy Protection (XCP), on its CDs.[144] These programs installed

136. *See* THE DIGITAL DILEMMA, *supra* note 6, at 142.

137. *See supra* Section III.B (discussing the nature of the reproduction right, and the uncertainty as to whether "making available" a copyrighted work over the internet is even actionable infringement).

138. *See supra* Section III.B (discussing the "making available" controversy).

139. London-Sire Records, Inc. v. Doe, 542 F. Supp. 2d 153 (D. Mass. 2008).

140. *See, e.g.*, Electronic Frontier Foundation, Sony BMG Litigation Info, http://www.eff.org/cases/sony-bmg-litigation-info (last visited Nov. 8, 2009) (explaining the Sony rootkit scandal, where Sony music CDs installed use-monitoring software on computers without customer knowledge or permission).

141. *See* THE DIGITAL DILEMMA, *supra* note 6, at 142; *supra* Section III.A.2.b)ii) (explaining *Sony* and other third party liability litigation).

142. Section III.A.3 (explaining Apple's ability to track personal use copies in its iTunes program).

143. *See generally* Richardson, *supra* note 123.

144. *See* Electronic Frontier Foundation, *supra* note 140.

hidden files onto the user's computer and monitored CD usage.[145] Additionally, MediaMax transmitted the user's listening habits back to SunnComm, which created a security vulnerability that exposed the user's computer to malicious attacks by third parties.[146] The security vulnerabilities raised public ire, and Sony agreed to a settlement.[147] Had Sony better hidden its behavior, the surveillance of its customers likely would have continued.

This ability and willingness of content owners to track all personal use copies and aggregate usage data in real time is a particularly intrusive and unnecessary invasion of privacy. To address this problem, some scholars have suggested that privacy law should be used to deter copyright holders' overuse of invasive protection efforts.[148] A privacy law solution, however, would likely require a privacy notice that users would not read or have the bargaining power necessary to contest.[149] Without addressing the copyright holders' underlying motivation to monitor consumer copying, the law would be perpetually playing catch-up with ever-enterprising copyright holders armed with engineers and lawyers intent on circumventing any privacy regulation.

In contrast, if copyright law made clear that personal use copies do not constitute actionable infringement and that making a work available to the public does,[150] then copyright holders would have little incentive to monitor private individual uses and a greater incentive to focus on public communications. Society should not sacrifice effective protection of personal privacy merely to preserve the antiquated idea that the physical "copy" is the locus of copyright infringement. Clear statutory protection for private personal use copies could push back against future surveillance efforts by content owners.

Most importantly, from a constitutional standpoint, the 1976 Act's insufficient protection for personal use copies hinders the development of new technologies that actually increase the economic value of copyrighted material.[151] A myriad of technologies from the VTR to the iPod have relied

145. Class Action Complaint at 2, Melcon v. Sony BMG Music Entm't, No. C 05-5084 MHP (N.D. Cal. Dec. 8, 2005), *available at* http://w2.eff.org/IP/DRM/Sony-BMG/ND_cal_complaint.pdf (last visited Nov. 8, 2009).

146. Electronic Frontier Foundation, *supra* note 140.

147. *Id.*

148. *See, e.g.*, Julie E. Cohen, *DRM and Privacy*, 18 BERKELEY TECH. L.J. 575 (2003).

149. *See generally* Edward J. Janger & Paul M. Schwartz, *The Gramm-Leach-Bliley Act, Information Privacy, and the Limits of Default Rules*, 86 MINN. L. REV. 1219 (2002) (explaining the ineffectiveness of the Gramm-Leach-Bliley Act's imposition of a privacy notice requirement in the financial sector).

150. *See infra* Sections IV.B and IV.C.

151. *See generally* von Lohmann, *supra* note 19.

on the understanding that personal use copies are fair use, and their development has launched other technological innovations as well.[152] If the copyright feedback loop were to continue, content owners would have even more opportunities to sue manufacturers of threatening technologies out of existence,[153] thereby depriving the public of the very progress the intellectual property clause seeks to promote.[154] It is tempting to leave the determination of what innovations will enhance the value of copyrighted works to copyright industries that already have a vested interest in maximizing the value of their works. However, because of risk aversion and growth seeking dynamics within large businesses, such as film studios or record labels, copyright industry players are notoriously unable to determine ex ante which personal-use-capitalizing innovations will increase or decrease the value of their copyrighted works.[155]

Therefore, it is unquestionably in the best interest of the copyright holder, the consumer, and the public at large to create a new overarching copyright principle that protects copyrighted works from online piracy, discourages copyright holders from violating user privacy to find possible infringement, and creates a healthy environment that fosters innovation in copyright-consuming technologies. The following Sections propose such a system.

