20:2 BERKELEY TECHNOLOGY LAW JOURNAL

Pages
985
to
1269

Spring
2005

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

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

Copyright © 2005 Regents of the University of California.
All Rights Reserved.

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

BERKELEY TECHNOLOGY LAW JOURNAL

VOLUME 20 NUMBER 2 SPRING 2005

TABLE OF CONTENTS

ARTICLES

PANEL: EVOLVING ANTITRUST TREATMENT OF DOMINANT FIRMS

DONORS

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

Benefactors ($25,000 and above)

COOLEY GODWARD LLP
San Francisco, CA

DLA PIPER RUDNICK GRAY CARY
Palo Alto, CA

FARELLA BRAUN + MARTEL LLP
San Francisco, CA

FENWICK & WEST LLP
Palo Alto, CA

HELLER EHRMAN WHITE & MCAULIFFE LLP
San Francisco, CA

LATHAM & WATKINS LLP
San Francisco, CA

MORRISON & FOERSTER LLP
San Francisco, CA

SKADDEN, ARPS, SLATE, MEAGHER & FLOM LLP
Palo Alto, CA

WEIL, GOTSHAL & MANGES LLP
Redwood Shores, CA

WILSON SONSINI GOODRICH & ROSATI PC
Palo Alto, CA

Members ($10,000 to $24,999)

ALSCHULER GROSSMAN STEIN & KAHAN LLP
Los Angeles, CA

BINGHAM MCCUTCHEN LLP
San Francisco, CA

COVINGTON & BURLING
San Francisco, CA

DAVIS POLK & WARDWELL
Menlo Park, CA

DAY CASEBEER MADRID & BATCHELDER LLP
Cupertino, CA

FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER LLP
Palo Alto, CA

FISH & RICHARDSON P.C.
Redwood City, CA

KIRKLAND & ELLIS
San Francisco, CA

KNOBBE, MARTENS, OLSON & BEAR, LLP
San Francisco, CA

MCDERMOTT, WILL & EMERY
Menlo Park, CA

MILBANK, TWEED, HADLEY & MCCLOY LLP
Palo Alto, CA

MORGAN, LEWIS & BOCKIUS LLP
San Francisco, CA

O'MELVENY & MYERS LLP
San Francisco, CA

ORRICK, HERRINGTON & SUTCLIFFE LLP
San Francisco, CA

PILLSBURY WINTHROP LLP
San Francisco, CA

TOWNSEND AND TOWNSEND AND CREW LLP
San Francisco, CA

WHITE & CASE LLP
Palo Alto, CA

Patrons ($5,000 to $9,999)

BAKER & MCKENZIE
Palo Alto, CA

DEWEY BALLANTINE LLP
Palo Alto, CA

HOWREY SIMON ARNOLD & WHITE LLP
Menlo Park, CA

KEKER & VAN NEST LLP
San Francisco, CA

KENYON & KENYON
San Jose, CA

MANATT, PHELPS & PHILLIPS LLP
Palo Alto, CA

MUNGER, TOLLES & OLSON LLP
San Francisco, CA

ROPES & GRAY LLP
Palo Alto, CA

VAN PELT & YI LLP
Cupertino, CA

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

ADVISORY BOARD

ROBERT C. BERRING, JR.
Walter Perry Johnson Professor of Law
Boalt Hall School of Law
Berkeley, California

ROGER BOROVOY
Fish & Richardson P.C.
Redwood City, California

JESSE H. CHOPER
Earl Warren Professor of Public Law
Boalt Hall School of Law
Berkeley, California

BRIAN C. CUNNINGHAM
Cooley Godward LLP
Palo Alto, California

MARK A. LEMLEY
Professor of Law and Faculty Scholar & Director of the Stanford Center for Law, Science & Technology
Stanford Law School
Palo Alto, California

REGIS MCKENNA
Chairman & CEO
Regis McKenna, Inc.
Palo Alto, California

PETER S. MENELL
Professor of Law & Executive Director of Berkeley Center for Law & Technology
Boalt Hall School of Law
Berkeley, California

ROBERT P. MERGES
Wilson Sonsini Goodrich & Rosati Professor of Law & Technology & Director of Berkeley Center for Law & Technology
Boalt Hall School of Law
Berkeley, California

JAMES POOLEY
Milbank, Tweed, Hadley & McCloy LLP
Palo Alto, California

MATTHEW D. POWERS
Weil, Gotshal & Manges LLP
Redwood Shores, California

DIANE WILKINS SAVAGE
Cooley Godward LLP
Palo Alto, California

LIONEL S. SOBEL
Professor of Law & Director of the International Entertainment & Media Law Summer Program in London, England
Southwestern University School of Law
Los Angeles, California

LARRY W. SONSINI
Wilson Sonsini Goodrich & Rosati
Palo Alto, California

MICHAEL TRAYNOR
Cooley Godward LLP
San Francisco, California

THOMAS F. VILLENEUVE
Gunderson, Dettmer, Stough, Villeneuve, Franklin & Hachigian, LLP
Menlo Park, California

SUBSCRIBER INFORMATION

The *Berkeley Technology Law Journal* (ISSN 1086-3818), a continuation of the *High Technology Law Journal* effective Volume 11, is edited and published four times each year (Spring, Summer, Fall, and Annual Review of Law and Technology) by the students of Boalt Hall School of Law, University of California, Berkeley. Application to Mail at Periodicals Postage Rate is Pending at Berkeley, California, and at additional mailing offices. POSTMASTER: Send address changes to Journal Publications Coordinator, Boalt Hall School of Law, 421 North Addition, University of California, Berkeley, CA 94720-7200.

Correspondence. Address all correspondence regarding subscriptions, address changes, claims for nonreceipt, single copies, advertising, and permission to reprint to Journal Publications Coordinator, Boalt Hall School of Law, 421 North Addition, Berkeley, CA 94720-7200; (510) 643-6600; journalpublications@law.berkeley.edu. Authors: see section entitled Information for Authors.

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

Form. The text and citations in the *Journal* conform generally to the UNITED STATES GOVERNMENT PRINTING OFFICE STYLE MANUAL (28th ed. 1984) and to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 17th ed. 2000). Please cite this issue of the *Berkeley Technology Law Journal* as 20 BERKELEY TECH. L.J. ___ (2005).

BTLJ ONLINE

Abstracts of all *Berkeley Technology Law Journal* and *High Technology Law Journal* articles as well as the full text of most articles published in previous issues can be found at http://www.btlj.org. Our site also contains subject, author and title indexes, general information about the *Journal*, selected materials related to technology law, and links to other related home pages. Subject, author and title indexes may also be found in Volume 10, Number 2 (1995) of the *Journal*.

INFORMATION FOR AUTHORS

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

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

Citations. All citations should conform to THE BLUEBOOK: A UNIFORM SYSTEM OF CITATION (Columbia Law Review Ass'n et al. eds., 17th ed. 2000). In addition, the author should include his or her credentials, including full name, degrees earned, academic or professional affiliations, and citations to all previously published legal articles.

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

Mailing Address. Please submit all hardcopy manuscripts to:

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

BOARD OF EDITORS 2004-2005

Editor-in-Chief
TITI NGUYEN

Managing Editor
JOSEPH V. MARRA

Senior Article Editors
JULIETA LERNER
MARC SHARP

Senior Executive Editor
MATTHEW C. HOLOHAN

Senior Annual Review Editors
BRIAN PAUL GEARING
ELIZABETH MILES

Submissions Editors
ANDREA FREEMAN
AARON PERZANOWSKI

Production Editor
ASSAD RAJANI

Symposium Editors
ANNA LEE
GEMMA SUH

Article Editors

TOMOMI HARKEY
CHARLES HUSE
CAROL JOHNS
OLGA V. KOTLYAREVSKAYA
SUNNY LU
KATIE NOLAN-STEVAUX
SHERWIN SIY
KATHERINE A. OYAMA
AARON PERZANOWSKI
RICHARD R. RONALD
KARA WARNER
ANNE WOOD

Executive Editors

DANIEL CASTRO
DEAN CHELEY
EUGENE CHOO
J. RYAN GILFOIL
TERESA HUANG
ANDREW LEE
JENNIFER LYNCH
LON SORENSEN
RANGANATH SUDARSHAN
TARA WHEATLAND

Annual Review Advisors

DAVID ALBAN
LARISSA BURFORD
STEPHEN BURDICK
MICHELLE CAI
AMANDA COHEN
GEORGE FOX
BRIAN PAUL GEARING
DANIEL HIGGS
JAE HONG LEE
ELIZABETH MILES
RICHARD R. RONALD
MARC SHARP
FELIX WU

MEMBERSHIP 2004-2005

Associate Editors

TIMOTHY P. BEST
PETER COLOSI
JAMEEL HARB
DAVID JANG
ERIN C. JONES
MICHAEL PASAHOW*
TORI PRAUL
DAVID SANKER
SARA TERHEGGEN
M. LEAH VELAZQUEZ
VIRGINIA ZAUNBRECHER
ANNA ZICHTERMAN*

Members

SARA ADIBISEDEH
CHRISTOPHER BALDWIN
JOCELYN BELLONI
JOSH BENSON
YONATAN E. BRAUDE
JAMIE CALLOWAY
BRIAN W. CARVER
CYNTHIA CHEN
ADRIENNE CHENG
STEPHEN DANG
WES DERRICK
CORRIE DRAKULICH
CATHY DUBOC
CARRIE DUONG
JOHN R. EASTBURG
DIANA FU
ALICIA GAMEZ
DOV GREENBAUM
MISTY S. HALSTEAD
GALEN HANCOCK
THOMAS KING
MARK LEZAMA
EMILY LUDMIR
REBECCA LYNN
KEVIN MACK
PETER MAYBARDUK
NARAYAN MELGIRI
CHRISTINE MOKHTARIAN
NIKI MOORE
FARBOD MORIDANI
SANDRA L. OLIVARES
ERIC OW
SARA POSTER
MICHAEL SCHLEMMER
LEWIS STILLER*
JOCELYN SUM
JOSEPH TELTSER
CECILIA WALSH
ALISON WATKINS
CHRISTOPHER WESTBERG
NICHOLAS WHILT
MARTIN WHITE
TASHICA WILLIAMS
BETHELWEL WILSON
MATTHEW WISE
DENISE USTARIZ
HAMADA ZAHAWI
LILY ZHANG

* Denotes Recipient of Outstanding Member Award

A FALLACY OF THE COMMONS IN BIOTECH PATENT POLICY

By David E. Adelman[†]

TABLE OF CONTENTS

Fears that the recent proliferation of biotech patents is undermining scientific norms and threatening innovation dominate the debate over biotech patent policy. Chief among these issues are protection of the public information commons and concerns about emerging "patent anticommons."[1] Yet, legal commentators have been surprisingly indifferent to whether the traditional model of the public commons accurately reflects the conditions of innovation in the biomedical sciences. This omission proves to be a critical one, for it obscures a central fallacy—that the com-

---

© 2005 David E. Adelman

† Associate Professor, Rogers College of Law, University of Arizona. For their helpful suggestions and advice, I would like to thank Graeme Austin, John Barton, Mark Lemley, Greg Mandel, Toni Massaro, Carol Rose, and Ted Schneyer; any errors are of course my own.

1. The standard model for the public commons is an area that is vulnerable to overexploitation by multiple actors because none of them bears the full impact of poor management. Examples include public lands and bodies of water. By contrast, a patent anticommons impedes development because narrow patent rights are dispersed among different entities too broadly, creating conditions under which no single entity has access to the technology needed to conduct research and development. *See* Rebecca S. Eisenberg, *A Technology Policy Perspective on the NIH Gene Patenting Controversy*, 55 U. PITT. L. REV. 633, 640 (1994); Michael A. Heller & Rebecca S. Eisenberg, *Can Patents Deter Innovation? The Anticommons in Biomedical Research*, 280 SCIENCE 698, 698 (1998).

mons for biomedical science is finite and congested. This often implicit presumption is contradicted by the overabundance of research opportunities created by recent advances in genomics (and other biotech fields), which have transformed biomedical science into an unbounded resource. The uniquely open-ended nature of biomedical science requires a reassessment of how patenting affects biotech research and innovation. Contrary to most legal scholarship, this reappraisal leads to the conclusion that the threats to biomedical innovation posed by biotech patenting are generally modest.

The principal gap in the literature on biotech patent policy centers on legal scholars' failure to examine the interplay between the underlying science and patent policy. Indeed, the science at the heart of the biotech revolution is conspicuously absent from the current debate over biotech patent policy. To the extent that science is considered, it is either filtered through an economic lens or treated generically.[2] Typically, unique features of biotech science are important only insofar as they affect the dynamics of innovation, such as whether biotechnology evolves discretely or cumulatively.[3] More often, legal commentators have focused their attention, often quite understandably, on the protection of scientific norms, such as communalism and free access to data, which are even further removed from the science itself.[4] Little, if any, of this discourse considers how the practical limits, specific research tools, and technical details of biomedical science shape patent strategy and incentives.[5]

The hyperbole surrounding advances in biotechnology, particularly in genomics and other "omic" sciences, contributes to the superficial treatment of biotech science in the legal policy debate. Overly optimistic claims obscure the technical barriers and experimental uncertainties that

---

2. *See, e.g.*, Robert P. Merges & Richard L. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 880 (1990).

3. *Id.* at 880-84. In particular, while Merges and Nelson argue that issues of patent policy, such as patent scope, "depend on the nature of the technology," they limit their consideration to "the relationship between technical advances in the industry, and the extent to which firms license technologies to each other." *Id.* at 843.

4. Arti K. Rai, *Regulating Scientific Research: Intellectual Property Rights and the Norms of Science*, 94 NW. U. L. REV. 77, 90-92 (1999); Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 289 (2003).

5. Economic data are similarly missing from the debate, although recent studies are beginning to have an impact. *See, e.g.*, John P. Walsh et al., *Effects of Research Tool Patents and Licensing on Biomedical Innovation*, *in* PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285 (Wesley M. Cohen & Stephen A. Merrell eds., 2003).

continue to plague biotech research and development.[6] Most importantly, this rosy vision hides the disparity that exists between the power of biotech methods to generate data, such as genome sequences, and to produce effective medical procedures and drugs. Research employing biotech methods has produced vast quantities of genetic data, which are often useful research tools (for example, drug targets and genetic probes), but the translation of this knowledge into new products has been far less impressive.[7] This dichotomy, when layered onto the complexity of biological processes themselves, has created an environment in which research opportunities far exceed the capacities of the scientific community. It is this basic dynamic that makes biotech science, in important respects, an effectively unbounded, uncongested common resource.[8]

The debate over biotech patent policy has favored economic theorizing that overlooks the defining characteristics of biomedical science that influence biotech patenting. Legal commentators have proposed numerous theories on patent scope and strategies for mitigating the negative impacts of patenting on biomedical innovation.[9] Contrasting proposals range from

---

6. The significance of the scientific barriers should not be underestimated, and is best illustrated by the declining rate of new drug development over the past decade, despite increased spending (and patenting) by the public and private sectors. *See* FOOD & DRUG ADMIN., INNOVATION OR STAGNATION: CHALLENGE AND OPPORTUNITY ON THE CRITICAL PATH TO NEW MEDICAL PRODUCTS 2 (2004), *available at* http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html; Richard S. Cooper & Bruce M. Psaty, *Genomics and Medicine: Distraction, Incremental Progress, or the Dawn of a New Age?*, 138 ANNALS INTERNAL MED. 576, 577 (2003); Robert F. Service, *Surviving the Blockbuster Syndrome*, 303 SCIENCE 1796, 1799 (2004) ("The plain truth is that many of the most dramatic scientific advances that have recently been made in the lab have not transformed medical care." (quoting FDA Commissioner Mark McClellan)).

7. Jenifer Couzin, *NIH Dives into Drug Discovery*, 302 SCIENCE 218, 219 (2003) (describing how drug companies are "barely dipping their toes into the 'data dump' that is the human genome sequence"); Service, *supra* note 6, at 1799.

8. In other contexts, particularly environmental regulation, property theorists have recognized that commons problems do not emerge until a commons becomes "congested," that is, the number of users rises beyond the point of sustainable exploitation of the resource. *See* Carol Rose, *Rethinking Environmental Controls: Management Strategies for Common Resources*, 1991 DUKE L.J. 1, 5-7. The distinction I make here is a simple variation on this basic insight, with the proliferation of patents restricting access to intellectual resources taking the place of the mounting numbers of resource extractors in the typical tragedy of the commons scenario.

9. *See, e.g.*, John H. Barton, *Non-Obviousness*, 43 IDEA 475 (2003); Rebecca S. Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1040-41 (1989); Maureen A. O'Rourke, *Toward A Doctrine of Fair Use in Patent Law*, 100 COLUM. L. REV. 1177 (2000); Arti K. Rai, *Engaging the Facts and Policy: A Multi-Institutional Approach to Patent System Reform*, 103 COLUM.

arguments that biotech patents should have narrow scope[10] to claims that federal agencies, such as the National Institutes of Health (NIH), should be empowered to protect biotech innovation from incipient patent anti-commons.[11] One of the more provocative proposals has argued for an eclectic approach premised on a technology-specific synthesis of patent policy, which maintains that biotech patents should be both broader and fewer in number.[12]

The simple economic narratives of the legal theories on biotech patenting starkly contrast with the conflicting messages of empirical studies. For example, contrary to the fears of many legal commentators, there are few signs that biotech patenting has impeded biomedical innovation.[13] Other studies have found that the relationship between patent policy and innovation is more nuanced and less predictable than anticipated.[14] On the other hand, some evidence suggests that broad patents on upstream discoveries have the potential to impede research.[15] In short, few of the predictions made or the solutions advocated by legal scholars are borne out consistently by empirical studies of biotech patenting. I argue that the dynamics created by the uncongested, open-ended status of biomedical science account for much of this divergence.

This Article proceeds in three parts. Part I provides a brief overview of the current debate over patent scope and evaluates the available data on biotech patenting. Part II explains the central features of biomedical science that are relevant to patent policy, paying particular attention to the roles of important research tools. I identify two classes of research tools,

---

L. REV. 1035 (2003); Lawrence M. Sung, *On Treating Past as Prologue*, 2001 U. ILL. J.L. TECH. & POL'Y 75.

10. Merges & Nelson, *supra* note 2, at 915.

11. Heller & Eisenberg, *supra* note 1, at 698.

12. Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1202-03 (2002) [hereinafter Burk & Lemley, *Technology-Specific*]; Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575 (2003) [hereinafter Burk & Lemley, *Policy Levers*].

13. Walsh et al., *supra* note 5, at 289, 331.

14. BRONWYN H. HALL, BUSINESS METHOD PATENTS, INNOVATION, AND POLICY (Dep't of Econ., Univ. of Cal., Berkeley, Working Paper No. E03-331, 2003), *available at* http://repositories.cdlib.org/iber/econ/E03-331. Professor R. Polk Wagner has also noted how the complex relationships between, and among, patent law doctrines creates indeterminacies in how they affect patent scope and incentives. R. Polk Wagner, *Of Patents and Path Dependency: A Comment on Burk and Lemley*, 18 BERKELEY TECH. L.J. 1341, 1342-43 (2003).

15. Walsh et al., *supra* note 5, at 331.

common-method and problem-specific,[16] and conclude that only the patents on common-method tools pose potential risks to biomedical innovation. Part III reexamines the competing legal proposals and argues that scholars misapply the standard finite-commons model because they fail to recognize how the complexity of biological processes and the power of existing biotech methods to produce genetic data make biomedical science, in crucial respects, an unbounded and uncongested common resource. Taken together, these findings imply that the potential adverse effects of biotech patenting are both qualitatively different from and often less significant than the effects many legal commentators have predicted.

I. PATENT POLICY AND ECONOMICS

Two central factors have combined to put biotech patents in the spotlight. First, the Bayh-Dole Act of 1980[17] arguably constitutes the single most important event, as it expanded both the range of entities patenting inventions and the types of inventions being patented.[18] Following passage of the Bayh-Dole Act, universities and research institutes increased their patent filings dramatically, further blurring the line between commercial and basic-science research.[19] This increase in university patenting has been accompanied by a steep rise in the patenting of basic-science research tools ("upstream technologies") that are integral to a broad cross-section of biotech research.[20] Second, rapid scientific developments during the 1980s and 1990s led to large influxes of private funding for biomedical research, outpacing government funding for the first time in 1992, and

16. Common-method research tools involve uniquely powerful methods of broad applicability (for example, the polymerase chain reaction (PCR), which is used to replicate DNA), whereas problem-specific research tools involve data or information that are of narrow applicability and available in many forms (for example, drug targets and gene probes).

17. 35 U.S.C. §§ 200-212 (2000).

18. Rebecca S. Eisenberg, *Bargaining over the Transfer of Proprietary Research Tools: Is This Market Failing or Emerging?*, *in* EXPANDING THE BOUNDARIES OF INTELLECTUAL PROPERTY: INNOVATION POLICY FOR THE KNOWLEDGE SOCIETY 226-27 (Rochelle Cooper Dreyfuss et al. eds., 2001) [hereinafter EXPANDING THE BOUNDARIES].

19. Walsh et al., *supra* note 5, at 295 (noting that the trend is viewed as a major change by both universities and the private sector).

20. NAT. ACADS. OF SCI., A PATENT SYSTEM FOR THE 21ST CENTURY 17, 20 (Stephen A. Merrill et al. eds., 2004) [hereinafter NAS REPORT]; Walsh et. al., *supra* note 5, at 295. Archetype examples of upstream technologies are the famous Cohen-Boyer patent, which covered the canonical methods for replicating and expressing foreign genes in microorganisms, and the PCR for copying DNA. Eisenberg, *supra* note 18, at 229-30.

continuing to do so today.[21] These trends have transformed the biomedical sciences, and have brought private and public sector research closer together and occasionally into conflict.[22]

Legal commentators have responded to these developments with a panoply of proposals and concerns.[23] They have paid particular attention to the increased patenting of research tools and the rapid growth in the number of biotech patents, both of which have the potential to impede innovation and research.[24] Broadly speaking, legal commentators are separable into two camps, one optimistic and the other pessimistic, about whether licensing and other market agreements can resolve these tensions. The optimists appeal to experience in well-established industries (for example, electronics and automobiles) to argue that the market will work out any tensions between patents and scientific progress.[25] The pessimists typically focus on anecdotal evidence and other incipient signs that aggressive patenting is threatening biomedical research and development.[26]

This Part discusses the legal and economic elements of the debate over biotech patenting. Section A provides an overview of the competing legal proposals and perspectives. Section B discusses the available economic information and recent survey data on biotech patenting. The first two sections expose the many differences between the legal perspectives and empirical observations. The sources of this divergence are discussed in Section C.

A. The Legal Debate over Patent Scope

Legal commentators tend to advocate one of three approaches to patent policy: (1) a traditional law-and-economics approach, emphasizing bright-line rules and market roles; (2) an agency-based approach, relying on experts to intervene when necessary to overcome market failures or to protect scientific norms; or (3) a judicial activist model, relying on so-called

---

21. Eisenberg, *supra* note 18, at 227 n.15.

22. NAS REPORT, *supra* note 20, at 20-21; Eisenberg, *supra* note 18, at 225-28; Walsh et al., *supra* note 5, at 29.

23. *See* Burk & Lemley, *Policy Levers*, *supra* note 12, at 1596-614.

24. Eisenberg, *supra* note 18, at 230-31 (suggesting that transaction costs of negotiating technology licenses are a significant problem in biomedical sciences); Robert P. Merges, *Institutions for Intellectual Property Transactions: The Case of Patent Pools*, *in* EXPANDING THE BOUNDARIES, *supra* note 18, at 129 ("[T]he key issue is the cost of integrating disparate rights.").

25. F. Scott Kieff, *Property Rights and Property Rules for Commercializing Inventions*, 85 MINN. L. REV. 697, 719-22 (2001); *see, e.g.*, Merges, *supra* note 24, at 130.

26. Kieff, *supra* note 25, at 719-22; *see, e.g.*, Merges, *supra* note 24, at 130.

patent "policy levers" latent in existing legal doctrines.[27] Notably absent is a legislative approach, which commentators widely agree would fall prey to undue public choice pressures from specific industry interests.[28] Each of the three competing views is discussed below.

Market-oriented commentators are unmoved by concerns about patent scope. They start from the premise that "there is no easily-defined 'ideal' menu of property rights for a given economy at a given moment in time."[29] They argue that determining an optimal level of patent protection is precluded by our lack of reliable metrics for analysis.[30] This belief leads them to reject legislative reform and judicial intervention as potential solutions and to embrace a market solution.[31] Their logic runs as follows: "Complex questions lack right answers. When there is no right answer—and when people bear the costs of their actions—we rely on those affected to make their own decisions."[32] Therefore, market-oriented commentators prefer clear rules, strong property rights, and institutions that promote negotiations (by reducing transaction costs) to solve property rights problems.[33] Moreover, they claim that recent technical advances reduce trans-

27. The Federal Circuit, for its part, has increasingly opted for bright-line rules and legal formalism over discretionary standards to promote clarity and predictability. *See* Burk & Lemley, *Policy Levers*, *supra* note 12, at 1672-73.

28. *Id.* at 1578, 1631-38; Rai, *supra* note 9, at 1128-30.

29. Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577, 588 (1999). Strong market skeptics, like Judge Frank Easterbrook, contend that refining existing patent laws creates more risks than potential benefits because (1) we know too little about the effect of the existing patent regime on innovation and (2) solutions developed in the dark are much more likely to be harmful than helpful. Frank H. Easterbrook, *Who Decides the Extent of Rights in Intellectual Property*, *in* EXPANDING THE BOUNDARIES, *supra* note 18, at 405-06.

30. *See, e.g.*, Easterbrook, *supra* note 29, at 406. As Judge Easterbrook has put it, "If firms that put millions of dollars on the line cannot make reliable decisions about technology, what would make us think that scholars [or policy makers] with *no* money on the line do well at devising legal rules to govern technology?" *Id.* at 408.

31. *See, e.g.*, *id.* at 408-11 (arguing that legislators fail for public choice reasons; that judges are unreliable because they are not accountable for their decisions and lack the necessary technical training and knowledge; and that courts create incoherent strategy because of limited time, differing perspectives, and uneven knowledge).

32. *Id.* at 411. This is just a restatement of the Coase Theorem: "[T]he more complex the problem, the more the 'right' answer varies over time and the affected population; and the easier it is to address the problem by private contract, the less we should attempt to resolve it by law." *Id.*

33. *See, e.g.*, *id.* at 411-13.

action costs already, making their standard Coasian approach all the more relevant.[34]

Modified market-oriented approaches also exist that endorse a competitive model for maximizing innovation. Robert P. Merges and Richard L. Nelson have led the move to reject prospect theory, which maintains that the central role of patenting is to allow for coordinated, rational development of a patent prospect (that is, field or area of invention), as opposed to providing ex ante incentives to invent.[35] Merges and Nelson instead assert that competition should be favored over prospect theory's reliance on broad pioneer patents in most technological areas.[36] They argue that prospect theory runs awry because it misstates the problem:

> But with the technological 'prospects' . . . no one knows for sure what possible inventions are in the technological pool. . . . Because of this uncertainty, development of technology is critically different from other common pool problems. The real problem is not controlling overfishing, but preventing underfishing after exclusive rights have been granted. The only way to find out what works and what does not is to let a variety of minds try.[37]

Accordingly, competition spurs innovation much more effectively than monopolies because it allows multiple inventors to work simultaneously and because use of an idea, unlike a traditional resource prospect, is not mutually exclusive.[38]

This modified market-oriented theory, unlike its counterpart, anticipates an active role for judges and, arguably, the Patent and Trademark Office (PTO). Judges should be guided by the modified theory's narrow-patent principle when applying patent doctrines, such as nonobviousness

---

34. *See, e.g.*, *id.*

35. Merges & Nelson, *supra* note 2, at 843 (arguing that patent scope policy is highly dependent on the nature of the technology with "[(1)] the relationship between technical advances in the industry, and [(2)] the extent to which firms license technologies to each other").

36. *Id.* at 843-44; *see* Walsh et al., *supra* note 5, at 291 n.11 ("[It is] well recognized in the economics of innovation [] that, given a technological objective (*e.g.*, curing a disease) and uncertainty about the best way to attain it, that objective will be most effectively achieved to the extent that a greater number of approaches to it are pursued.").

37. Merges & Nelson, *supra* note 2, at 873.

38. Merges and Nelson single out "science-based" technologies, such as biotech, in this regard, arguing that "the patent system should be particularly careful in awarding patents of broad scope" in such areas. *Id.* at 884. Their fear is that "a real danger [exists] that allowing patent scope to be overbroad may enable the individual or firm who first came up with a particular practical application to control a broad array of improvements and applications." *Id.*

and patent disclosure requirements. Merges and Nelson refine their theory further by establishing categories of inventive activity, which are governed by distinct rules for applying patent doctrines and optimizing patent scope.[39] They argue, for example, that judges should limit patent scope and apply the doctrine of equivalents narrowly for cumulative and science-based technologies.[40] Consistent with their basic insight, the greater the scientific uncertainty and complexity, the more innovation will benefit from a diversity of approaches and the competition engendered by numerous narrow patents.[41]

The anticommons theory of Michael A. Heller and Rebecca S. Eisenberg challenges these competitive market models.[42] A patent anticommons threatens to emerge when patent rights become so fragmented that no single scientist or company can access the technology necessary to conduct research in their field.[43] Heller and Eisenberg expose a major defect in the view that narrow patents necessarily promote competition and a broad range of inventive activity. The two demonstrate that, while this may be true, it is valid only up to the point at which fragmentation of patent rights causes the transaction costs associated with obtaining multiple technology licenses (for example, search costs and negotiations) to limit access to needed technologies.[44] Heller and Eisenberg argue that the risk of a patent anticommons emerging is exacerbated by individual cognitive biases and cultural differences between private and public sector researchers that impede negotiations.[45] They conclude that these tendencies are most acute in biotech patenting because the inventions at issue, particularly research tools such as drug targets, are both costly to identify and initially very difficult to value.[46]

---

39. Their approach also maintains that several different models of technological invention exist: (1) discrete invention for inventions like the safety razor, some pharmaceuticals, and toys, which are relatively insensitive to patent scope; (2) cumulative technologies, such as those associated with aircraft, automobiles, lights, semiconductors, and computers, which are highly sensitive to patent scope; (3) chemical technologies, particularly manufacturing processes, which are similar to cumulative technologies; and (4) science-based technologies, such as biotechnology, which tend to be a blend of the other categories and reliant on unpredictable factors. *Id.* at 880-83.

40. *Id.* at 906-11, 915.

41. *Id.* at 915.

42. Heller & Eisenberg, *supra* note 1, at 698.

43. *Id.*

44. *Id.* at 698-99.

45. *Id.* at 700.

46. *Id.* at 698. Eisenberg also raises important issues that seem to derive from the high uncertainty of the science and the huge windfalls that can accrue from the research. Eisenberg, *supra* note 18, at 225, 229, 248-49; *see also* REPORT OF THE NATIONAL

The threat of a patent anticommons has inspired numerous proposals. Arti K. Rai and Rebecca S. Eisenberg argue that federal grant-making agencies, such as the NIH, should be empowered to abrogate patent rights when an anticommons threatens innovation and research.[47] The technical complexities and scientific uncertainties of biotechnology lead Eisenberg to conclude that it is better to "leave[] more discretion for agencies to determine on a case-by-case basis whether the public interest would be better served by obtaining patent protection or by leaving certain basic research discoveries in the public domain."[48] In effect, they opt for a quasi-Kitchian prospect theory approach that replaces private entities with a federal agency to coordinate research and development. Complementing this agency-based model, Eisenberg, Rai, and others have argued for an expanded research exemption to patent infringement to mitigate further the adverse effects of patent anticommons.[49]

In contrast, Dan L. Burk and Mark A. Lemley propose a broad policy synthesis that is as expansive as it is provocative. Their proposal stands out because, among other elements, it relies on judges actively calibrating patent doctrines, which they refer to as "policy levers," on a technology-specific basis.[50] Drawing on the earlier work of Merges and Nelson, they make the case that patent policy ought to be technology specific—after all,

---

INSTITUTES OF HEALTH (NIH) WORKING GROUP ON RESEARCH TOOLS (1998), *available at* http://www.nih.gov/news/researchtools.

47. Rai & Eisenberg, *supra* note 4, at 310-13 (suggesting recommendations for NIH action and modifications to the Bayh-Dole Act to avoid problems of overly aggressive patenting of research tools). Rai has also argued that greater deference to the PTO would improve biotech jurisprudence. *See* Arti K. Rai, *Genetic Technology: Social Values and Personal Autonomy in the 21st Century*, 34 WAKE FOREST L. REV. 827, 828 (1999). More generally, Rai has proposed a number of broad institutional and procedural changes within the PTO and the courts. Rai, *supra* note 9, at 1041, 1075-78, 1091, 1097-101, 1124-25.

48. Eisenberg, *supra* note 1, at 640.

49. Rai, *supra* note 9, at 1041. Taking a slightly different tactic, Rochelle Dreyfuss has suggested an elegant refinement to the research exemption by tying it to a commitment that any inventions generated by the research will not be patented. Rochelle C. Dreyfuss, Varying the Course in Patenting Genetic Material: A Counter-Proposal to Richard Epstein's Steady Course 8-11 (Apr. 2003), *available at* http://ssrn.com/abstract=394000. More specifically, Dreyfuss provides that a university or nonprofit research institute would be immune from prosecution for using patented materials in research if (1) "those materials were not available on reasonable terms" (Nelson amendment), and (2) "the university or other research organization agrees not to patent anything that came out of the research." Richard R. Nelson, *The Market Economy and the Scientific Commons* 27-28 (Nov. 2003) (restating and adding to Dreyfuss's plan), *available at* http://ideas.repec.org/p/ssa/lemwps/2003-24.html.

50. Burk & Lemley, *Policy Levers*, *supra* note 12, at 1578-79.

research and development costs, patent incentives, and rates of technological progress vary greatly across industries.[51] Therefore, making the natural inference from these observations, they conclude that legal incentives must be optimized to promote innovation on an industry-by-industry basis.[52] Burk and Lemley marshal further support by pointing out that the Federal Circuit already applies legal doctrines in a technology-specific manner.[53]

Having made a case for technology-specific patent law, Burk and Lemley develop a systematic program for calibrating various patent doctrines, or policy levers, for specific technologies.[54] Incorporating prospect theory, Merges and Nelson's competition-based model, and the Heller and Eisenberg anticommons theory, Burk and Lemley describe several policy lever regimes and link them to specific technologies.[55] Thus, for example, biotechnology raises issues that span prospect theory (broad patents) and anticommons problems (too many patents), resulting in a policy favoring a relatively small number of broad patents. To achieve this, Burk and Lemley propose that the patentability standard of "nonobviousness" be raised and the written description required of an invention be reduced for patentability purposes.[56]

The evolution of more than a decade of patent policy debate is brought almost full circle in a short article by R. Polk Wagner.[57] Wagner accepts that differences exist in the application of patent law to distinct areas of

---

51. *Id.* at 1581-86, 1588-89.