B. COMMUNICATION TO THE PUBLIC AS THE PROXY FOR INFRINGEMENT

The World Intellectual Property Organization Copyright Treaty (WCT), negotiated in 1996, provides a much better starting point to create an infringement enforcement mechanism compatible with the realities of the digital era. Article 8 of the WCT provides that

> authors of literary and artistic works shall enjoy the exclusive right of authorizing any communication to the public of their works, by wire or wireless means, including the making available to the public of their works in such a way that members of the public may access

152. *See generally id.*

153. *See generally* Paramount Pictures Corp. v. RePlayTV, 298 F. Supp. 2d 921 (C.D. Cal. 2004); *Reining in Tech*, *supra* note 50; *cf.* Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417 (1984) (finding that recording television broadcasts to "time shift" was fair use).

154. U.S. CONST. art. I, § 8, cl. 8.

155. *See* von Lohmann, *supra* note 19, at 854–55.

> these works from a place and at a time individually chosen by them.[156]

With the exception of the derivative work right, all of the exclusive rights listed in § 106 can be classified as specific categorizations of the author's overarching right to control the communication of her work to the public.[157] At their core, the rights of reproduction, distribution, public performance, and public display are simply different means of preserving the author's incentive to create by granting the author control over the communication of her work to the public. This is particularly evidenced by the fact that private performances and displays are not protected by the 1976 Act.[158]

Collapsing these separate exclusive rights into a single overarching right of communication would simplify the Copyright Act and make it more accessible to the public. This would aid copyright holders in their efforts to educate the public about copyright infringement.[159] Currently, under § 106, where all copies, no matter how private or commercially insignificant, are technically infringement, the public determines for itself which personal use copies are acceptable and which are not.[160] The public consensus has been that copying is acceptable as long as the copy is not sold.[161] This normative line of demarcation is unacceptable in light of P2P file sharing, where one copy that is not sold is obtainable by thousands of potential customers free of charge. Although the copyright industries have embarked on large-scale

156. World Intellectual Property Organization Copyright Treaty, art. 8, Dec. 20, 1996, 112 Stat. 2860, 2186 U.N.T.S. 152, *available at* http://www.wipo.int/treaties/en/ip/wct/trtdocs_wo033.html.

157. *See* 17 U.S.C. § 106(1) (2006) (reproduction right); § 106(3) (distribution right); § 106(4), (6) (public performance rights); § 106(5) (public display right).

158. *See* 17 U.S.C. § 106(4)–(6) (2006); *see, e.g.*, *In re* Application of Cellco P'ship, 663 F. Supp. 2d 363 (S.D.N.Y. 2009) (finding that ASCAP was not entitled to royalties from cell phone ringtone plays because the ringtones did not constitute a "public" performance).

159. Education is one of the main strategies that the content industries use to combat widespread digital piracy by the public at large. *See, e.g.*, Recording Industry Association of America, Piracy: Online and on the Street, http://www.riaa.com/physicalpiracy.php (last visited Nov. 8, 2009) [hereinafter Piracy: Online and on the Street]; Motion Picture Association of America, Respect Copyrights, http://www.respectcopyrights.org/ (last visited Nov. 17, 2009).

160. For an overview of just how far the reproduction right extends, *see generally* Tehranian, *supra* note 18.

161. OFFICE OF TECHNOLOGY ASSESSMENT, *supra* note 4, at 3.

education campaigns to combat this problem,[162] the sheer complexity of the 1976 Act stands as a formidable obstacle.[163]

A simplified and streamlined right of communication to the public maps much more logically onto digital file sharing realities and increases public comprehension and acceptance. Although many would disagree with the idea that a record label is entitled to royalties every time you make a CD for your car or a music mix for a friend,[164] the public, particularly in light of the advent of P2P file sharing, would likely understand and agree with the proposition that public dissemination of copies should be the sole prerogative of the copyright holder. Under this regime, because acts such as including the original text in an email reply unquestionably would not be copyright infringement,[165] copyright law would become more credible and compliance with it would likely increase.[166]

Furthermore, a broad public communication right would make § 106 more easily adaptable to new technologies, thereby avoiding much of the folly of the 1976 Act's enforcement in the current digital era.[167] Therefore, the WCT obligation should be adapted to U.S. copyright law as follows:

> § 106 Exclusive Rights in Copyrighted Works:
>
> Subject to sections 107 though 122, the owner of a copyright under this title has the exclusive rights to do and to authorize any of the following:
>
> (1) any communication to the public of their works by any means, including a public performance, a public display, or the making available to the public of their works in such a way that members of the public may access these works from a place and time individually chosen by them;

---

162. *See, e.g.*, Piracy: Online and on the Street, *supra* note 159; Respect Copyrights, *supra* note 159.

163. *See, e.g.*, OFFICE OF TECHNOLOGY ASSESSMENT, *supra* note 4, at 163 (finding that a consumer's perceived familiarity with copyright law had no effect on home taping habits).