52. *Id.* at 1588-89.

53. *Id.* Again biotech patents are showcased: in this case, for the notably weak obviousness standard and exceptionally stringent written description requirements the Federal Circuit has applied to biotech patents. Burk and Lemley also argue that such judge-based policies are unavoidable, as failure to articulate a clear policy is a policy decision itself. *Id.* at 1669, 1678.

54. Policy levers are context-specific standards, and include the rule against patenting abstract ideas, utility, experimental use, level of skill in the art, secondary considerations for obviousness, written description, reasonable interchangeability, pioneer patents, reverse doctrine of equivalents, presumption of validity, new secondary considerations, patent misuse, and injunctions. *Id.* at 1641-68.

55. In an earlier article, Burk and Lemley also argue that the person having ordinary skill in the art (PHOSITA) for obviousness and disclosure requirements should be "decoupled" because of the reciprocal relationship between obviousness and enablement. Burk & Lemley, *Technology-Specific*, *supra* note 12, at 1202-03.

56. This approach follows directly from their inferences that patent scope is directly proportional to the standard for obviousness, but indirectly proportional to the enablement and written description requirements. Burk & Lemley, *Policy Levers*, *supra* note 12, at 1593-94.

57. Wagner, *supra* note 14, at 1341.

technological development, but he questions whether these differences align with the technologies ("macro-specificity") as opposed to the substantive differences ("micro-specificity") within each technology.[58] Wagner asserts that the inter-technology differences that Burk and Lemley identify can be explained by micro-specificity.[59] Of equal importance, Wagner uses several elegant logical arguments to demonstrate that the Burk and Lemley synthesis fails because the application of their policy levers is hopelessly indeterminate.[60] He then returns to traditional economic theory, arguing for clearer, more stable patent doctrine and less, not more, judicial discretion.[61]

It is worth briefly stepping back to take stock of the legal policy arguments. Prospect theory argues for broad patents because, by granting control of a technology, they ensure efficient coordination of innovation. Standard economic theory maintains that patent scope is secondary to clear rules, strong property rights, and low transaction costs. However, Merges and Nelson's modified market approach is premised on innovation being proportional to the number of independent inventors, which entails narrow patents. Qualifying this theory, Heller and Eisenberg argue that patent rights can become too fragmented, creating an anticommons to which no one has ready access; consequently, they call for intervention by experts in federal agencies. Finally, the Burk and Lemley synthesis maintains that "everyone is right," though only for specific technologies, and relies on judges to manage patent policy on a technology-specific basis. Stated succinctly, these proposals reduce to: (1) more is more (broad patents preferred); (2) less is more (narrow patents preferred); (3) more is sometimes more and sometimes less; and (4) neither (patent scope is irrelevant). Although none of these proposals has prevailed, their logical completeness is gratifying.

---

58. *Id.* at 1343-44.

59. *Id.* at 1347.

60. *Id.* at 1348-49. This indeterminacy arises because disclosure requirements and the obviousness standard each affect the scope of a patent and, at the same time, are linked through the legal construct of the PHOSITA. Further, both are indirectly proportional to the scope of a patent, but patent disclosure is directly proportional to the skill level of the PHOSITA, whereas obviousness is indirectly proportional. *Id.* at 1348. These relationships create tradeoffs that nullify the simple relationships Burk and Lemley describe. In an earlier article, however, Burk and Lemley do advocate decoupling the PHOSITA standard in the disclosure and obviousness regimes. Burk & Lemley, *Technology-Specific*, *supra* note 12, at 1202-05.

61. Wagner, *supra* note 14, at 1358-60. Indeed, Wagner persuasively argues that Burk and Lemley's many criticisms of Federal Circuit doctrine demonstrate that there is little reason to have confidence that judges will make the right judgments about aligning legal doctrine with good policy. *Id.* at 1359.

B. Biotech Patent Data and Qualitative Surveys

Studies of patenting in the biomedical sciences remain very limited. Nonetheless, we do know that there has been both a striking increase in the number of patents on research tools and a significant rise in defensive patenting, particularly in the genomic sciences.[62] However, recent data suggest that the surge has slowed and that biotech patent applications may be declining.[63] In addition, we know that university patenting accounts for a significant fraction of this increase in patenting. For example, universities' share of the patents issued in three key biomedical utility classes increased from eight to twenty-five percent between the early 1970s and the mid-1990s.[64] Thus, the limited data available on patenting in the biotech sector lend support to concerns about emerging patent anticommons.

Anecdotal evidence lends support to the existence of an anticommons in biotech patenting as well. One of the most publicized and debated cases from the biotech sector has involved efforts to reduce the incidence of blindness due to vitamin A deficiency by genetically modifying rice to produce vitamin A.[65] In order to undertake this research, researchers had to negotiate licenses to seventy patents and obtain access to fifteen pieces of technical property spread over thirty-one institutions.[66] Although ultimately resolved through a collective set of agreements for royalty-free licenses, this case has become the poster child for critics who have been critical of expansive biotech patenting and its potential to impede research.[67]

Similar examples have been identified in more traditional areas of the biomedical sciences. Two legal commentators identified one hundred patents related to the andrenergic receptor, which is important in metabolic

---

62. NAS REPORT, *supra* note 20, at 10; Walsh et al., *supra* note 5, at 295. Consistent with Walsh et al., "research tools" will be defined broadly to include "any tangible or informational input into the process of discovering a drug or any other medical therapy or method of diagnosing disease." Walsh et al., *supra* note 5, at 287.

63. *See, e.g.*, FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP, BIOTECHNOLOGY INNOVATION REPORT 2004, at 8-9 (Arie M. Michelsohn ed., 2004), *available at* http://www7.nationalacademies.org/step/STEP_Projects_Proteomics.html.

64. Walsh et al., *supra* note 5, at 295.

65. *See, e.g.*, Richard C. Atkinson et al., *Public Sector Collaboration for Agricultural IP Management; Intellectual Property Rights*, 299 SCIENCE 174, 174 (2003); Roger N. Beachy, *IP Policies and Serving the Public*, 299 SCIENCE 473, 473 (2003); Ronald L. Phillips, *Intellectual Property Rights for the Public Good: Obligations of U.S. Universities to Developing Countries*, 6 MINN. J.L. SCI. & TECH. 177, 181 (2004).

66. Walsh et al., *supra* note 5, at 288 n.6.

67. *Id.* at 298.

pathways.[68] Fourteen patents controlled by several organizations cover the hepatitis-B vaccine; thus, stacking royalties totaling $1.47 per dose, or thirteen to fifteen percent of sales, add to the cost of the vaccine's production.[69] Concern also persists about patents that restrict access to critical drug targets (for example, receptors and mutated genes) or biotech techniques (for example, genechips and diagnostic tests).[70]

Prompted by concerns about patent anticommons and anecdotal experiences about biotech patenting, the research arm of the NAS commissioned a study on the effects of patenting in the biomedical sciences ("Walsh Study").[71] Surprisingly, the authors of the Walsh Study found "little evidence of routine breakdowns in negotiations over rights, although research tool patents are observed to impose a range of social costs and there is some restriction of access."[72] They also concluded that although "access . . . to foundational upstream discoveries has not yet impeded biomedical innovation significantly, . . . interviews [by the authors] and prior cases suggest that the prospect exists and ongoing scrutiny is warranted."[73] The authors opined that, in addition to several "working solutions" that had evolved over time, the large number of opportunities in

68. Heller & Eisenberg, *supra* note 1, at 699. However, a subsequent review identified 135 patents using the search term "andrenergic receptor," but concluded that, at most, a handful of patents needed to be licensed for typical research on the receptor. Walsh et al., *supra* note 5, at 294-95 (discussing research done by R.K. Seide and J.M. MacCloud in response to Heller and Eisenberg).

69. Walsh, *supra* note 5, at 298 n.18. Other upstream patents that have garnered attention include those on DNA probes based on expressed sequence tags (ESTs) and single nucleotide polymorphisms (SNPs), although recent tightening of the PTO's written description and utility requirements have defused some of these concerns. *Id.* at 287, 299.

70. NAS REPORT, *supra* note 20, at 62; Rai & Eisenberg, *supra* note 4, at 302.

71. According to the NAS committee, "there was only one area—biotechnology research and development, primarily where applied to humans' health—where it was repeatedly suggested that there might be a significant problem of access to patented technology." NAS REPORT, *supra* note 20, at 59.

72. Walsh et al., *supra* note 5, at 289, 331. This includes the risks from patent anticommons that were paramount in many people's minds. *Id.* at 317. The NAS committee also "found little evidence, one way or the other, of the economic effects of the many steps taken during the 1980s and 1990s to extend and strengthen intellectual property rights." NAS REPORT, *supra* note 20, at 9. For example, a patent term may be extended to account for the time consumed to obtain FDA approval to market and sell a drug in the United States. NAS REPORT, *supra* note 20, at 10, 24.

73. Walsh et al., *supra* note 5, at 331. The study further noted that it is important to distinguish between research tools with only rival uses (such as Geron's stem cells and diagnostic tests) and those that have nonrival uses as well, since the latter are much less likely to be used exclusively by the patentee. *Id.* at 332-33.

biotech research had neutralized much of the potential for patents to impede innovation.[74]

The more detailed findings of the Walsh Study provide insight into these broader conclusions about the rise in biotech patenting. In a series of interviews, the Walsh Study found near unanimity among those interviewed that addressed the issue that the patent landscape has become more complex and requires more extensive due diligence.[75] Yet, while some respondents acknowledged that a large number (sometimes hundreds) of patents may need to be considered initially, "in [general] practice there may be, in a complicated case, about 6-12 that they have to seriously address, but that more typically the number was zero."[76] In sum, the number of patents one must evaluate is generally manageable. Consistent with this general result, the Walsh Study found that, although time consuming, negotiations over licensing agreements rarely halted projects[77] and that royalty payments rarely threatened the commercial viability of downstream products and virtually never halted research projects.[78]

The Walsh Study is equivocal about patents on research tools. The primary issue it identified in this context is ensuring access to important research tools, such as unique drug targets and stem cells that are covered by one or a few patents.[79] Half of the Walsh Study's respondents complained of licensing fees on research tools, but nevertheless conceded that

---

74. *Id.* at 331-32.

75. *Id.* at 294.

76. *Id.* at 294-95.

77. *Id.* at 315-16. In fact, fifty-four of fifty-five respondents could not even identify a specific incident. *Id.* Further, although only indirectly related to patenting, material transfers were found to be "a source of some concern and vexation," as the process is very bureaucratic and time intensive. *Id.* at 319-21. Even when parties are willing to share, the process is complicated by agreements, with time often increasing from days to months. *Id.*

78. *Id.* at 299. The norm for royalty payments on drug development programs, for instance, is one to five percent of sales, with exclusive royalties being higher, and royalties of ten percent being viewed as "high" or "ridiculous." *Id.* at 300. Royalty stacking could affect a decision made at the margins (where there were two equally viable candidates); probability of success and the size of the market are more central concerns. *Id.* at 304.

79. *Id.* at 305-06 ("[T]his is not a problem of accessing multiple rights but one of accessing relatively few—perhaps even one—patent on a key tool or discovery."). Particular concern has been expressed about exclusivity arrangements for drug targets (that is, "any cell receptor, enzyme, or other protein implicated in a disease, thus representing a promising locus for drug intervention"). *Id.* at 310. Indeed, one of the genes with a strong association with breast cancer, BRCA1, has been the subject of substantial controversy because of the limited access the patent owner for the gene is permitting. *Id.* at 312.

the costs did not preclude projects.[80] Further, while royalties are often high, respondents acknowledged that fees on research tools were within reason given productivity gains.[81] Redirecting research projects around research tool patents was also found to be common, but in most cases did not entail shifting to an entirely new research area (for example, new disease or technical approach).[82] The complexity of most diseases apparently permits a range of different research strategies.

The Walsh Study concluded by describing a number of working solutions that have mitigated the potential impact of strategic patenting in the biomedical sciences. Most of them are obvious, such as licensing, inventing around, and court challenges, but a few are more unexpected. Two significant working solutions are the use of technology without a license and the resurgence of support for public databases in the public and private sectors.[83] Norms of the research community, as Arti Rai has argued, play an important role in these developments.[84] In particular, researchers, whether public or private, are less likely to enforce their patents when it will erode the personal relationships and the information exchange integral to the scientific community.[85] As a consequence, university researchers, and to a lesser extent even those in the private sector, routinely use patented inventions without obtaining a license under the guise of a "research exception" to patent liability.[86] The viability of this working solution is aided by the difficulty of enforcing patent rights against research in-

---

80. *Id.* at 300.

81. *Id.* at 301, 335. Exceptions to this general finding do of course exist. DNA chips were singled out as particularly expensive and beyond the reach of most small labs, forcing—for better or worse—collaborations between companies. *Id.* at 302. There also have been some efforts by companies to allow access to research tools at reduced costs to academic researchers. *Id.*

82. *Id.* at 303. Moreover, redirection was generally associated with patents on specific compounds, not on processes or techniques. *Id.*

83. *Id.* at 331.

84. Rai, *supra* note 4, at 90-92.

85. Walsh et al., *supra* note 5, at 331 (stating that companies, in particular, rarely sue universities for fear of the bad press that would ensue). The importance of scientific norms is also reflected in the reluctance of early university inventors to patent landmark biotech methods (for example, the Cohen-Boyer process). Rai, *supra* note 4, at 93-94.

86. Walsh et al., *supra* note 5, at 324-28, 334. The recent Federal Circuit opinion in *Madey v. Duke University*, 307 F.3d 1351, 1361-62 (Fed. Cir. 2002), may foreclose this working solution. The court held that the research exception does not apply to basic science research at universities because research is, in effect, the business of universities. *Madey*, 307 F.3d at 1362. The court further held that only acts "solely for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry, . . . qualify for the very narrow and strictly limited experimental use defense." *Madey*, 307 F.3d at 1362.

fringement because it is difficult to detect given its small scale and the absence of an open sale or manufacture of an infringing product.

The recent growth of public databases and efforts to promote public access to information and techniques play an important role as well. Major databases exist for genes (such as GenBank), proteins (such as Blueprint Worldwide and Protein Data Bank), and genetic probes (such as the quasi-public Merck Gene Index and SNPs Consortium).[87] Similarly, Merck has initiated a program to provide to the research community, at cost, 150 patent-free transgenic mice without use restrictions.[88] The scientific community has also been instrumental in preserving and enhancing openness and access to technologies. Biology journals, for example, often require authors to deposit gene and protein sequences in public databases.[89] More recently, the NIH has negotiated generic license agreements for academic researchers to ensure that they obtain access to important privately-owned research tools, has provided funding for development of new research tools (for example, transgenic lab animals), and has conditioned receipt of grants on commitments not to patent inventions that derive from NIH-supported research.[90] These developments, coupled with the results of the Walsh Study, suggest that forces beyond those identified by legal scholars are having a major effect on biotech patenting.

C. The Divergence of Legal Theory and Biotech Patenting

None of the previously described legal theories readily captures the dynamics of patenting in the biomedical sciences. First, although the Walsh Study found that the expanding number of patents requires more negotiations for licenses and increases the costs of biomedical research,[91] it has not led to Heller and Eisenberg's dire anticommons predictions. Biomedical research has not been markedly impeded by the growing number of biotech patents.[92] It appears instead that working solutions aided, as

87. Walsh et al., *supra* note 5, at 329. Celera Genomics, founded to complete the task of sequencing and assembling the human genome, has recently decided to make all of its genomics information publicly available after acknowledging that its business model for selling genomics information was failing. *See, e.g.*, Andrew Pollack, *Celera to Quit Selling Genome Information*, N.Y. TIMES, Apr. 27, 2005, at C2. Some have argued that this openness benefits the large pharmaceutical companies at the expense of the biotech sector, as it is in the interest of the pharmaceutical companies to undercut business opportunities of biotech firms and then to "compet[e] on the exploitation of this shared information." Walsh et al., *supra* note 5, at 329.

88. Walsh et al., *supra* note 5, at 329.

89. *Id.*

90. *Id.*

91. *See* discussion *supra* Part I.B.

92. *Id.*

I will argue below, by the characteristics of the science itself have mitigated many of the negative effects of this trend.

Second, the most serious threats the Walsh Study identified were from discrete patents on key research tools.[93] This finding undercuts Merges and Nelson's narrow-patents theory because, even where narrowly drawn, patents on key research tools can be used to limit a diverse range of work by competitors. For example, even a narrow patent on the foundational technology for all recombinant DNA processes discovered by Stanley Cohen and Herbert Boyer, or other irreplaceable research tools, would have a broadly preclusive effect if access to the technology were denied. Conversely, where numerous alternative research tools are available, promotion of narrow patents is unnecessary, as alternative avenues for conducting research already exist. For similar reasons, the prescriptions that Burk and Lemley advocate are inappropriate.[94] Raising the obviousness standard, as Burk and Lemley advise, will have little or no effect, because key research tools by their very nature represent major advances beyond the prior art. Worse still, the loosening of disclosure requirements urged by Burk and Lemley stands to aggravate technology-access problems by allowing patents on research tools to claim a broader constellation of uses.[95]

The failings of the other legal theories might incite market enthusiasts to claim victory for traditional economic theory. According to this line of argument, the defects of these other theories demonstrate that patent scope is secondary to maintaining clear rules, strong property rights, and low transaction costs for technology licensing. The problem with this view is that transaction costs for licensing biotech patents are in fact significant and are not diminishing.[96] More importantly, the two working solutions identified by the Walsh Study—reliance on the now defunct research exemption and dedication of research tools to the public domain—do not center on reducing transaction costs or clarifying the law, but instead in-

93. Walsh et al., *supra* note 5, at 305-06.

94. Burk and Lemley conclude that "[b]iotechnology is in part about pharmaceuticals—and therefore prospect theory—and in part about DNA research—and therefore anticommons theory." Burk & Lemley, *Policy Levers*, *supra* note 12, at 1676.

95. *Id.* These problems also lend credence to Wagner's micro-specificity argument, for it appears that different biotech inventions pose different sets of problems for patent policy. *See supra* Part I.A. If this is the case, it makes no sense to treat biotech patents on a technology-specific basis.

96. Walsh et al., *supra* note 5, at 314-15 (noting that more than a third of the respondents in the study commented that transaction costs and delays associated with licensing technology were significant).

volve abrogating property rights and abandoning private ownership altogether.[97]

Biotech patenting dynamics are not solely attributable to legal or economic factors, though both are important.[98] Making sense of the interplay between law, economics, and science in the biomedical sciences requires that the third prong of this trio be more fully understood and taken into account. The Walsh Study includes two revealing observations in this regard. First, one of the respondents remarked that "[w]e have more targets than we have chemists to work on them" and noted later that the value for targets has decreased over time due to their abundance.[99] Second, another respondent commented:

> I have never worked with a disease where one particular protein makes the only difference. A patent gets you exclusive rights to a class of drugs, but there may be other classes. . . . I could imagine a genetic disease where a single target was involved, but I don't think that the big medical problems fall into this case.[100]

Both observations highlight the diverse set of research options that have emerged in the biomedical sciences.[101] This diversity helps explain the

---

97. *Id.* at 324-27, 329. In a recent article that challenges the "Conventional Critique in the intellectual property (IP) world," Merges refers to strategies that involve dedicating technology and data to the public domain as "property-preempting investments," which he argues are motivated by the desire of firms and individuals to "preempt[] or undermin[e] the potential property rights of economic adversaries." *See* Robert P. Merges, *A New Dynamism in the Public Domain*, 71 U. CHI. L. REV. 183, 183 (2004). This Article ascribes these strategies to factors that go beyond broad economic theorizing, which does not necessarily imply that the scientific factors discussed here have a significant impact on individual economic calculations.

98. For example, recent cases that have invalidated or narrowed the scope of important patents have eased some concerns about patenting. Walsh et al., *supra* note 5, at 330. The recent case of *Rochester v. G.D. Searle & Co.*, 358 F.3d 916 (Fed. Cir. 2004), *cert denied*, 125 S. Ct. 629 (2004), is a case in point. The Federal Circuit invalidated the broadly claimed patent based on a stringent application of the written description requirement. *Rochester*, 358 F.3d at 928-29.

99. Walsh et al., *supra* note 5, at 304-05; *see also* Allison Abbott, *Geneticists Prepare for Deluge of Mutant Mice*, 432 NATURE 541, 541 (2004) (forecasting a "glut of new mouse strains" for use as experimental models of human disease).

100. Walsh et al., *supra* note 5, at 324.

101. These statements are borne out by estimates that the total number of "druggable" targets is about 5,000 to 10,000—notably, the number of targets for safe and effective existing drugs is a mere 483. Jurgen Drews, *Drug Discovery: A Historical Perspective*, 287 SCIENCE 1960, 1961-62 (2000); *cf.* Andrew L. Hopkins & Colin R. Groom, *The Druggable Genome*, 1 NATURE REVS. DRUG DISCOVERY 727, 729 (2002) (estimating that the number of small molecule targets is likely between 600 and 1500 whereas the likely druggable targets number approximately 3000). However, regardless of what the exact

(apparently) lowered protectionist tendencies of inventors in the private and public sectors. Just as importantly, these observations point to the numerous opportunities for developing new research tools.[102] As argued more fully below, the nature of biomedical science itself has played a critical role in reducing potential friction between biotech innovation and the patent system.

II. HUMAN GENETICS AND BIOTECH INNOVATION

The systemic technical barriers inherent in the biomedical sciences that are responsible for the diversity of research opportunities and the uncertainty of success rarely are factored into patent policy.[103] This Part aims to redress this oversight.[104] The Part begins by highlighting the gulf that exists between the popularized version of biotech science and the more complex, less deterministic factors that shape biomedical science. This Part also challenges the common portrayal of genes as rigid blueprints that fully determine an individual's susceptibility to disease. These distinctions are critical to appreciating the relationship between biomedical science and biotech patenting.

The public and scientific images of human genetics are chronically estranged. In its most simplistic form, public perception is that genes determine and control an individual's susceptibilities to disease. This view is analogous to the claim that the food one eats fully determines who one is and what one does. Literally speaking, we certainly are constructed out of what we eat, but it is equally true that we are much more than these constituent parts—traits and characteristics emerge at the level of a whole or-

---

number of druggable targets is, a great deal of sifting of genetic data will be required given that scientists estimate there are 20,000 to 25,000 genes in the human genome and that the number of proteins is possibly about 1 million. Christopher P. Austin et al., *NIH Molecular Libraries Initiative*, 306 SCIENCE 1138, 1138 (2004).

102. *See* Walsh et al., *supra* note 5, at 304-05.

103. I do not mean to suggest that the interplay between science and patenting is absent from the legal literature as a whole, only that it has not been adequately considered in the biotech context. *See, e.g.*, Mark A. Lemley, *The Economics of Improvement in Intellectual Property Law*, 75 TEX. L. REV. 989, 1035-37 (1997) (describing how "subsequent developers . . . must work within the parameters of the physical laws, and hence may be forced to build on the original inventor's work").

104. For a more detailed discussion of factors inhibiting genomics research, see David E. Adelman, *The False Promise of the Genomics Revolution for Environmental Law*, 29 HARV. ENVTL. L. REV. 117, 151-67 (2005).

ganism that cannot be reduced to the elements that make them up.[105] What we eat or do not eat may influence our behavior or fate, but it would be absurd to infer that humans are fully determined by what they eat. The relationship between an individual's genetic makeup and disease susceptibilities is no different; a few susceptibilities have strong genetic influences, while the majority are affected by genetic factors only weakly.[106]

This less deterministic understanding is illustrated by the story of the BRCA1 and BRCA2 genes, which are strongly associated with breast and ovarian cancers. The BRCA genes represent a best-case scenario for biotech methods because they involve single genes that have a large impact on risk.[107] However, while these mutations are strongly correlated with cancer, only about ten percent of women with breast cancer develop the disease because they have a mutation in either of these genes.[108] Moreover, the estimates of the genetic link (eighty-five percent for breast cancer; forty-five percent for ovarian cancer) are in doubt, as other genetic and environmental factors play a role in disease etiology.[109] As a result,

---

105. *See* Kenneth M. Weiss & Anne Buchanan, *Evolution by Phenotype: A Biomedical Perspective*, 46 PERSP. BIOL. MED. 159, 178 (2003).

106. A number of diseases such as cancer have been associated with genetic variations in tens of genes. *Id.* at 172. Further, the relationship between genetics and disease can be complicated by much more mundane factors, such as physiological differences that may aggravate or neutralize the effect of a genetic mutation. *See, e.g.*, Samuel M. Cohen, *Risk Assessment in the Genomics Era*, 32 TOXICOLOGIC PATHOLOGY, at 5-6 (Supp. 1 2004) (describing how basic physiological differences between animal models and humans are determinative of whether certain chemicals heighten this risk of bladder cancer and concluding that "[g]enomics will contribute little to this risks assessment").

107. Weiss & Buchanan, *supra* note 105, at 167, 175.

108. Ruth Hubbard & Richard C. Lewontin, *Pitfalls of Genetic Testing*, 334 NEW ENG. J. MED. 1192, 1192 (1996). Over one hundred variants of the two genes have been identified, but only a few have been linked to tumor growth, and predominantly in women whose family histories provide independent grounds for finding a high familial risk of breast cancer. *Id.*

109. *Id.* Recent work, for example, has shown that lifetime risks vary significantly (for example, depending on the decade when the woman was born), suggesting that the cancer risks associated with these mutations may be overstated. Weiss & Buchanan, *supra* note 105, at 175. An important potential source of error is confounding factors in multiply-affected families. Weiss & Buchanan, *supra* note 105, at 175. In other words, scientists have not yet demonstrated that mutations BRCA1 and BRCA2 result in the increased susceptibility, as other genes or factors in these families could be the putative "cause." Hubbard & Lewontin, *supra* note 108, at 1192. This ambiguity arises because one cannot know *a priori* whether a trait is common because of an inherited characteristic or because of a functional genetic reason. Weiss & Buchanan, *supra* note 105, at 174. A disease may be common because an environmental factor affects a particular genetic or molecular-level pathway shared by everyone or because a specific underlying genetic variant is common. Weiss & Buchanan, *supra* note 105, at 174.

even a positive test provides a highly ambiguous guide for a doctor in counseling her female patient given the underlying uncertainties of the causal link.[110]

These qualifications have led patients and doctors alike to view genetic testing and genomic methods skeptically.[111] They also highlight the many complexities and uncertainties that underlie biomedical research—even after specific genetic anomalies have been identified. Legal commentators often fail to appreciate fully the seriousness of these obstacles, or the open-ended nature of the biomedical sciences, at this point in their development.[112] The following discussion explains the scientific origins of these uncertainties and barriers to development in order to examine how they impact biotech patenting.

A. Complexity in the Biomedical Sciences

The human genome and the translation of genes into biologically active molecules (usually proteins) are far more complex than popularized versions of genetics would lead one to believe. First, less than two percent of the human genome codes for proteins, and more than fifty percent consists of repeat sequences with currently undefined functions.[113] Second, genes themselves are oddly constructed—most are interspersed with long

110. *See* Weiss & Buchanan, *supra* note 105, at 175. Furthermore, in the case of relatively rare genetic disorders, such as BRCA1 and BRCA2, broad public genetic testing may not even be cost effective, particularly given the risk of false positives. Neil A. Holtzman & Theresa M. Marteau, *Will Genetics Revolutionize Medicine?*, 343 NEW ENG. J. MED. 141, 142-44 (2000); Hubbard & Lewontin, *supra* note 108, at 1193-94; *see* Paolo Vineis et al., *Misconceptions About the Use of Genetic Tests in Populations*, 357 LANCET 709, 710-11 (2001).

111. Richard S. Cooper & Bruce M. Psaty, *Genomics and Medicine: Distraction, Incremental Progress, or the Dawn of a New Age?*, 138 ANN. INTERNAL MED. 576, 577-78 (2003) ("[T]he available empirical data support the argument against a clinical role for susceptibility testing for chronic disease."); Hubbard & Lewontin, *supra* note 108, at 1192-93.

112. *See, e.g.*, Burk & Lemley, *Policy Levers*, *supra* note 12, at 1677-78 ("The availability of research tools has made the isolation and characterization of biological macromolecules routine."); Eisenberg, *supra* note 9, at 1070 ("[D]educing a gene sequence from its amino acid sequence [is] not . . . a particularly risky or uncertain step."); Rai & Eisenberg, *supra* note 4, at 289 ("Once largely a matter of serendipity or trial-and-error, drug discovery is now critically dependent on basic knowledge of genes, proteins, and associated biochemical pathways.").

113. Alan E. Guttmacher & Francis S. Collins, *Genomic Medicine—A Primer*, 347 NEW ENG. J. MED. 1512, 1514 (2002).

segments of non-coding DNA.[114] Third, many critical processes that alter the activity of a gene or its protein product are not controlled by DNA sequences.[115] Cellular processes may include complex feedback mechanisms that act directly on the protein or influence gene activity levels, for example.[116] These complex "epigenetic" dynamics are a distinguishing feature

114. *Id.* In fact, different coding sequences of a gene may be linked together in a variety of ways, such that the 30,000 to 35,000 genes in the human genome code for more than 100,000 proteins. *Id.*

115. *Id.* (citing examples such as the signals that turn genes on and off and molecules that activate and deactivate critical proteins). These alternate control mechanisms have emerged because "[t]he evolution of additional complex attributes is essentially an organizational one," not a product of major genetic modifications. Gerald M. Rubin et al., *Comparative Genomics of the Eukaryotes*, 287 SCIENCE 2204, 2214 (2000). Current evidence suggests that "the majority of phenotypic variation between individuals (and species) results from differences in the control architecture, not the proteins themselves." John S. Mattick & Michael J. Gagen, *The Evolution of Controlled Multitasked Gene Networks: The Role of Introns and Other Noncoding RNAs in the Development of Complex Organisms*, 18 MOLECULAR BIOL. EVOLUTION 1611, 1622-23 (2001); *see also* David K. Gifford, *Blazing Pathways Through Genetic Mountains*, 293 SCIENCE 2049, 2049 (2001).

116. A recent issue of *Science* contained a special section on "Mathematics in Biology," and other recent articles have also highlighted the rising importance of mathematical modeling and the study of complexity in biological systems. *See* Gilbert Chen et al., *Biology by the Numbers*, 303 SCIENCE 781, 781 (2004); Ronald N. Germain, *The Art of the Probable: System Control in the Adaptive Immune System*, 293 SCIENCE 240, 244 (2001) ("[I]t is now time to add the power of mathematics, systems analysis, and quantitative cell-based modeling [to the study of complex biological systems (e.g., the immune system)]."); Hiroaki Kitano, *Systems Biology: A Brief Overview*, 295 SCIENCE 1662, 1662 (2002); Robert F. Service, *Exploring the Systems of Life*, 284 SCIENCE 80, 82 (1999) (arguing that scientists will need to develop complex models for biological systems). This focus on modeling is motivated both by the needs of genomics research and the realization that biological systems often operate more as networks, with different pathways interacting, than as systems driven from the smallest level up by the same fundamental forces.

of biological systems.[117] They also cannot be ignored because epigenetic processes play an important role in certain diseases.[118]

These features of human biology make the process of identifying genes far from trivial, and the task of linking genes to specific diseases extraordinarily challenging.[119] Finding the genetic origin of a disease is further complicated by the fact that genetic mutations, if present, are not necessarily determinative of disease onset.[120] Environmental factors, such as nutrition, exercise, and chemical exposures, are believed to be more important for common diseases.[121] This multifactor etiology follows from

---

117. Epigenetics is the study of heritable changes in gene expression that occur other than those resulting from a change in DNA sequence. Rebecca E. Watson & Jay I. Goodman, *Epigenetics and DNA Methylation Come of Age in Toxicology*, 67 TOXICOLOGICAL SCI. 11, 11 (2002) (stating that "adaptive epigenetic inheritance challenges the 'central dogma' that information is unidirectional from DNA to protein" and that epigenetic processes are unimportant in assessing potential chemical toxicity). Examples of epigenetic phenomena include silencing of tumor genes through chemical modifications, short double-stranded RNA (RNAi) segments that mediate gene expression, and DNA-DNA, DNA-RNA, and RNA-RNA interactions that trigger gene silencing. Alan P. Wolffe & Marjori A. Matzke, *Epigenetics: Regulation Through Repression*, 286 SCIENCE 481, 483 (1999).

118. Frederica P. Perera & I. Bernard Weinstein, *Molecular Epidemiology: Recent Advances and Future Directions*, 21 CARCINOGENESIS 517, 521 (2000) (stating that many carcinogenic chemicals act "through indirect genotoxic or epigenetic mechanisms").

119. A test of gene detection methods on the Drosophila genome, for instance, was mixed. The accuracy of the methods used to find genes varied between five to ninety-five percent, and they incorrectly identified up to fifty-five percent of the genes studied. Teresa K. Attwood, *The Babel of Bioinformatics*, 290 SCIENCE 471, 471 (2000) (citing M.G. Reese et al., *Genome Annotation Assessment in* Drosophila melanogaster, 10 GENOME RES. 483 (2000)).

120. Weiss & Buchanan, *supra* note 105, 172-73; Walter C. Willett, *Balancing Life-Style and Genomics Research for Disease Prevention*, 296 SCIENCE 695, 696 (2002) ("[T]he majority—probably the large majority—of important cancers in Western populations are due to environmental rather than genetic factors."). Indeed, critics of genomics methods reject the view "that the genetic determinants of complex traits are tractable, and that knowledge of genetic variation will materially improve the diagnosis, treatment or prevention of a substantial fraction of cases of the diseases that constitute the major public health burden of industrialized nations." Kenneth M. Weiss & Joseph D. Terwilliger, *How Many Diseases Does It Take to Map a Gene With SNPs?*, 26 NATURE GENETICS 151, 151 (2000).