164. *See id.* at 3, 12, 157 (1989) (finding that the most common use of personal copies was shifting to a different playback device, and that most thought giving a copy to a friend was acceptable); THE DIGITAL DILEMMA, *supra* note 6, at 134 (explaining the sentiment of private use advocates that what a consumer did with his own copy in his own home was none of the copyright holder's business).

165. *Cf.* Tehranian, *supra* note 18, at 543, 547 (listing forwarding an email as an example of an unexpected act of infringement).

166. *See* THE DIGITAL DILEMMA, *supra* note 6, at 212–13 ("When popular attitudes and practices are out of synch with laws, the enforcement of laws becomes more difficult. . . . There are also political dangers associated with criminalizing generally accepted behavior.").

167. *Cf.* MAI Sys. Corp. v. Peak Computer, Inc., 991 F.2d 511 (9th Cir. 1993) (finding that RAM copies were infringing).

(2) to prepare derivative works[168] based on the copyrighted work.

The precise boundary between what is "public" and what is "private" will be discussed in the next Section.

C. A ROBUST PERSONAL USE EXEMPTION TO DEFINE WHAT IS NOT PUBLIC

Much of what makes a public communication right ideally suited to the technological age is its adaptability. To function properly, however, a public communication right under U.S. law, unlike the WCT provision, must define what is meant by "the public."[169] Instead of attempting to determine precisely what is and is not public, a more feasible alternative is to follow the example of other countries and define what is "private" and therefore not actionable.[170]

U.S. copyright law should delineate the scope of what is a public communication by codifying a clear personal use right, naming very specifically what are not "public communications," and leaving the "gray areas" between what is obviously public and obviously private to the ordinary common law process. This safe harbor for personal uses would also foster innovation in copyright consumptive technologies, which have the potential to increase the value of copyrighted content, which in turn increases the incentives for authors to create.[171]

The personal use exemption in Swiss copyright law provides a suitable starting point. Article 19 of the Swiss copyright law, in pertinent part, holds that:

> 1. Published works may be used for private purposes. Private use shall mean:
>
> a. any use of a work in the personal sphere or within a circle of persons closely connected to each other, such as relations or friends;
>
> b. any use of a work by a teacher for teaching in class;

168. An examination of digital technology's implications on the derivative work right is beyond the scope of this Note.

169. Guido Westkamp, *Transient Copying and Public Communications: The Creeping Evolution of Use and Access Rights in European Copyright Law*, 36 GEO. WASH. INT'L L. REV. 1057, 1083 (2004) (explaining that because public communication is an obscure concept "the right stands and falls depending only on an interpretation of 'the public' ").

170. *See id.* at 1081 (explaining that the U.K. focuses on what is "private" to define what is "public").

171. *See generally* von Lohmann, *supra* note 19 (arguing that fair use incentivizes investment in technologies that are complementary goods to copyrighted works).

> c. the reproduction of copies of a work in enterprises, public administrations, institutes, commissions and similar bodies for internal information or documentation.[172]

Section 1.a. is easily adaptable to the U.S. legal regime and provides a logical basis for determining what uses are private and thus noninfringing. However, the remaining sections, which exempt from liability copies made for teaching in both academic and corporate settings, are more problematic in an age where businesses are multinational conglomerates and universities include several campuses of tens of thousands of faculty and students.[173] Exempting all copies made within these very large spheres from any sort of remuneration to the copyright owner could harm the incentive structure the Copyright Act seeks to create. At the same time, subjecting instructional copies made in business and educational institutions, entities that create much of the progress in arts and sciences, to statutory damages penalties[174] would not serve the purpose of copyright law.

Therefore, copies made within businesses and educational institutions for commercialized purposes, such as creating course readers for purchase by students,[175] or for purposes of securing a commercialized patent, or for product development, should be subject to a compulsory licensing scheme similar to that contained in § 115.[176] The licensing fee would be set by the Register of Copyrights at a sufficiently low level as to allow unfettered use of material in course readers without allowing entire works to be used and commercially distributed without the copyright holder's permission. The compulsory license would serve as the maximum that a copyright owner could demand from a research institution. It would in no way bar publishers, particularly university publishers, from authorizing free use of their materials. Fees for orphan works would be held in trust by the Register of Copyrights, should the author be located at a later date.[177]

---

172. Loi fédérale sur le droit d'auteur et les droits voisin, [Federal Law on Copyright and Neighboring Rights], Oct. 9, 1992, RS 101, art. 19 (Switz.).

173. The University of California, for example, has ten university campuses which collectively enroll more than 220,000 students. University of California Home Page, http://www.universityofcalifornia.edu/campuses/welcome.html (last visited Nov. 8, 2009).