121. Cooper & Psaty, *supra* note 6, at 578-79 ("The primary disease-producing forces are rooted in our technology based lifestyle and the resulting patterns of consumption, behaviors, and environmental exposures."); Weiss & Buchanan, *supra* note 105, at 174 ("[E]nvironmental factors play a major—probably *the* major—role in risk for common diseases."); Willett, *supra* note 120, at 695-96 (arguing that "most diseases contributing to mortality in Western populations" are dominated by nongenetic factors).

genes being a necessary but only very rarely a sufficient contributor to disease.[122]

Two central barriers to biomedical innovation emerge from this understanding: (1) genes do not have a fixed negative or positive impact on human health, and (2) a weak causal association exists between a person's genetic makeup and her susceptibility to disease.[123] These barriers place the most significant limits on the commercialization of biotechnology. They are also responsible for the disparity that exists between the power of biotech methods to generate data, such as genome sequences and probes, and their ability to promote the discovery of new medical procedures and drugs. Biotechnology is in the somewhat paradoxical position that it can produce vast quantities of genetic data, much of which are useful as research tools, but has so far had great difficulty overcoming these two fundamental barriers to innovation.[124] The following discussion describes each of these barriers to innovation in greater detail.

A canonical principle in biology is the dependence of a gene's function on other genes and environmental factors. According to this principle, known as the "Genetic Theory of Relativity," a gene may be highly beneficial in a person with one set of genetic attributes or genetic background, and extremely harmful to another person having a different genetic background.[125] As a consequence, genes typically have multiple effects depending on one's genetic background and the environment in which one lives.[126] This variation creates two central challenges: first, it negates the

---

122. Jonathan Rees, *Complex Disease and the New Clinical Sciences*, 296 SCIENCE 698, 699 (2002); Richard Strohman, *Maneuvering in the Complex Path from Genotype to Phenotype*, 296 SCIENCE 701, 701 (2002).

123. Weiss & Terwilliger, *supra* note 120, at 152 (stating that the central inferential problem is that a specific genotype does not imply a specific phenotype nor does a specific phenotype imply a specific genotype; they are not necessarily correlated).

124. *Id.*

125. Anne Glazier et al., *Finding Genes That Underlie Complex Traits*, 298 SCIENCE 2345, 2345 (2002) ("The propensity of genetic background to modify the phenotypic expression of most if not all Mendelian traits suggests that few if any traits are truly monogenic and that instead most are genetically complex."); Elliot Sober & Richard C. Lewontin, *Artifact, Cause, and Genic Selection*, 49 PHIL. SCI. 157, 159 (1982) ("[N]o gene has a fixed selective value, the same gene may confer high fitness on one genetic background and be virtually lethal on another.").

126. Mark S. Boguski, *Biosequence Exegesis*, 286 SCIENCE 453 (1999). Some scientists have argued that,

> If only 1 in 10,000 of the [mutations] present in the human population has some quantifiable [tangible] effect, then there would be more than enough unique combinations of these [mutations] to assure that every

central genomic mission of ascribing fixed disease susceptibilities to genes; second, it introduces variability that undermines biotech methods designed to link disease states using gene sequence or expression levels.

Even in single-gene diseases, the disease phenotype depends on an individual's genetic background and environmental conditions.[127] The effect of the genetic mutation that causes sickle cell anemia provides a simple example of this variability. The mutation involved in sickle cell anemia has counterbalancing effects—when present in two copies, it degrades the functioning of red blood cells and when present in only one copy, it provides resistance to malaria.[128] Symptoms consequently range from severe anemia for individuals with two copies of the mutation, to resistance to malaria for individuals who have one mutated and one normal copy, to no effects for individuals with two normal copies of the gene who are not exposed to malaria.[129] For complex diseases, the variation will be more intricate because a number of interacting genes will be involved.[130] The end result is the same. Genes rarely exert fixed effects in individuals with different genetic backgrounds or across different environments.

Thus, the causal relationship between genetic makeup ("genotype") and disease susceptibility ("phenotype") is not a simple one.[131] First, natural selection acts primarily on phenotype, and only indirectly on genotype.[132] This indirect relation decouples genotype from phenotype, such

---

> human being (with the exception of identical twins) should have a unique [set of tangible effects] . . . .

John L. Hartman IV et al., *Principles for the Buffering of Genetic Variation*, 291 SCIENCE 1001, 1001 (2001).

127. Hartman et al., *supra* note 126, at 1001 (stating that there is no "wild-type" in nature; all disease and chemical toxin susceptibilities are "arbitrarily defined [at a] point along a spectrum"); Julian Little et al., *The Human Genome Project Is Complete. How Do We Develop a Handle for the Pump?*, 157 AM. J. EPIDEMIOLOGY 667, 669 (2003); Weiss & Buchanan, *supra* note 105, at 167.

128. Sober & Lewontin, *supra* note 125, at 163-64.

129. *Id.* at 163-67.

130. Ernst Mayr, *The Objects of Selection*, 94 PROC. NAT'L ACADS. SCI. USA 2091, 2092 (1997).

131. Robert Milliken, *The Changing Face of Epidemiology in the Genomics Era*, 13 EPIDEMIOLOGY 472, 474 (2002) (arguing that a huge gap exists in understanding the relationship between genotype and phenotype because unmeasured genetic and environmental factors can influence expression); Strohman, *supra* note 122, at 701 (stating that the progression from genotype to phenotype extends over four basic levels of control, "[e]ach control level [of which] is defined by a dynamic system of self-organizing proteins, the output of which is governed by laws that are still poorly understood").

132. Mayr, *supra* note 130, at 2093; Weiss & Buchanan, *supra* note 105, at 160. This distinction is important because the indirect relationship between natural selection and genotype allows genetic drift (that is, selectively neutral genetic variation) to propagate

that a phenotype may remain fixed under the pressures of natural selection, while the underlying genotype varies significantly.[133] As a consequence, it generally is not possible to infer genotype from an observed phenotype because the same phenotype can arise from multiple genotypes.[134] The absence of a unique, or even well-defined, genotype-phenotype relationship complicates the process of identifying meaningful genetic signatures of disease, and may erode the association between genetics and disease altogether.

Second, biological processes actively buffer phenotype from variations in genotype.[135] A genetic mutation that, for example, inhibits the activity of an important metabolic enzyme may be neutralized by processes that counteract the impact of the mutation on the enzyme's function or by redundancies built into the specific metabolic process.[136] Buffering mechanisms may also cause specific genotypes to be associated with diverse

---

over time. Weiss & Buchanan, *supra* note 105, at 159 ("[G]enetic variation is abundant in all natural species, and most is expected to be neutral or nearly neutral with respect to fitness."). The term "natural selection" is used in its traditional sense to mean the process in which environmental factors impose pressures on species that create differential levels of survival and reproduction, such that, in the long run, the best adapted ("fittest") members of the species are selected for survival.

133. Hubbard & Lewontin, *supra* note 108, at 1192 (stating that appearance of the same trait in different people need not be associated with the same genetic polymorphism; for example, 200 different nucleotide variations appear to produce hemophilia B); Weiss & Buchanan, *supra* note 105, at 164. In fact, "even strong [natural] selection favoring a specific phenotype closely tied to specific genes does not usually purify [genetic] variation." Weiss & Buchanan, *supra* note 105, at 170-71.

134. Weiss & Buchanan, *supra* note 105, at 165. Moreover, once the classical two-allele (abnormal, wild-type) classification scheme is abandoned, substantial phenotypic equivalence and a broad genotype-phenotype distribution results from numerous alleles. *Id.* at 167-68.

135. *See* Suzanne L. Rutherford, *Between Genotype and Phenotype: Protein Chaperones and Evolvability*, 4 NATURE REVS. GENETICS 263, 263-64 (2003) (stating that specific biological molecules exist that buffer the "expression of genetic variation as phenotypic variation"). The prevalence of genetic buffering is driven by the important role it plays in natural selection. Genetic buffering allows a reserve of neutral genetic variation to build up in a population under stable conditions. *Id.* at 271. This reserve is critical to a species' resiliency to environmental change and alters natural selection pressures because it provides a genetic reservoir upon which a species can draw in response to changed circumstances. *Id.* at 263, 271.

136. Suzanne L. Rutherford, *From Genotype to Phenotype: Buffering Mechanisms and the Storage Genetic Information*, 22 BIOESSAYS 1095, 1095 (2000) (stating that many ways exist in which phenotypes are buffered from perturbation by genomic and environmental variation).

phenotypes.[137] Buffering of the genetic trait therefore weakens the association between gene activity levels, which are central to genomic studies, and disease susceptibility.[138]

Third, a simple one-to-one relationship does not exist between genotype and phenotype because they are separated by intervening epigenetic and apparently random processes.[139] For example, epigenetic processes may determine whether or not a gene is activated and may play a significant role in the toxicity of certain compounds.[140] Similarly, growing evidence indicates that stochastic processes are integral to disease-response mechanisms.[141] This innate uncertainty adds to the complexity of the genotype-phenotype relation because, owing to variations in environmental conditions, a given genotype will have multiple corresponding phenotypes.[142]

All three of these factors—natural selection acting on phenotype, active buffering processes, and stochastic biological processes—stand in the way of discovering effective medical procedures and drugs. Each of these

---

137. Hubbard & Lewontin, *supra* note 108, at 1192 (explaining that having the same DNA sequence in a gene does not guarantee that different people will display the same phenotype; for example, autosomal dominant retinitis pigmentosa display a range of effects from complete blindness to completely functional vision); Rutherford, *supra* note 136, at 1095 (noting that the genotype of essential biochemical pathways, for instance, can be disrupted in some strain backgrounds with minimal phenotypic effect, while in other genetic backgrounds the organism is severely affected).

138. Rutherford, *supra* note 136, at 1097. An important problem for genomics methods is that they cannot distinguish between cases in which a phenotype arises because of limited genetic variation due to a high degree of buffering or because the trait is constrained in some other way (for example, biochemical constraints). *Id.* at 1102.

139. Michael B. Elowitz et al., *Stochastic Gene Expression in a Single Cell*, 297 SCIENCE 1183, 1183, 1186 (2002) (explaining that random epigenetic signaling can generate long-term heterogeneity among animals with identical genetic backgrounds); Germain, *supra* note 116, at 241 ("[T]he difference between health and disease could be the 'stochastic' activation of a single cell, followed by positive feedback in the form of a gain in [] sensitivity and multiplication of the responding cells to high numbers."); Simon A. Levin et al., *Mathematical and Computations Challenges in Population Biology and Ecosystems Science*, 275 SCIENCE 334, 337 (1997) ("[S]tochastic effects become paramount [in biological systems]."); Mayr, *supra* note 130, at 2092.

140. Elizabeth Pennisi, *Behind the Scenes of Gene Expression*, 293 SCIENCE 1064, 1065 (2001) (describing how epigenetic deactivation of tumor-suppressor genes can cause cancer); Watson & Goodman, *supra* note 117, at 12-13 (describing how chemical modifications to DNA that affect gene activity have been connected to chemical toxicity).

141. Germain, *supra* note 116, at 240-41; Hartman et al., *supra* note 126, 1001 (noting that diseases or toxic susceptibilities can be influenced differentially by environment factors, stochastic events, or interactions with other genes); Rutherford, *supra* note 136, at 1100.

142. Rutherford, *supra* note 136, at 1100.

processes complicates the interpretation of genetic studies by attenuating and, in some cases, completely obscuring the connection between gene-expression levels and the biological processes relevant to the disease responses and susceptibilities that scientists are attempting to monitor and understand. This decoupling makes the process of identifying useful drug targets and understanding the biology of diseases very challenging and uncertain, often necessitating extensive trial-and-error research.[143]

These three factors are compounded by the fact that most human health conditions are complex and multigenic.[144] The simple cases in which biotech methods have been applied successfully are the relatively rare exceptions.[145] Further, the most important, typically chronic, diseases in the United States have late onsets (that is, after an individual's reproductive years) which are even less likely to be determined by genetic traits alone.[146] This additional barrier arises because the late onset of these diseases makes them selectively neutral.[147] The end result is that biotech methods will have great difficulty overcoming the complex etiologies of

143. *See* Austin et al., *supra* note 101, at 1138-39 (describing the importance of small molecules in the trial-and-error research that will be necessary to exploit the Human Genome Project).

144. *See* Vineis et al., *supra* note 110, at 710-11 (stating that mutations in genes coding for proteins that metabolize environmental toxins are prototypical of common, but weakly associated, genetic defects that affect individual susceptibility to toxins); Weiss & Buchanan, *supra* note 105, at 174 (stating that evidence to date indicates that simple genetic variants with major effects on risk are the exception, not the rule).

145. Even simple organisms, such as yeast, display a high degree of complexity in their gene-gene interactions. In one recent study, scientists found an average of 34 gene-gene interactions per gene based on an analysis of 143 genes in yeast mutants. Lee Hartwell, *Robust Interactions*, 303 SCIENCE 774, 775 (2004). As the study's author acknowledges, "[f]or those interested in uncovering the genetic basis of disease susceptibility in the human population, this result is daunting." *Id.*

146. Julian Peto, *Cancer Epidemiology in the Last Century and the Next Decade*, 411 NATURE GENETICS 390, 390 (2002) (citing studies that suggest, for example, that cancer risks in old age may depend on diet in early life as much as a person's habits when she tract the cancer); Weiss & Buchanan, *supra* note 105, at 156 ("Late-onset chronic diseases—whose elimination through genetics is currently the supreme object of our affections—are much more complex by comparison.").

147. *See* Weiss & Buchanan, *supra* note 105, at 153. Type 2 diabetes, for example, has become much more severe in recent times, implying that environmental factors dominate. *Id.* at 154. Further, even the pandemic among certain Native American tribes, which has been shown to have a genetic component, is still aggravated by environmental factors because it was rare in the same populations sixty years ago. *Id.* at 175.

many important diseases (such as cancer, heart disease, and diabetes) found in the United States and elsewhere.[148]

The complexity of biological systems has important implications for patent policy. Above all, it means that multiple approaches will exist for understanding and treating most diseases, as multiple genes, biochemical pathways, and epigenetic factors are likely to be involved. However, it also has an obvious downside. Because human biology does not fit into a simple Newtonian model of science in which genes are the elementary objects that define biological systems as a whole, gene discovery does not lead inexorably towards an understanding of disease processes and viable treatment options.[149] Biotech methods are uncertain in large part because individual genes have a highly variable and typically very limited effect on disease processes.[150] Because of this, developing effective methods for monitoring and understanding common diseases will ultimately require scientists to address these more complex dynamics.[151] Until then, and likely beyond, biomedical science will be subject to large, unavoidable uncertainties that will require a great deal of trial-and-error research, creat-

148. *See* Nelson Freimer & Chiara Sabatti, *The Human Phenome Project*, 34 NATURE GENETICS 15, 16 (2003) ("[T]here is still relatively little known about how to integrate this effort with investigation of environmental influence on phenotype."); Strohman, *supra* note 122, at 701, 703 (noting that scientists are particularly ignorant of the interplay between disease and environmental factors); Weiss & Buchanan, *supra* note 105, at 153 ("[G]enetic factors are not likely to explain [common, chronic] diseases in the usual causal sense.").

149. *See* RICHARD C. LEWONTIN, THE TRIPLE HELIX 113-14 (2000). Lewontin writes:

> It is not new principles that we need but a willingness to accept the consequences of the fact that biological systems occupy a different region of the space of physical relations than do simpler physicochemical systems, a region in which the objects are characterized, first, by very great internal physical and chemical heterogeneity and, second, by a dynamic exchange between processes internal to the objects and the world outside of them. That is, organisms are internally heterogeneous open systems.

Id.

150. Biological signaling processes that control cellular responses to environmental exposures, for example, involve networks with "complex properties that are independent of genetic factors." Upinder S. Bhalla & Ravi Iyengear, *Emergent Properties of Networks of Biological Signaling Pathways*, 283 SCIENCE 381, 386 (1999).

151. Strohman, *supra* note 122, at 703 (objecting to policies that "continue to see complex phenotypes as primarily derivable from genomic and proteomic databases"). Biomedical scientists also acknowledge the need to come to terms with complex biological processes. Geoffrey Duyk, *Attrition and Translation*, 302 SCIENCE 603, 603-04 (2003) (arguing that the shrinking number of drugs being discovered is attributable to the failure of scientists to address biological complexity in a systematic manner).

ing a few islands of significant advances and insights in an ocean of rapidly proliferating genetic data.

B. Implications for Innovation in the Biomedical Sciences

The deficiencies of a heavily genetics-oriented approach to understanding human disease are evident in the limited successes of genomics thus far. The ideals of rational drug design and personalized medicine hyped by scientists are currently more aspiration than reality. The decline in new drug therapies, despite large infusions of public and private support, exposes the seriousness of these technical barriers,[152] as does the recent stream of published reports for which experimental results could not be reproduced.[153] Ironically, the power of biotech methods—particularly their ability to monitor thousands of genes simultaneously—comes at a significant price. The vast quantities of data generated raise extremely challenging problems for its analysis.[154] Indeed, discerning meaningful results from the masses of background noise requires novel methods that are computationally intensive and highly complex.[155] Predictably, the few

152. Rees, *supra* note 122, at 698; Strohman, *supra* note 122, at 702.

153. Habibul Ahsan & Andrew G. Rundle, *Measures of Genotype Versus Gene Products: Promise and Pitfalls in Cancer Prevention*, 24 CARCINOGENESIS 1429, 1429 (2003); Kirk E. Lohmueller et al., *Meta-Analysis of Genetic Association Studies Supports a Contribution of Common Variants to Susceptibility to Common Disease*, 33 NATURE GENETICS 177, 177 (2003) (finding that, in a meta-analysis of 301 published studies covering twenty-five different disease associations, "less than half of [the] genetic association studies [were] statistically significant"); Weiss & Buchanan, *supra* note 105, at 171-72; Kenneth M. Weiss & Andrew G. Clark, *Linkage Disequilibrium and the Mapping of Complex Human Traits*, 18 TRENDS GENETICS 19, 22 (2002).

154. PIERRE BALDI & WESLEY HATFIELD, DNA MICROARRAYS AND GENE EXPRESSION: FROM EXPERIMENTS TO DATA ANALYSIS AND MODELING, at viii (2002) ("The bioinformatics solutions to problems associated with the analysis of data on this scale are a major current challenge."); Hopkins & Groom, *supra* note 101, at 729 (finding that the viability of "targets identified by proteomic or transcriptional-profiling studies is likely to be low"); Richard D. Irwin et al., *Application of Toxicogenomics to Toxicology: Basic Concepts in the Analysis of Microarray Data*, 32 TOXICOLOGIC PATHOLOGY 72, 72 (2004) ("The amount and complexity of the data means that confounding factors can be easily missed and potential[ly] important changes may be overlooked.").

155. BALDI & HATFIELD, *supra* note 154, at 55-56 (discussing the multidimensional nature of biological systems and the sophisticated probabilistic methods that are being used to analyze them); Christopher P. Austin, *The Completed Human Genome: Implications for Chemical Biology*, 7 CURRENT OPIN. CHEM. BIO. 511, 512 (2003) (describing the huge number of genetic targets to be evaluated and the diverse range of experimental methods needed to validate them).

successful treatments that have emerged from biotechnology involve relatively simple cases.[156]

The difficulty of these challenges is perhaps best illustrated by the genetic variation found in DNA repair genes, which play an essential role in correcting mutations. Scientists have identified over 450 variants of DNA repair genes using genetic screens of a representative sample of the U.S. population.[157] The large number of genetic variants, each of which may have multiple phenotypes or no phenotype at all, creates a near-intractable problem for biotech methods:

> The complexity of . . . associating genetic variation with risk becomes apparent when it is realized that these repair pathways require activity of 20-40 different proteins to complete the repair process. Thus, given the large number of different variant[s], the typical individuals will be variant for 10-15 proteins required for repair of a specific class of damage. But, these typical individuals will not have similar pathway genotypes as these 10-15 variants will be drawn from a pool of 100-200 different [genetic variants].[158]

The numerous combinations possible imply that few people will have the same genetic variants, making genetic associations much harder to detect. Further, because the pathway as a whole determines disease risk, causal links between individual genetic variants in the pathway and disease susceptibility will be obscured. Moreover, the process is confounded by the underlying biological mechanisms discussed in the preceding Section. In essence, identifying gene-disease associations is analogous "to search[ing] for a needle in a needle stack," where the challenge is to identify the subset of genes that is causally related to the disease in question from a far greater number that are not.[159]

Three central areas pose particularly difficult problems for biotech researchers: (1) identifying genes and linking proteins to genes, or vice

---

156. Service, *supra* note 6, at 1799; *see also* Cooper & Psaty, *supra* note 6, at 577 ("To date, both [gene expression] studies and genome-wide scans have identified only weak and inconsistent genetic signals [for common diseases in the United States such as cardiovascular disease and cancer].").

157. Harvey Mohrenweiser, *Genetic Variation and Exposure Related Risk Estimation: Will Toxicology Enter a New Era? DNA Repair and Cancer as a Paradigm*, 32 TOXICOLOGIC PATHOLOGY 136, 139 (2004).

158. *Id.*

159. Weiss & Buchanan, *supra* note 105, at 155 ("[C]urrent genetic approaches . . . have been likened to a search for a needle in a needle stack (a great many individually modest effects).").

versa; (2) characterizing potential drug targets; and (3) using genomic technologies to understand disease etiology and effects. The scientific uncertainty and complexity feed into the dynamics that make patenting so important. Namely, a large differential exists between the cost of discovery, which requires much trial-and-error research, and the cost of copying and producing an invention, which utilizes standardized processes for reverse engineering the patented product. In other words, because scientists are reliant on the same research tools, they face the same barriers, which, once overcome, open the door to easy reproduction of newly discovered products.

Drug discovery, which is the ultimate objective of most biotechnology, adds several additional challenges and constraints. Viable drugs must do more than effectively modify relevant biological processes. They must remain active long enough to have an effect and be biologically available—that is, be adequately absorbed into a patient's blood stream, readily enter the cell types to which they are targeted, bind to their intended target for a sufficient period of time, and be metabolized at an appropriate rate.[160] These factors place significant constraints on potential drug targets and drugs themselves, which often require determination of a drug target's structure before it can be addressed.[161] The "vast majority of successful drugs" are directed at modifying the activity of proteins, not DNA.[162] More importantly, they narrow the range of therapeutic options available, thus reducing the number of viable needles in the genomics needle stack.

Biotechnology is still a science dominated by statistics and probabilities, rather than one driven by deterministic models and a rigorous understanding of human biology. Genetic data are therefore the starting point for more arduous and extended research. More importantly for the present discussion, the dichotomy between genetic-data production and invention creates an environment in which research opportunities are, as a practical matter, unbounded because they far exceed the capacities of the scientific community.

---

160. Austin et al., *supra* note 101, at 1139 (observing that drug candidates typically will have to be modified to ensure that they have the proper pharmacokinetic (that is, absorption and permeation) and metabolic properties); Hopkins & Groom, *supra* note 101, at 727 (describing the attributes of chemicals necessary for them to be good candidates for drugs).

161. Hopkins & Groom, *supra* note 101, at 727-28 (noting, for example, that "many targets have failed to show any evidence of binding compounds that are potent and 'drug-like'").

162. *Id.* at 727. However, scientists are working towards expanding what has become known as the "druggable genome." Austin et al., *supra* note 101, at 1138.

III. THE INTERPLAY BETWEEN BIOMEDICAL SCIENCE AND BIOTECH PATENTING

Biotech patenting has evolved in seemingly unpredictable ways when viewed from a purely legal perspective. As one might anticipate from its importance in the biomedical sector, patents have been pursued aggressively from the start, beginning with the seminal Cohen-Boyer process.[163] This patent set the stage for the subsequent surge in patenting of research tools, which arguably reached its apex in the mid-1990s with the rush to patent DNA probes. Recently, the number of patent applications on gene sequences has flattened and even begun to decline.[164] Further, while concerns have been raised about patents on important drug targets, significant patent anticommons have not emerged. To the contrary, there has been a rise in the dedication of research tools to the public domain.[165] None of the legal theories anticipated these developments.

Developments in the underlying science, as well as its limitations, provide a number of important insights into the evolution of biotech patenting and licensing. The single most important factor is the most obvious one. Research and development in the biomedical sciences are shaped by the high cost and uncertainty of biotech methods, which derive from the extensive trial-and-error research required to evaluate the large number of potential drug targets available and to navigate the complexity of the biochemical interactions involved.

However, the biological complexity that makes discovery so challenging also mitigates the potential for patents to create broad monopoly power. The diversity of human biology makes the biomedical sciences relatively open-ended and less susceptible to patent anticommons. Many biological systems have built-in redundancies that protect against failures of specific processes, and this redundancy is more prevalent in more important processes. For example, DNA repair includes these types of parallel functions.[166] Further, scientists have found a huge range of genetic variants and a multigenic origin for common diseases. This complexity

163. Rai & Eisenberg, *supra* note 4, at 293-94.

164. FINNEGAN, HENDERSON, FARABOW, GARRETT & DUNNER, LLP, *supra* note 63, at 6-10. The arguments made here do not presume that this trend in biotech patent filings is attributable solely to factors emanating from the underlying science. Clearly a number of factors, particularly the heightened standard for utility promulgated by the PTO in 2001, affect biotech patent filing and issuance as well. *See* Utility Examination Guidelines, 66 Fed. Reg. 1092, 1098-99 (Jan. 5, 2001).

165. *See* discussion *supra* Part I.B.

166. *See* discussion *supra* Part II.B.

defies an atomistic, gene-by-gene analysis of disease processes.[167] More importantly for patent policy, common diseases will, as a consequence, be associated with multiple pathways or molecules, implying that most important diseases will have numerous potential drug targets.[168] Thus, by both affording numerous opportunities for research and a variety of treatment options, the complexity of biological processes reduces the potential for conflict between patenting and biomedical innovation.

The Walsh Study corroborates the importance of complex biological traits. Respondents acknowledged that few, if any, common diseases will have only a single drug target, and one commented that "we have more [drug] targets than [personnel needed] to work on them."[169] The increasing dedication of information to the public domain also reflects the fruitlessness of protecting overabundant research tools. As another respondent in the Walsh Study observed, dedicating research tools, including drug targets, to the public domain likely benefits established pharmaceutical companies.[170] By making the information freely available, pharmaceutical companies that are generally better positioned to exploit it are spared the cost of acquiring rights to use research tools from smaller biotech companies.[171]

Biomedical science therefore plays an important role in shaping patent strategy and business models in the biotech sector. These effects are evident in the evolution of biotech patenting and changing biotech business patterns. This Part returns to the legal debate, recasting it in light of the scientific constraints that have been discussed above. Section A examines the traditional commons metaphor in light of the open-ended nature of biomedical science. Section B draws on this discussion, but focuses on factors that mitigate the threats from patentees limiting access to key common-method research tools.

---

167. *See* discussion *supra* note 116.

168. Walsh et al., *supra* note 5, at 323-24; Weiss & Terwilliger, *supra* 120, at 153 (discussing how common disease traits will be associated with many potential candidate genes that scientists will have to investigate). Examples of the richness of potential drug targets include the DNA repair processes discussed in Part II.B. Indeed, even relatively simple diseases like malaria reveal complex underlying genetic susceptibilities that offer several potential targets. *See* discussion *supra* Part II.B.

169. Walsh et al., *supra* note 5, at 304-05, 324.

170. *Id.* at 329.

171. *Id.* Merges has argued that there is a natural counterbalancing that occurs in markets between those interests that are seeking to privatize certain types of inventions and those whose business interests actually favor dedicating them to the public domain. Merges, *supra* note 97, at 185-86. Merges claims that recent efforts to dedicate biotech data and materials to the public domain fit this model. Merges, *supra* note 97, at 185-86.

A. The Open Frontier of Biomedical Science and the Public Commons

The metaphor of the finite public commons provides the principal conceptual framework for biotech patent policy. Prospect theory employed the commons metaphor to argue for broad patents.[172] Merges and Nelson argued for granting narrow patents on the ground that the knowledge commons, unlike traditional physical commons, cannot be overexploited and the belief that the rate of innovation increases with the number of investigators.[173] Finally, Heller and Eisenberg used the commons metaphor to expose the risk that highly fragmented and broadly dispersed patent rights can impede innovation.[174]

All three theories assume implicitly that the underlying science is strictly finite and congested (or congestible).[175] Biomedical science is distinct in that, although some types of research tools are not plentiful, the many potential avenues for research create conditions in which others are practically unbounded—at least at this time. In this context, two types of research tools exist: (1) the relatively small number of common methods (for example, Cohen-Boyer, Kohler-Milstein, and PCR) that are critical to a broad range of biotech research, and (2) problem-specific tools that are plentiful (for example, ESTs, SNPs, and drug targets).[176] The differences between the two classes of research tools are critical to patent policy. Restricting access to patented common-method tools has the potential to im-

---

172. *See* discussion *supra* Part I.A.

173. Mergers & Nelson, *supra* note 2, at 843-44, 884. They also singled out biotechnology as a field in which particular care should be taken to constrain the scope of patents because such science-based technologies are particularly vulnerable to the negative effects of broad pioneer patents. *Id.* at 915.

174. Heller & Eisenberg, *supra* note 1, at 698.

175. Prospect theory presumes that biotech science can be bounded and managed. Merges and Nelson's proposal assumes some bounding and congestion; otherwise their advocacy of narrow patents is superfluous. Finally, Heller and Eisenberg's theory is explicitly premised on intellectual resources being "scarce," and in any event, only makes sense if the commons are both finite and seriously congested.

176. Admittedly borderline or variable cases will exist. For example, certain "promoters," which are DNA segments that turn genes on and off, are probably best categorized as common-method research tools because they are so broadly used. Powerful constitutive promoters, that is, those that cause a gene to be continuously expressed in all cell types, would fall into this category because they are so broadly applicable and valuable in genetic engineering. However, other promoters, such as those that are cell-specific or activated only under certain circumstances, may be more accurately categorized as problem-specific because their activation properties are so circumscribed. Judgments about how to categorize research tools will therefore by necessity have to take into account context, which will naturally change over time.

pede scientific research and innovation, whereas limiting access to problem-specific tools, because of their abundance, is unlikely to negatively affect biotech innovation.[177]

Two distinct policy regimes emerge from the two categories of research tools. The first regime falls within traditional patent theorizing, where a difficult balance must be struck between open access to key upstream research tools and maintaining the incentives to develop research tools in the first place. However, the risk here is not of an anticommons emerging, but whether access to uniquely powerful technologies will be limited. Because patents on common-method research tools are essential to conducting a wide range of research activities, such tools raise the most difficult challenges for balancing patent incentives and protecting scientific norms.[178] A large body of legal scholarship already exists on these issues, and I will not address them here. Instead, the section that follows discusses several scientific factors that mitigate—but by no means negate—the potential for biotech patents to limit access to key common-method research tools.

The second category deviates significantly from standard public commons-based policy arguments. The unbounded nature of the field—scientists have more drug targets than they know what to do with—neutralizes the central problem created by a finite public commons. In the traditional commons scenario, individual self-interest inexorably leads to overexploitation of the resource. However, the tragedy of the commons disappears when individuals cannot collectively overexploit a common resource.[179] For similar reasons, the threat of an anticommons emerging is also neutralized, as areas of dense patenting can always be avoided for uncongested resources. In fact, one might argue that the broad distribution of research activity caused by extensive patenting is a positive outcome. The abundance of problem-specific research tools also has the salutary effect of diminishing the value of those patents, consistent with recent trends toward dedicating these types of research tools to the public domain.[180]

---

177. Rose, *supra* note 8, at 5-7 (noting that uncongested resources do not suffer from commons problem by virtue of the fact that they cannot be adversely impacted by individual actions—control of the resource is beyond the capacity of either individual or collective action).

178. *See, e.g.*, NAS REPORT, *supra* note 20, at 62; Rai & Eisenberg, *supra* note 4, at 302.

179. Rose, *supra* note 8, at 5-7.

180. *See* discussion *supra* Part I.B. The reduced value of the patents also has important implications for cases, which apparently occur routinely, where a successful scientist conducts research without obtaining a license. The low value of the patents limits the damages a patentee could extract from an inventor in a patent infringement suit after a

The debate over biotech patent policy and fears about patent anticommons center on speculative biotech patenting of problem-specific research tools. Understanding the unbounded nature of biotech research reveals the irrationality of this form of strategic patenting. Inventors cannot predict ex ante which problem-specific research tools will be valuable for drug development. Further, many problem-specific research tools exist for prospective research, but only a tiny subset will be necessary for the development of a viable drug product. In this regard, the current state of biotech research and development represents the worst conditions for strategic patenting. If the number of potentially valuable patents were relatively circumscribed and the potential value of any given research tool still highly uncertain, inventors could hedge their bets through expansive speculative patenting. Here, however, the number of patentable research tools is virtually unlimited, making the expected utility of such a strategy diminishingly small.[181] Indeed, the most economically viable option for such research tools is licensing them for use in commercial microarrays used in biological assays, where they may end up being one of many thousands of probes, as opposed to being a high-value application.

The variety of potential approaches to studying or treating a disease creates a further disincentive for speculative patenting. As respondents in the Walsh Study observed,[182] the redundancy and intricacy of biological processes allow for multiple lines of research that enable scientists to circumvent existing problem-specific patents. Thus, although biological complexity offers many opportunities for strategic patenting, potentially enhancing the likelihood that a patent anticommons will emerge, this characteristic is a double-edged sword: it also affords many potential routes for engineering around existing patents.[183] Patentees, as a result, cannot be sure that their patent rights will be sufficient to exclude competitors because little will be known about the relevant biological processes when a speculative gene patent is first filed.

---

valuable product is discovered. Other factors also mitigate against such windfalls occurring, most important among these being the extensive research that will be required even after a research tool (such as a drug target) is discovered and the weak connection that will exist between the research tool and the final product, whose discovery will only be nominally attributable to the research itself.

181. *See* discussion *supra* Part II.A. This is not to suggest that there may be localized areas of scarcity (for example, access to gene therapy patents), but this is not inherently a commons problem, or at least it seems that the commons metaphor is inapt. More importantly, the data that do exist suggest that such problems have yet to emerge. *See* discussion *supra* Part I.B.

182. *See* discussion *supra* Part I.B.

183. *See* discussion *supra* Part I.C.

Finally, enforcement of problem-specific research tools is challenging. Except where a sequence is used as a probe in a commercial microarray assay, detection of infringing uses of a problem-specific research tool will be costly and onerous. In the absence of an infringing product or sale, infringing uses will occur in specific labs and generally will be undetectable in the publicly available research produced. This situation holds particularly true if the lab is constructing its own microarrays in which the patented research tool is just one of hundreds or thousands of probes. Where the specific sequence is integral to the reported results, the rare case, infringement may be detectable without gaining access to the lab itself.[184] Moreover, the ex ante value of the patented sequence, presumably in microarrays, will be small, making damage claims commensurately modest and reducing the incentive to spend valuable time and money enforcing speculative patents in the first place. None of these factors supports speculative genetic patenting as a viable business model. The Walsh Study's failure to identify significant anticommons problem further confirms that speculative patenting is far less of a threat than legal commentators have presumed.