174. *See* 17 U.S.C. § 504(c) (2006).

175. *See, e.g.*, Princeton Univ. Press v. Mich. Document Servs., Inc., 99 F.3d 1381 (6th Cir. 1996) (en banc) (finding liability where copyrighted material was used to create commercially sold course packs).

176. 17 U.S.C. § 115 (2006).

177. The logistics of how to deal with the orphan works problem, particularly the question as to how long the Register of Copyrights should be required to hold any royalties paid, is beyond the scope of this Note.

Because this proposed personal use exemption is not a definition per se, but an exemption meant to serve as a guideline as to what uses are "public," this section should be codified as § 107A, to follow fair use. The section should be as follows:

> § 107A Personal Use Exemption:
>
> Notwithstanding the provisions of sections 106 and 106A,
>
> (1) private uses of works protected under this title shall not give rise to any cause of action. Private uses are to include any use of a work in the personal sphere or within a circle of persons closely connected to each other, such as relations or friends. Third parties who enable such private uses are not subject to liability under this title;
>
> (2) unauthorized internal uses of works protected under this title within a single business or educational institution shall be permitted upon payment to the author of a fee to be fixed by the register of copyrights. In instances where a good faith search does not determine the author of a protected work, the register of copyrights shall collect the fee to hold in trust for the unnamed author.

V. CONCLUSION

At the intersection of the entertainment and consumer electronics industries, the 1976 Act fails to fulfill its constitutional purpose to promote progress. Instead, the Act actually disincentivizes the creation of personal-use-enabling technologies through the threat of contributory infringement liability, encourages content producers to waste money by suing consumers, and leaves ordinary citizens, who often do not see their conduct as infringement, open to lawsuits for which they are unprepared. To combat this problem, the reproduction right should be amended as follows:

> SECTION 102. RIGHT OF COMMUNICATION TO THE PUBLIC
>
> Section 106 of title 17, United States Code is amended to read as follows:
>
> "Subject to sections 107 through 122, the owner of copyright under this title has the exclusive rights to do and to authorize any of the following:
>
> (1) any communication to the public of their works by any means, including a public performance, a public display, or the making available to the public of their works in such a way that members of the public may access these works from a place and time individually chosen by them; and
>
> (2) to prepare derivative works based on the copyrighted work."

SECTION 103. PERSONAL USE COPIES

The following is added to title 17 of the United States Code as Section 107A:

"Notwithstanding the provisions of sections 106 and 106A,

(1) private uses of works protected under this title shall not give rise to any cause of action. Private uses are to include any use of a work in the personal sphere or within a circle of persons closely connected to each other, such as relations or friends. Third parties who enable such private uses are not subject to liability under this title;

(2) unauthorized internal uses of works protected under this title within a single business or educational institution shall be permitted upon payment to the author of a fee to be fixed by the register of copyrights. In instances where a good faith search does not determine the author of a protected work, the register of copyrights shall collect the fee to hold in trust for the unnamed author."

MEMO TO CONGRESS

A DISTRICT JUDGE'S PROPOSAL FOR PATENT REFORM: REVISITING THE CLEAR AND CONVINCING STANDARD AND CALIBRATING DEFERENCE TO THE STRENGTH OF THE EXAMINATION

William Alsup[†]

Patent reform is hot these days. This short Article gives my perspective, as a federal district court judge in a patent-intensive venue, on what is wrong, and my proposal for a simple, if only partial, fix. Over my decade on the bench in San Francisco, I have presided over more than one hundred patent infringement actions, ten of which went through trial. In my earlier career as a trial lawyer, I litigated two patent trials to completion. In this Article, I speak individually and for no one else. Although this Article sets forth a recommendation, my duty as a judge to faithfully uphold the law, whether or not Congress adopts my recommendation, will be foremost.

The foremost advantage of our patent system is that it protects one of the crown jewels of the United States—its intellectual property. The patent system's foremost problem is that too many invalid and weak patents get through the U.S. Patent and Trademark Office (PTO), which are then lorded over competitors and competitive products—without advancing any worthwhile interest. This is an unwelcome clog on commerce.

The vast explosion in patent litigation began in the mid-1980s, resulting from the reinvigoration of the patent system by Congress, including its establishment of the Court of Appeals for the Federal Circuit.[1] That court, which hears all patent appeals, did what Congress presumably wanted and breathed great strength into patents and their protection. Perhaps too much

† William Alsup has served as a U.S. District Court judge since 1999 in the Northern District of California. Before that, he was a trial lawyer at Morrison & Foerster in San Francisco. He also served as a law clerk to justice William O. Douglas (1971–72) and as an Assistant Solicitor General (1978–80). The author thanks his extern Alexander C. Touma for his assistance in preparing this Article.

1. United States Court of Appeals for the Federal Circuit, About the Court, http://www.cafc.uscourts.gov/about.html (last visited Mar. 16, 2010).

so. So much reinvigoration occurred that patent litigation, once a quiet backwater of the federal courts until the 1980s, has become a swelling sea.

A central reason for the litigation boom is the presumption of validity and the "clear and convincing" standard: Patents are presumed valid under 35 U.S.C. § 282 and can only be set aside in litigation upon "clear and convincing" proof of invalidity.[2] This presumption of validity applies equally to all patents—even those that are almost certainly invalid. This is a huge advantage for the patent holder—and it is often an unfair advantage, given the ease with which applicants and their agents can sneak undeserving claims through the PTO. Because of the burnish of this presumption, patentees can use a weak, arguably invalid patent, to force an accused infringer through years of litigation. This is more than just a nuisiance. Legal defense costs run, at the low end, about three million dollars per case, and range well over ten million dollars in some actions.[3] In the United States, the number of patent infringement suits filed annually nearly doubled between 1994 and 2004.[4] According to the Phoenix Center for Advanced Legal and Economic Public Policy Studies, patent litigation costs the economy 4.5 billion dollars annually.[5]

The presumption of validity and the clear and convincing standard would be wise if almost all patents were, in fact, valid. My own experience, however, has been that at least one-third of patent claims asserted in litigation should never have issued. I believe most lawyers, academics, and judges would agree that far too may invalid patents slip through the PTO, even though they would vary up or down from my percentage estimate.[6] But invalid or not, they all receive the forensic advantage of a legal earthwork fortified by a protective moat, namely the presumption of validity in tandem with the clear and convincing standard of proof that is required to overcome that presumption.

---

2. 35 U.S.C. § 282 (2006).

3. George S. Ford et al., *Quantifying the Cost of Substandard Patents: Some Preliminary Evidence* 27 (Phoenix Ctr. Policy Paper Series, Paper No. 30, 2007).

4. *See* J. SHAWN MCGRATH & KATHLEEN M. KEDROWSKI, AM. BAR ASS'N, TRENDS IN PATENT DAMAGES, *available at* http://www.docs.piausa.org/ABA/07-06-01-ABA-Report-On-Patent-Damages.pdf.

5. Ford, *supra* note 3, at 3.

6. *See, e.g.*, FTC, TO PROMOTE INNOVATION: THE PROPER BALANCE OF COMPETITION AND PATENT LAW 6 (2003), *available at* http://www.ftc.gov/os/2003/10/innovationrpt.pdf (noting that Professor Mark Lemley, a panelist at a 2002 FTC and DOJ panel on patent reform, found that 45–46% of all patents litigated to final results are held invalid).

Notably, Congress has not adopted the clear and convincing standard. It evolved via caselaw.[7] The standard arrived before the patent litigation explosion—indeed, it probably helped to bring on that explosion. Congress (or perhaps the Supreme Court) could intervene to conform the standard to the normal evidentiary standard used in trademark and other areas of the law, which is, as stated, by a preponderance of the evidence. No less than the Federal Trade Commission, not to mention distinguished professors, have urged this change![8]

In 2003, the FTC released a careful critique of our patent system and suggested how to improve it.[9] Their report listed many possible fixes. As a concession to the shortness of life, this Article focuses sharply on only one fix, a small change that would do much good: (1) to reduce the level of proof required to invalidate a patent to a preponderance of the evidence, the normal standard of proof in civil cases, and (2) to modulate the degree of deference due to the PTO by the extent to which the invalidity question surfaced during examination, and if the applicant prevailed on that invalidity question, the extent to which it was addressed. Under my proposal, Section 282 would be revised to state, in part,

> A patent claim is presumed to be valid and may be found invalid only upon proof of invalidity by a preponderance of the evidence. In evaluating the question and in evaluating the degree of deference to be accorded a Patent and Trademark Office action or allowance, the trier of fact may take into account the extent to which the examiner was afforded a reasonable opportunity to address the specific question of invalidity and the extent to which the examiner reasonably addressed the specific question of invalidity.

Note well that my proposal would also expressly reduce or increase the deference accorded a PTO examination by the extent to which the same invalidity issue was or was not vetted by the agency. This would codify a wise suggestion the Supreme Court made recently when it said that the rationale for the presumption was "much diminished" when the prior art in question was not disclosed during the PTO's examination.[10]

---

7. Doug Lichtman & Mark Lemley, *Rethinking Patent Law's Presumption of Validity*, 60 STAN. L. REV. 45, 51–52, 61 (2007); Mark D. Janis, *Reforming Patent Validity Litigation: The "Dubious Preponderance"*, 19 BERKELEY TECH. L.J. 923, 924–25 (2004).