Biomedical science remains a relatively unexplored territory in which the frontier is nowhere near an obvious geographical boundary. The standard commons arguments therefore simply do not apply. This framework explains why anticommons have not been a major factor in biotech patenting, and why they are unlikely to arise anytime soon. Simply put, the public commons model at the heart of the debate over biotech patent policy must be readjusted to reflect the important respects in which biotech patenting is uncongested and biomedical science is unbounded at this stage of its development. On the other hand, the biotech analog of a railroad or other technology providing access to this emerging territory, serves a unique purpose and is limited in number. Accordingly, just as control of railroads determined who had access to the American West, patents on common-method research tools can be used to restrict access to emerging areas of biomedical research.

B. Scientific Factors Mitigating the Detrimental Limits on Access to Patented Upstream Biotechnologies

The combination of common-method and problem-specific research tools has two practical implications for patent policy. First, fears about the patenting of abundant problem-specific research tools are unwarranted be-

184. Indeed, for problem-specific research tools, it is not even clear whether the patentee could prevent production of a down-stream drug product, which in most cases will be distinct from the research tool itself.

cause the public commons-based arguments that have provoked concern erroneously assume that biomedical science is a bounded and congested resource. This conclusion is borne out not only by the move to dedicate these types of research tools to the public domain, but also by the rapid growth in the number of problem-specific research tools over the past few years.[185] Second, as a number of commentators have recognized, patents on common-method research tools do present potentially significant risks to innovation and warrant continuing scrutiny.[186] Nevertheless, while these risks are substantial, several intrinsic scientific factors mitigate this event. The influence of these factors, as I argue below, turns on the fact that powerful common-method research tools typically have many nonrivalrous uses.[187]

This Section draws on two examples to examine the factors that mitigate against patents on common-method research tools being used to exclude access to them altogether. The first example involves a class of proteins, the Nuclear Factor κB (NF-κB) family, which aids in the regulation of several important biochemical pathways. The second example is the CD34 antigen, which has significant applications in stem cell research and cancer treatment. Both of these research tools have a broad range of potential applications that make them suitable for treatment as common-methods. They also have been cited repeatedly in the legal literature and have been used as examples of how biotech patents threaten research and development.[188]

1. The NF-κB Patent

In 2002, a patent on methods for using the NF-κB proteins issued and was exclusively licensed to Ariad Pharmaceuticals ("Ariad"), which has granted nonexclusive licenses and sought royalty payments from numer-

185. The human genome contains 3 billion base pairs and an estimated 30,000 to 35,000 genes, and rapidly evolving genomics methods are generating information at a stunning speed. Guttmacher & Collins, *supra* note 113, at 1514.

186. The NAS Report, as well as other commentators, also identifies this second issue as one deserving special attention. NAS REPORT, *supra* note 20, at 59-63. Robert Merges and Richard Nelson also highlighted this as a major issue years ago. Merges & Nelson, *supra* note 2, at 843-44.

187. I am using "rivalrous" in the sense that a rivalrous use would involve applications of patented technology in the same market(s), whereas a nonrivalrous application would arise in a distinct market. Uses of certain proteins, for example, can span completely different disease categories.

188. *See, e.g.*, Avital Bar-Shalom & Robert Cook-Deegan, *Patents and Innovation in Cancer Therapeutics: Lessons from* CellPro, 80 MILBANK Q. 637, 638 (2002) (specifically discussing the CD34 patents); Rai & Eisenberg, *supra* note 4, at 302.

ous companies.[189] NF-κB proteins are important in the regulation of a broad array of critical processes in mammals, including cell death, cell proliferation, immune responses, and tumor development,[190] and are consequently implicated in a large number of diseases, including cancer and immunological deficiencies.[191] For this reason, and because of their regulatory functions (that is, turning important biological processes on and off), the NF-κB proteins are uniquely powerful research tools.

The NF-κB patent claims more than a dozen methods of treating and studying diseases using NF-κB regulatory proteins.[192] The patent claims are very broad, covering essentially all of the potential methods for using the NF-κB proteins conceived at the time the patent was filed—despite the fact that patents on drugs that operated via some of these same mechanisms preceded the filing date of the NF-κB patent.[193] Moreover, the patent is broad, not only because the claims anticipate the many ways in which NF-κB proteins can be used, but also because the NF-κB proteins help regulate several critically important biological pathways. Nevertheless, the NF-κB patent is limited in two key respects: it covers neither the NF-κB proteins themselves nor the various biological pathways in which they are active.[194] In essence, the NF-κB patent claims a number of general recipes for using NF-κB proteins for therapeutic and research purposes.

Two standard objections exist for the NF-κB patent. First, it marries standard methods with a novel compound of unique biomedical importance. The inventors' primary contribution was in discovering the biological role of the NF-κB proteins. By contrast, the specific methods claimed by the patent were not novel. Once the function of the NF-κB proteins was understood, it naturally revealed how the proteins could be used in biomedical research and treatments. Second, the NF-κB patent is an upstream patent on basic, government-financed research that is far removed from practical applications. Put another way, the patent represents a clever way

189. U.S. Patent No. 6,410,516 (issued Jun. 25, 2002) [hereinafter "'516 patent"].

190. Diana Bolotin & Elaine Fuchs, *More Than Skin Deep*, 421 NATURE 594, 594 (2003); Michael Karin et al., *NF-κB in Cancer: From Innocent Bystander to Major Culprit*, 2 NATURE REV. CANCER 301, 301 (2002) ("[NF-κB] is not a single protein, but a small menagerie of closely related protein dimers that bind a common sequence motif known as the κB site.").

191. Karin et al., *supra* note 190, at 302.

192. Naomi Aoki, *Patenting Biological Pathways Ariad Suit Raises Questions on Laying Claim to Processes in Treating Disease*, BOSTON GLOBE, July 24, 2002, at C4.

193. '516 patent, *supra* note 189, cols.82-90.

194. *See id.*

of indirectly claiming a conceptual advance that, once made, had readily foreseeable practical implications.[195] As such, the NF-κB patent is a perfect example of a toll levied on upstream research that would have occurred even if patenting was not possible. These objections raise difficult policy questions, and I have no wish to diminish their significance or merit. My interest here, however, centers on whether access to the NF-κB proteins in fact will be restricted and, if so, how the restriction will affect biotech research and innovation.

The numerous methods covered by the NF-κB patent ensure that a wide range of nonrivalrous research is made possible using the patented methods. This finding is consistent with Ariad's efforts to license the methods broadly, which follows naturally from the fact that Ariad could not possibly support (or limit) all of the feasible related research.[196] NF-κB proteins are exceptional, both with respect to their specific regulatory functions and the diversity of biological pathways they affect. These attributes make them a distinctive research tool, but not because they are likely to be the only drug target available. To the contrary, the complex diseases and number of biological pathways with which NF-κB proteins are associated imply that multiple targets will exist.[197] Moreover, scientists are discovering that the intricate network of NF-κB interactions has its downsides, as it can complicate research and diminish the value of the NF-κB proteins as drug targets.[198] These factors, and the cost and uncertainty of biotech research and development, stand to mitigate the potential negative impacts of the NF-κB patent. In short, the scope of the patent, which derives from the diverse effects of NF-κB proteins, has a significant downside—precisely because they interact with multiple processes, using NF-κB proteins clinically is far more complicated and unpredictable. These realities will limit any tendency that Ariad may have to limit access to or charge exorbitant licensing fees on its patented methods.

2. The CD34 Antigen Patent

Researchers at Johns Hopkins University patented and later licensed the CD34 antigen to Baxter Healthcare Corp. ("Baxter").[199] The CD34 an-

---

195. *See* Merges & Nelson, *supra* note 2, at 906-07.

196. Aoki, *supra* note 192, at C4.

197. *See supra* Part II.B.

198. Bolotin & Fuchs, *supra* note 190, at 595. Scientists, for example, have found that the many complex interactions associated with NF-κB "raise caveats about the use of NF-κB inhibitors to treat cancer." *Id.*

199. U.S. Patent No. 4,965,204 (issued Oct. 23, 1990) [hereinafter "CD34 patent"]; Bar-Shalom & Cook-Deegan, *supra* note 188, at 639-40. Johns Hopkins University actu-

tigen is found on the surface of human stem cells, which are immature cells that can be transformed into a variety of human tissues.[200] This plasticity makes stem cells potentially valuable for treatment of individuals with cancer, Alzheimer's disease, and other conditions.[201] However, one of the major challenges in using stem cells is their scarcity. Analogous to the Cohen-Boyer process, the CD34 antigen overcomes this problem by providing a generic means of purifying stem cells. Because CD34 antigens exist only on stem cells, the antigens provide an ideal molecular target for isolating stem cells from a sample dominated by other cell types. In this process, stems cells are purified from a sample containing many cell types by tagging them with an antibody that selectively binds to CD34 antigen.[202]

The principal claim of the CD34 patent covers *all* antibodies that selectively bind to the CD34 antigen.[203] In the patent specification, this genus claim is supported by a single example, the My-10 antibody. The open-ended nature of this genus claim constitutes its most troubling aspect—it implicates both undiscovered binding sites on the CD34 antigen and unknown antibodies, created by scientists, that bind to the CD34 antigen. Tellingly, much of the frustration associated with the litigation over the CD34 patent[204] derived from the perception that the patent's coverage is too broad,[205] to which the CellPro litigation lent further support because the defendant had independently discovered a distinct antibody that bound

---

ally obtained four patents on the CD34 antigen. My focus here is on just one of these, the CD34 patent.

200. Johns Hopkins Univ. v. Baxter Healthcare Corp., 152 F.3d 1342, 1346 (Fed. Cir. 1998).

201. Bar-Shalom & Cook-Deegan, *supra* note 188, at 640-41; Jennifer L. Enmon, *Stem Cell Research: Is the Law Preventing Progress?*, 2002 UTAH L. REV. 621, 624.

202. Bar-Shalom & Cook-Deegan, *supra* note 188, at 640-41. The basic process is simple: (1) a sample of cells is treated with an antibody that binds to the CD34 antigen; (2) the stem cells are isolated by selectively binding them to a filter based on the properties of the bound antibody; and (3) the stem cells are released in a purified form from the filter. CD34 patent, *supra* note 199.

203. CD34 patent, *supra* note 199, cols.20-22; Bar-Shalom & Cook-Deegan, *supra* note 188, at 641.

204. *See, e.g.*, Johns Hopkins Univ. v. CellPro, Inc., 152 F.3d 1342 (Fed. Cir. 1998); Johns Hopkins Univ. v. CellPro, Inc., 931 F. Supp. 303 (D. Del. 1996).

205. *See CellPro*, 152 F.3d at 1350-51; Bar-Shalom & Cook-Deegan, *supra* note 188, at 643-44. The process the defendant used to purify stems cells also represented an improvement beyond that covered by the Johns Hopkins patents. *CellPro*, 152 F.3d at 1350-51.

to an entirely new (and unanticipated) binding site on the CD34 antigen and had superior binding properties.[206]

Concerns about the CD34 patent derive from its potential to restrict access to purified stem cells, not because it has particular biological significance in and of itself. The potential for the CD34 patent to impact biotech innovation consequently arises at two levels. First, Baxter will have a strong incentive to restrict development of any other antibodies that could compete with its My-10 antibody. Baxter will therefore use its patent to control the market for CD34 stem cell purification processes. This market was at issue in the protracted CellPro litigation over the CD34 patent.[207] Consequently, to the extent that Baxter's strategy retards development of new antibodies and purification methods using the CD34 antigen, biotech research and development will be impacted negatively. However, the CD34 patent is not meaningfully distinct from patents on many patented research instruments, which have not been subject to significant criticism in the legal literature. Moreover, to the extent that there is reason for heightened concern, it does not derive from the CD34 patent being an upstream patent per se, but rather from it being unjustifiably broad on substantive technical grounds.

Second, Baxter will have a strong incentive to make the CD34 stem cell purification process broadly available for a fee. In light of the company's actual business interests, it simply is not plausible that Baxter would limit access to CD34 purification process. The CD34 antigen is found on all stem cells, which have a diverse range of applications that no single entity could possibly fully exploit; similar to the NF-κB example, there will be many nonrivalrous uses of the technology. Further, the large number of applications for stem cells also implies that the potential market for the CD34 purification process is significant. While this may evoke Kitchian prospect theory, the conditions under which biotech research and

206. *See CellPro*, 931 F. Supp. at 311-12; Bar-Shalom & Cook-Deegan, *supra* note 188, at 643-44. These differences reaffirm the fallacy of defining a biomolecular genus functionally when predicting the function of a biomolecule from its structure is more art than science.

207. The litigation over the CD34 patent ultimately led to the demise of CellPro and was driven by each company's interest in selling purification systems for purposes of administering purified stems cells to cancer patients. Bar-Shalom & Cook-Degan, *supra* note 188, at 638. Chemotherapy and radiation therapy, both of which are used to treat cancer patients, kill stem cells. *Id.* at 640-41. Scientists believed that purifying a patient's stems cells from tissue samples prior to cancer treatment could be used to replenish the bone marrow cells lost. *Id.* Interestingly, the demand for stem cell separation methods ultimately collapsed after scientists discovered that bone marrow transplantation had little benefit relative to its costs and toxicity. *Id.* at 655.

development occur mitigate the concerns raised by Merges and Nelson.[208] First, patentees will have a high level of technical sophistication, suggesting that they will appreciate the value of the nonrivalrous uses of their technology. Second, patentees will have strong economic incentives to license research-tool patents because of the long period that typically exists between development of research tools and actual product development; virtually any opportunity to cultivate an early-stage source of revenue will be welcome.

This example reveals an important dynamic: the broader the range of applications for a research tool, the less likely a patent owner will be able to exploit its research potential and the greater the market-size incentives will be to make the technology broadly available. As a consequence, access to research tools of broad importance to biomedical research and development is unlikely to be restricted. Nevertheless, patent premiums still could function as *de facto* restrictions on access, although concern about this occurrence is allayed somewhat by the lack of corroborating evidence.[209]

The preceding discussion illustrates the interplay among scientific factors, economics, and legal rules. The two categories that I have identified, common-method and problem-specific research tools, provide a useful framework for analyzing whether the patenting of specific research tools exerts adverse effects on biomedical research and development. It also provides several bases for understanding why biotech patenting has not conformed as predicted in the various competing legal theories. Analysis of the underlying science reveals several mitigating factors that explain why the rapid growth in biotech patenting has not led to the negative outcomes predicted. Although these findings provide grounds for optimism about biomedical innovation and hope that dedication of data and materials to the public domain will continue, my primary objective is to advance the current understanding of biotech patent dynamics, not to advocate the kind of rose-colored vision of biotech patent policy that I have criticized the scientific community for promoting about biotechnology.

IV. CONCLUSION

The legal debate over biotech patent policy rightly has focused attention on the patenting of important biotech research tools and the threats to innovation posed by patent anticommons. However, the important influ-

208. *See* discussion *supra* Part I.A.

209. Walsh et al., *supra* note 5, at 331-33.

ence of biomedical science itself on biotech patenting has been notably absent from this discourse.

I have sought to remedy this oversight by developing three central points. First, the uncertainty and complexity of biomedical science help explain several important trends in biotech patenting, including recent shifts towards free public access to important classes of research tools. Second, the standard finite commons model is not representative of the essentially unbounded opportunities in biotech research that exist at this early stage of development. Further, once the premise of a finite commons is abandoned, the potential for patent anticommons to emerge largely disappears and patents on most research tools pose less of a threat than predicted. Third, I identify two central classes of biotech research tools, common-method and problem-specific, but conclude that only patents on common-method tools pose potentially significant risks to biomedical innovation. However, even these risks can be mitigated by factors, such as the existence of nonrivalrous uses, that reduce the likelihood that access will be limited.

In sum, although it would clearly be wrong to infer that biotech patenting has no adverse effects on biomedical innovation, these findings suggest that the actual risks are significantly lower and more circumscribed than many have predicted. Thus, the current legal regime works reasonably well with respect to problem-specific research tools. Potential bottlenecks, however, exist with respect to common-method research tools. In particular, despite the mitigating factors identified above, the high-stakes nature of biomedical science suggests that targeted interventions likely will be needed to address the potential threats posed by patents on critical common-method research tools that lack nonrivalrous uses.

ADDRESSING GLOBAL HEALTH INEQUITIES: AN OPEN LICENSING APPROACH FOR UNIVERSITY INNOVATIONS

By Amy Kapczynski,[†] Samantha Chaifetz,[‡] Zachary Katz[‡‡] & Yochai Benkler[††]

TABLE OF CONTENTS

© Amy Kapczynski, Samantha Chaifetz, Zachary Katz & Yochai Benkler

† Post-Doctoral Fellow in Law and Public Health, Yale Law School, Yale School of Public Health.

‡ J.D. candidate, Yale Law School; Steering Committee, Universities Allied for Essential Medicines.

‡‡ J.D. candidate, Yale Law School.

†† Professor of Law, Yale Law School.