8. *See* FTC, *supra* note 6, at 32; Lichtman, *supra* note 7, at 49; Janis, *supra* note 7, at 926, 932–35.

9. FTC, *supra* note 6.

10. KSR Int'l Co. v. Teleflex Inc., 550 U.S. 398, 426 (2007).

I recognize that this change would not have a dramatic effect at the summary judgment stage, because changing the standard of proof and the degree of deference cannot overcome plausible, if weak, evidence of validity at the summary judgment stage. But one impact of the amendment would be at trial. No longer would juries woodenly accord great deference to PTO examinations wherein the key prior art issues never even surfaced or, if surfaced, were given short shrift. This, in combination with application of the lower burden of proof, would make the invalidity trial a fair contest and, in turn, deter at least some infringement actions based on weak patents. Another important impact would occur before patents ever reached the courthouse. In PTO prosecutions, the amendment would encourage more disclosure, and more *pointed* disclosure, thereby giving the PTO examiner a better chance to reach the right outcome.

♦ ♦ ♦

How do so many weak patents slip through the PTO? The main reason is that the patent-application process before the PTO is largely one-sided. The only side that gets to present a case is the applicant, who has a manifest interest in slanting matters in a light most favorable to allowance. There is no adversary to present or argue the opposing view to the examiner. In fact, the competitors against whom the claims will eventually be asserted are often kept unaware of the entire proceeding.[11]

Patent claims, once granted, have the force of law, that is, of rules and regulations. An issued patent claim amounts to an agency determination to exclude everyone but the patent holder (and its licensees) from practicing any commerce within the scope of the claim. One premise of the Supreme Court's *Markman* decision in 1996 was that patent claims are matters of "law," whose scope a judge must construe, rather than "fact" for a jury to construe.[12] Yet issuance of patent claims by the PTO is one of the few times in our legal system that rulemaking occurs without opportunity for public comment. Indeed, it occurs ex parte, with only the applicant's case presented to the agency.

In the United States, we usually believe that there are two sides to every story—that is certainly the premise of the adversary system and the even-handed way in which the courts try cases and Congress conducts

11. Applications are published eighteen months after filing, unless the applicant requests that the application remain unpublished and certifies that he has not and will not file for patent protection in a foreign jurisdiction that requires publication within eighteen months of filing. *See* 37 C.F.R. § 1.213 (2010).

12. Markman v. Westview Instruments, Inc., 517 U.S. 370, 388 (1996).

hearings. But the opposite premise is used in obtaining a patent. Even if there are two or more sides, the PTO listens to only one side, and trusts that side to state all positions fairly. The truth is that the applicants and their attorneys are skilled at prosecuting the case for patentability without revealing the best opposing arguments. Although they are obligated to disclose adverse prior art known to them, they sometimes fail to do so, or, if they do, they minimize these references with one-sided arguments or bury them in a long list of prior art references. No adversary is there to keep them honest and present the other side of the story. PTO examiners are overwhelmed with work, and often do not have the time to develop the best counterarguments. They are no match for the professionals who earn large sums for guiding patent applications through the PTO. In short, the core problem is that the agency engages in what amounts to rulemaking that is binding on the entire economy, yet gives only one party, the one with the greatest incentive to distort, an opportunity to be heard. No other patent system in the world is so one-sided.[13]

One of the advantages of my proposal is that it would incentivize patent applicants to search out the prior art, to explain it to the examiner, and to call attention to the most difficult questions of invalidity, all with a view toward overcoming the references and obtaining a stronger prosecution record for litigation. This would encourage more disclosure to the examiner and would benefit deserving patents. Conversely, it would work against weaker patents by assisting the examiner to see through them, or, if that opportunity was denied to the examiner, it would be easier for an adversary to "undo" the patent in court. This small change would be easy to write into Section 282 and would in no way interfere with continuing consideration of longer-range reforms.

To be sure, recent reforms now allow a third party to have a limited voice to the PTO. Pre-issuance, i.e., prior to allowance of a patent, current PTO procedures allow a third party to submit prior art in two ways: "protests"[14] and "third-party submissions."[15] In post-issuance, i.e., post-allowance,

13. *See, e.g.*, Convention on the Grant of European Patents arts. 99–105, Oct. 5, 1973, 1065 U.N.T.S. 255 (establishing post-grant opposition procedure); *see generally* Dale L. Carlson & Robert A. Migliorini, *Past as a Prologue for Patent Reform: Experience in Japan with Oppositions Suggests an Alternative Approach for the U.S.*, 88 J. PAT. & TRADEMARK OFF. SOC'Y 101 (2006) (describing the Japanese patent opposition system); Anna Mayergoyz, *Lessons from Europe on How to Tame Trolls*, 42 CORNELL INT'L L.J. 241, 260–63 (describing the European patent opposition system).

14. U.S. PATENT AND TRADEMARK OFFICE, MANUAL OF PATENT EXAMINATION PROCEDURE §§ 1900–1920 (8th ed. 7th rev., 2008) [hereinafter MPEP].

15. *Id.* §§ 1134–1134.01.

procedures, a third party may make a challenge through the "reexamination" process.[16] It is worth pausing over these reforms to explain why they leave the basic problems unsolved.

These limited procedures do not cure the problem. A protest allows a any third-party member of the public to submit information challenging a pending patent application. It may include a listing of patents, publications, or other information relevant to the prosecution process. The third party may also explain the relevance of each reference. The protest, however, must be submitted *prior to the date* the application is published or a notice of allowance is mailed, whichever occurs first. A patent is published, pursuant to 35 U.S.C. § 122 (subject to exceptions), eighteen months from the filing date. Thus, to challenge a pending patent, the challenger must somehow already be aware of the existence of the application,which is usually a secret. (Patent applicants are understandably unlikely to make this information available to parties who would be adverse to their position.) Notably, this procedure does not allow for an open third-party submission process wherein the general public is put on notice of a pending application and provided an opportunity to give input. Further, once the protest is filed, the applicant will take pains to "distinguish" all of the references, but the protestor is not allowed to participate in any follow-up. In short, there is no public notice or opportunity for public comment.

If the PTO issues a patent notwithstanding the protest, the patent holder is able to argue in litigation that the reference was "before the examiner" and the patent issued regardless, meaning that the examiner must have felt the reference did not stand in the way of patentability. Given the presumption of validity, this is a hard argument to overcome in litigation by clear and convincing evidence. Rather than protest in this way, opponents[17] of the proposed patent will thus usually prefer to "keep their powder dry" and wait to be sued in a courtroom where they will have a more fair, two-sided contest on the prior art (despite the burden of the presumption of validity and the necessity to overcome it with clear and convincing evidence).

The third-party submission option is also problematic. Like protests, a third-party submission may be filed by any member of the public against a pending application. Submissions must be submitted within two months from the date of publication, or prior to the mailing of a notice of allowance, whichever comes first. Unlike protests, however, each third-party submission

16. *Id.* §§ 2209, 2609.

17. Opponents of the patent are usually those parties who are likely to be sued by the eventual patent holder.

is limited to ten patents and publications. *Curiously, explanations of the submissions are not allowed*.[18] Without explanations, the examiner must evaluate the submissions without the guidance of the third party who may be well versed in the art. Again, potential litigants are afraid to utilize this alternative, realizing that if the patent issues anyway, the patent holder will argue in later litigation that the reference was "before the examiner," and the examiner nevertheless allowed the claim. Furthermore, in the event a patent is allowed within the eighteen-month window prior to publication, this avenue never becomes available at all. Illustrating how seldom the third-party submission is used, for every 500 patent applications published in 2007, the PTO received only *one* such submission.[19]

Post-issuance, a third party may challenge a patent through the reexamination process.[20] This process allows a third party to attack an issued patent claim on the ground that it is invalid based on prior patents and publications. There are two types of reexamination: ex parte and inter partes. Ex parte reexamination can be requested by any party, including a third party. But other than a response to an initial statement by the patent owner addressing the request,[21] the reexamination will be limited to a dialogue between the patent owner and the PTO. There is *no* other opportunity for third-party input. By comparison, inter partes reexaminations, available since 1999, provide for *continued* examiner communication with the third party. Throughout the process, within thirty days of the patent owner's responses to the PTO, the third party is permitted to respond by written comment. Moreover, the third party may appeal an examiner's decision to the Board of Patent Appeals and Interferences, and later to the Federal Circuit. Significantly, however, the third party is estopped, in any later district court civil action, from asserting the invalidity of any claim on any ground that the he raised or could have raised during the proceeding. Unlike in court proceedings, however, the process does not provide for interrogatories, depositions, subpoenas, live testimonies, or witness cross-examinations. This is a significant downside: critics take issue with the estoppel aspect of inter partes reexamination.[22] Even the PTO has admitted that this is the most

18. *See* MPEP, *supra* note 14, § 1134.01(d) ("A submission under this section shall not include any explanation of the patents . . . [and] is also limited to ten total patents or publications.").

19. CTR. FOR PATENT INNOVATIONS, N.Y. LAW SCH., PEER-TO-PATENT: FIRST ANNIVERSARY REPORT 6 (2008), *available at* http://dotank.nyls.edu/communitypatent/P2Panniversaryreport.pdf.

20. MPEP, *supra* note 14, §§ 2209, 2609.

21. *Id.* § 2212.01.

22. *See, e.g.*, Joseph Farrell & Robert P. Merges, *Incentives to Challenge and Defend Patents:*

frequently identified inequity that deters third parties from filing requests for inter partes reexamination of patents.[23]

Ultimately, all sides should have a fair opportunity to be heard *before* issuance and *before* heels are dug in. Professor Jay Kesan of the University of Illinois has argued that from a psychological standpoint, a pre-grant opposition system makes sense.[24] Post-decisional cognitive dissonance refers to the idea that once a person (the examiner) is committed to an outcome, the amount of evidence needed to change the person's viewpoint is greater than if the evidence was being presented prior to a decision being made.[25] By allowing parties to offer evidence of unpatentability *before* a patent is issued, this problem could be reduced.[26]

♦ ♦ ♦

Although many reform proponents favor more fundamental improvements in the way patents are granted, I believe such reforms may be too contentious to be adopted in the foreseeable future. To take one example, if U.S. patent law was amended to allow full and fair notice and opportunity for public comment prior to allowance of any patent claim, then

---

Why Litigation Won't Reliably Fix Patent Office Errors and Why Administrative Patent Review Might Help, 19 BERKELEY TECH. L.J. 943, 967 (2004) ("This [reexamination estoppel] creates huge risks for challengers, who must trust that the USPTO will not make any mistakes in handling the reexamination. There is no opportunity to litigate the issue again in court. The broad consensus among patent experts is that these risks are too great."); Sherry M. Knowles et al., *Inter Partes Patent Reexamination in the United States*, 86 J. PAT. & TRADEMARK OFF. SOC'Y 611, 627 (2004) ("The *inter partes* reexamination procedure places so many constraints on third-party requesters of such reexamination that, as some patent attorneys have stated, 'It would be legal malpractice to recommend a client initiate an *inter partes* reexamination.' ").

23. U.S. PATENT AND TRADEMARK OFFICE, REPORT TO CONGRESS ON *INTER PARTES* REEXAMINATION 7 (2004), *available at* http://www.uspto.gov/web/offices/dcom/olia/reports/reexamreport.pdf.

24. *See* Jay P. Kesan, *Carrots and Sticks to Create a Better Patent System*, 17 BERKELEY TECH. L.J. 763, (2002); *see also* telephone interview with Jay P. Kesan, Professor of Law, Univ. of Illinois (Nov. 19, 2009). Most foreign countries have implemented a post-grant system. Kesan refers to empirical data from Germany and Japan to support his argument that a pre-grant system is more effective. Kesan, *supra*, at 781. Data shows that in Germany and Japan, two countries that have switched from a pre-grant to post-grant system, more patents are likely to be challenged in a pre-grant opposition setting. *Id.* When both countries had post-grant opposition systems in place, there was a decrease in opposition proceedings and increase in court initiated invalidation trials. *Id.* Kesan claims that "the perception that opponents are more likely to mount a successful challenge to a patent in a pre-grant system seems to have played a role in the decreasing number of challenges in the post-grant system." *Id.* at 782.

25. Kesan, *supra* note 24, at 780.

26. *Id.*

the usual rule favoring exhaustion of administrative remedies, if carried forward as well, would require at least all competitors within the cross-hairs of the claims to present all grounds for invalidity or other challenges to the PTO for consideration. And, in subsequent litigation, a court would be bound to uphold the agency determination so long as it was supported by substantial evidence or had a rational basis. Thus, while such a reform would improve the chances that the PTO would "get it right," such an amendment would seriously restrict the ability of the federal courts to set aside invalid patents. Such an amendment would make the PTO the almost exclusive arbiter of invalidity and relegate the courts to deciding issues of infringement and damages. This would, in turn, necessitate a large expansion in the PTO examiner staff and the resources available to them, a shift in resources that might be quite expensive. Certainly, it would slow down the timeline for obtaining a patent and invite competitors to throw up roadblocks against allowance. Those who profit from the status quo would surely lobby against such sweeping reforms. Perhaps most importantly, would American industry and congressional experts be comfortable giving up the safeguard offered by the federal courts as a check against invalid patents?

With or without more sweeping reforms, I return to my original, more modest proposal, one easily adaptable to the existing statutory framework: that Congress (1) reduce the standard for proving invalidity to a preponderance of the evidence, and (2) modulate the degree of deference to be accorded to PTO actions in accordance with the extent to which the examiner was afforded a reasonable opportunity to consider the specific question of invalidity, and the extent to which the examiner in fact reasonably addressed it. As stated, this will incentivize patent applicants and their counsel to lay bare the relevant prior art, to draw attention to the specific points of possible invalidity (while, of course, stating why allowance should nonetheless be granted), and thus enable an office action explaining how, if at all, the references do or do not restrict the proposed invention. This will strengthen deserving patents and winnow out at least some undeserving ones. Based on my immersion in many patent cases, I am convinced that this would discourage litigation based on weak patents without reducing the protection that we all wish to be accorded to deserving patents.