The authors would like to thank Professors Jack Balkin, Brook Baker, James Boyle, Harold Koh, Jean Lanjouw, and Arti Rai for their thoughtful comments. The authors also benefited from the help of numerous individuals, including university technology transfer officers, advocates, scientists, and academics who provided invaluable advice and counsel to them in their efforts to develop a model Equitable Access License. That license was the product of an independent, informal working group based at Yale University and convened by Universities Allied for Essential Medicines. The license can be found at Model Provisions for an "Equitable Access License" Version 1.0 (Jan. 2005), *available at* http://www.essentialmedicine.org/EAL.pdf.

I. INTRODUCTION

Each year, millions of people in low- and middle-income (LMI) countries die from preventable and treatable diseases.[1] AIDS provides one of the starkest examples: it killed more than three million people in 2004[2] and has become the world's leading cause of death for adults aged fifteen to fifty-nine.[3] These deaths continue despite the fact that we have known for years that antiretroviral combination therapy (ARVs) can substantially improve the lives of those living with HIV/AIDS, and even reverse the tide of death associated with the disease.[4]

But the drugs that we take for granted in the United States have long been out of reach for most of those living with HIV/AIDS around the

1. *See, e.g.*, WORLD HEALTH ORG., WORLD HEALTH REPORT 2002, at 186-91.

2. UNAIDS, AIDS EPIDEMIC UPDATE 1 (Dec. 2004) [hereinafter UNAIDS 2004], http://www.unaids.org/wad2004/EPI_1204_pdf_en/EpiUpdate04_en.pdf.

3. WORLD HEALTH ORG., KEY FACTS FROM THE WORLD HEALTH REPORT 2004, at 1, http://www.who.int/whr/2004/en/facts_en.pdf.

4. In the United States, in the two years after ARVs were adopted for widespread use, AIDS-related mortality dropped by more than seventy percent. *See* Frank J. Palella et al., *Declining Morbidity and Mortality Among Patients with Advanced Human Immunodeficiency Virus Infection*, 338 NEW ENG. J. MED. 853 (1998); *see also* Paulo R. Teixeira et al., *The Brazilian Experience in Providing Access to Antiretroviral Therapy*, *in* ECONOMICS OF AIDS AND ACCESS TO HIV/AIDS IN DEVELOPING COUNTRIES 69 (2003) (describing a similarly dramatic reduction in mortality in Brazil following the introduction of ARV therapy).

world.[5] One crucial reason has been their cost. In 2000, the average worldwide price for patented ARVs was more than $10,000 per patient per year.[6] Today, the same medicine is sold in generic form for as little as $168 per year.[7] This drastic reduction in price has enabled governments and international agencies to initiate programs designed to bring these medicines to millions of HIV-positive individuals around the world who otherwise lack access to them.[8] These programs still have a long way to go before they meet existing need,[9] but they would not have begun at all if prices had not come down so dramatically.

---

5. Approximately ninety-five percent of AIDS-related deaths occur in the developing world. *See* UNAIDS, AIDS EPIDEMIC UPDATE 5 (Dec. 2003), http://www.unaids.org/html/pub/publications/irc-pub06/jc943-epiupdate2003_en_pdf.pdf. In 2001, a survey of seventy low-income countries found that only two percent of those with advanced HIV infection had access to treatment. INT'L HIV TREATMENT ACCESS COALITION, WORLD HEALTH ORG., A COMMITMENT TO ACTION FOR EXPANDED ACCESS TO HIV/AIDS TREATMENT 2 (2002). The following year, the World Health Organization (WHO) reported that on average only five percent of all people in need of ARVs worldwide received them; within sub-Saharan Africa, just one percent were treated. INT'L HIV TREATMENT ACCESS COALITION, *supra*, at 1.

6. *See* MÉDECINS SANS FRONTIÈRES, UNTANGLING THE WEB OF PRICE REDUCTIONS: A PRICING GUIDE FOR THE PURCHASE OF ARVS FOR DEVELOPING COUNTRIES 9 (7th ed. 2005) [hereinafter MSF, UNTANGLING THE WEB], http://www.accessmed-msf.org/documents/untanglingtheweb%207.pdf. At the same time, in Brazil, generics were being produced for less than $3000 per patient per year. *Id.*

7. *Id.*

8. A number of national ARV programs explicitly rely on generics. *See, e.g.*, Charles Wendo, *Uganda Begins Distributing Free Antiretrovirals*, 363 LANCET 2062 (2004). International agencies have also found generics important to their program objectives. *See* Asia Russell, *The Bush Administration's Global AIDS Promises—and Praxis*, 4 YALE J. HEALTH POL'Y L. & ETHICS 133, 138 (2004) (citing GLOBAL FUND TO FIGHT AIDS, TUBERCULOSIS, & MALARIA, GUIDELINES FOR PROPOSALS (2003)); Keith Alcorn & Theo Smart, *Fixed Dose ARV Combinations: Choices and Challenges*, HIV & AIDS TREATMENT IN PRACTICE (NAM, United Kingdom, Mar. 2004) (noting that the WHO's 3 x 5 Initiative favors generics because they are believed to make the program affordable), *at* http://www.aidsmap.com/en/docs/3FE6E952-3B09-494A-96E0-200381027DA0.asp; *cf.* Juan Rovira, *Trade Agreements, Intellectual Property, and the Role of the World Bank in Improving Access to Medicines in Developing Countries*, 4 YALE J. HEALTH POL'Y L. & ETHICS 401 (2004).

9. The programs are, however, headed in the right direction. The number of people in developing countries receiving treatment purportedly increased by nearly two-thirds in the second half of 2004. *Compare* Press Release, WHO/UNAIDS/Global Fund/U.S. Government, 700,000 People Living with AIDS in Developing Countries Now Receiving Treatment (Jan. 26, 2005), http://www.who.int/mediacentre/news/releases/2005/pr07/en/print.html, *with* UNAIDS 2004, *supra* note 2, at 5 (reporting that 440,000 low- and middle-income country residents were receiving treatment as of June 2004, according to WHO statistics).

These recent price reductions have also generated a storm of controversy regarding the contribution that patents and other exclusive rights[10] make to the inequities in global availability of life-saving medicines. The problem that patents can pose for access to medicines and medical technologies[11] is complex and cannot be understood without a nuanced assessment of the political economy in which the key players operate.

Consider one example: in 2000, only an estimated one percent of the 500,000 South Africans in need of ARV medicines received them.[12] The humanitarian organization Médecins Sans Frontières (MSF), better known in the United States as Doctors Without Borders, wanted to begin a pilot program in a township outside Cape Town to demonstrate that, contrary to popular belief (and the claims of South African President Thabo Mbeki[13]), AIDS medicines could be used effectively in resource-poor settings if they could be made affordable.[14] MSF faced a practical problem: a limited budget and a seemingly unlimited supply of patients. In South Africa, the price of stavudine, just one of the drugs then used in ARV therapy, was over $1600 per year.[15] An Indian company offered to sell MSF generic stavudine for approximately three percent of the branded version's price,

10. In the pharmaceutical field, patents are increasingly supplemented by other exclusive rights, such as rights in regulatory data. For the sake of clarity, this Article will refer to the range of patent and patent-like exclusive rights that may apply to medical technology collectively as "patent rights." Also, when we discuss these exclusive rights we refer to them as they are usually used—that is, to secure a monopoly and extract supra-marginal returns.

11. We include the range of non-pharmaceutical products important to the practice of medicine, such as vaccines, diagnostics, and monitoring tools, when referring to medicines or medical or biomedical technologies.

12. *See* Jennifer Barrett, *A Major Step*, NEWSWEEK (WEB EXCLUSIVE), Nov. 24, 2003, *at* http://msnbc.msn.com/id/3606125; *see also* Edwin Cameron, *The Deafening Silence of AIDS*, 5 HEALTH & HUM. RTS. 7 (2000) (describing the lack of access to treatment in South Africa in 2000).

13. *Cf.* Barrett, *supra* note 12 (mentioning Mbeki's refusal to accept that antiretrovirals worked, as well as his subsequent reversal on the issue).

14. MSF and others have succeeded in establishing this principle. *See* Paul Farmer et al., *Community-Based Approaches to HIV Treatment in Resource-Poor Settings*, 358 LANCET 404 (2001); Toby Kasper et al., *Demystifying Antiretroviral Therapy in Resource-Poor Settings*, 32 ESSENTIAL DRUGS MONITOR 20 (2003); Donald G. McNeil Jr., *Africans Outdo U.S. Patients in Following AIDS Therapy*, N.Y. TIMES, Sept. 3, 2003, at A1.

15. Letter from Eric Goemaere, Representative of Médecins Sans Frontières—South Africa, to Jon Soderstrom, Managing Director, Office of Cooperative Research, Yale University (Mar. 9, 2001) [hereinafter Goemaere MSF Letter] (on file with authors).

but because the drug was subject to a South African patent, MSF could not legally accept the offer.[16]

Though Bristol-Myers Squibb (BMS) had an exclusive license to sell the drug, Yale University was the key patent holder.[17] MSF approached Yale in February 2001, requesting a license to use generic stavudine. MSF simultaneously asked BMS for a price reduction that would lower the price to the generic level.[18] In addition to its immediate, pragmatic objectives regarding its pilot program, MSF likely also had a broad strategic goal in mind. At that time, patent-based pharmaceutical firms[19] refused to offer transparent and comprehensive price reductions for AIDS drugs for developing countries,[20] threatened to sue generic companies that supplied ARVs to developing countries where the firms believed they held patents,[21] and sued the South African government over a statute intended to allow cheaper medicines into the country.[22] MSF no doubt hoped that Yale would act differently than the drug companies, setting a precedent that would ultimately demonstrate that, contrary to drug company assertions, price discounts and patent concessions in countries like South Africa would not destroy the patent-based pharmaceutical industry.[23]

---

16. *Id.*; *see also* Melody Petersen, *Lifting the Curtain on the Real Costs of Making AIDS Drugs*, N.Y. TIMES, Apr. 24, 2001, at C1 (noting that Cipla, an Indian generic company, had offered to sell generic stavudine to health organizations for $40 per year).

17. Bristol-Myers Squibb brought the drug (also known as d4t) to market in 1994 under the brand name Zerit. *See* John Curtis, *Hunting Down HIV*, YALE MED., Summer 1998, http://info.med.yale.edu/external/pubs/ym_su98/cover/cov_hunting11.html

18. Goemaere MSF Letter, *supra* note 15.

19. We use the terms "patent-based," "originator," or "proprietary" to denote pharmaceutical companies, including biotech firms, that develop, produce, and/or market patented medicines. The Pharmaceutical Research and Manufacturers of America (PhRMA) represents the interests of these companies in the United States, and "PhRMA companies" is another common synonym. These terms specifically exclude generic companies, both in the developed and developing worlds.

20. *See infra* notes 150-54 and accompanying text.

21. *See infra* note 160.

22. *See* Medicines and Related Substances Control Amendment Act, No. 90 (1997) (S. Afr.); *see also* Mark Heywood, Debunking 'Conglomo-Talk': A Case Study of the Amicus Curiae as an Instrument for Advocacy, Investigation and Mobilization, Paper Presentation at Health, Law and Human Rights: Exploring the Connections Conference 13 (Sept. 29, 2001), http://www.tac.org.za/Documents/MedicineActCourtCase/Debunking_Conglomo.rtf.

23. *Cf.* Barton Gellman, *A Turning Point That Left Millions Behind*, WASH. POST, Dec. 28, 2000, at A1 (citing the Chairman of Pfizer, in 2000, who argued in favor of a donation program in developing countries, instead of differential pricing or generic competition, for its important AIDS drug diflucan because the industry "lives and dies on intellectual property").

This approach paid off. Within weeks of receiving MSF's request, Yale and BMS jointly announced that they would permit the sale of generics in South Africa and that BMS would lower the price of its brand-name stavudine to approximately $55 per year throughout sub-Saharan Africa for governments and nongovernmental organizations (NGOs).[24]

The Yale/BMS decision garnered significant media attention,[25] and may have helped create a tipping point in the campaign for access to affordable ARVs—shortly after the announcement, pharmaceutical manufacturers dropped their lawsuit against the South African government.[26] Major price reductions in sub-Saharan Africa followed from other companies,[27] as well as additional concessions on intellectual property rights (IPRs).[28] This in turn enabled activists in countries such as South Africa to turn the spotlight on their own government's inaction, and eventually obtain commitments to provide ARVs in the public sector.[29]

---

24. *See* Press Release, Bristol-Myers Squibb, Bristol-Myers Squibb Announces Accelerated Program To Fight HIV/AIDS in Africa (Mar. 14, 2001), http://www.prnewswire.co.uk/cgi/news/release?id=64424. While a steep discount, this price was still higher than the price offered by generic companies. *See* Goemaere MSF Letter, *supra* note 15. Generic forms of stavudine have been available in South Africa since 2003, and two companies have been awarded a tender to supply generic d4t to the South African government. *See* Amy Kapczynski et al., Editorial, *Global Health and University Patents*, 300 SCIENCE 1629 (2003); *South African Generic Drug Maker To Produce Country's First Generic Antiretroviral Drug*, KAISER DAILY HIV/AIDS REP., Aug. 7, 2003, *at* http://kaisernetwork.org/daily_reports/rep_index.cfm?hint=1&DR_ID=19240; Press Release, South Africa Department of Health, ARV Drug Tender Awarded (Mar. 3, 2005), http://www.doh.gov.za/docs/pr/pr0303-f.html.

25. *See, e.g.*, Karen DeYoung & Bill Brubaker, *Another Firm Cuts HIV Drug Prices*, WASH. POST, Mar. 15, 2001, at A1; Michael Waldholz & Rachel Zimmerman, *Bristol-Myers Offers To Sell Two AIDS Drugs in Africa at Below Cost*, WALL ST. J., Mar. 15, 2001, at B1.

26. *See* Rachel L. Swarns, *Drug Makers Drop South Africa Suit over AIDS Medicine*, N.Y. TIMES, Apr. 20, 2001, at A1.

27. *See, e.g.*, *Drug Company Cuts AIDS Drug Prices in S. Africa*, REUTERS NEWSMEDIA, Nov. 30, 2001, http://www.emro.who.int/asd/WhatsNew-GlobalEvents-Reuters3011.htm; Rachel Zimmerman & Michael Waldholz, *Abbott To Cut Prices on AIDS Drugs Distributed in Sub-Saharan Africa*, WALL ST. J., Mar. 27, 2001, at A3

28. *See, e.g.*, Press Release, Aspen Pharmacare Ltd., Aspen Pharmacare Receive Voluntary License from GlaxoSmithKline on Anti-Retroviral Patents in South Africa (Oct. 8, 2001) [hereinafter Aspen Pharmacare Press Release], http://www.aspenpharmacare.co.za/showarticle.php?id=135.

29. Treatment Action Campaign (TAC), a South African NGO, for example, aligned with the ruling African National Congress (ANC) party during the drug company lawsuit. Directly after the victory, the ANC made clear that it had no plans to take advantage of the potential for lower prices by creating a national treatment program. *See* Ben Hirschler, *Glaxo Gives Up Rights to AIDS Drugs in South Africa*, REUTERS NEWSMEDIA,

In renouncing the enforcement of its South African patent, Yale went further in making intellectual property (IP) concessions on an AIDS medicine than any proprietor had done before, and demonstrated that patent holders could trigger substantial, immediate price reductions. The Yale/BMS deal may also have been a watershed event for public sector institutions (which we define as public and private universities, governmental agencies, and nonprofit organizations). Over the past few years some such institutions have taken steps to ensure that their patents do not contribute to what we call the "access gap"—the systematic inability of individuals in developing countries to obtain existing medicines.

Public sector institutions have also begun to address a related problem—one we term the "research and development gap" (R&D gap)—of massive underinvestment in medicines for diseases that primarily impact the global poor, known as neglected diseases.[30] The scale of the inequality is immense: "only 10% of the world expenditure on health R&D is spent on health conditions that represent 90% of the global [disease] burden . . . ."[31] Although the R&D gap has received far less attention than the access gap, its implications are no less grim.

Consider one vivid example: the most commonly used drug to treat African sleeping sickness is arsenic-based and kills up to five percent of

---

Oct. 6, 2001 (describing the government's resistance and TAC's response), http://www.aegis.com/news/re/2001/RE011009.html. TAC then shifted its focus to the government, filing and winning a landmark constitutional case establishing the government's obligation to create programs to provide medicines to HIV-positive women to prevent the transmission of HIV to their children. Minister of Health v. Treatment Action Campaign, 2002 (5) SALR 721 (CC) (S. Afr.). In 2003, the South African government finally launched a national antiretroviral program, employing generic stavudine as a key component of the formulary. *See* Julian Meldrum, *South African HIV Treatment To Depend on Generic Drugs*, AIDSMAP NEWS, Aug. 7, 2003, http://www.aidsmap.com/en/news/F5E96962-F1B4-40F2-8969-624AC8A7D424.asp. Approximately 27,000 people are reportedly now receiving treatment from the public sector. *See* Ben MacIennan, *Aids Activists Are Govt's 'Conscience'*, MAIL & GUARDIAN (S. Afr.), Feb. 16, 2005, http://www.mg.co.za/articlePage.aspx?articleid=197660&area=/breaking_news/breaking_news_national. While this represents a significant advance, it remains far shy of the government's announced intention to treat 50,000 people by the end of 2004. *See* Sharon LaFraniere, *South Africa Approves Plan To Offer Free AIDS Medication*, N.Y. TIMES, Nov. 20, 2003, at A3.

30. There is no standard definition of "neglected diseases," and other terms such as tropical diseases or developing-country diseases are often used interchangeably.

31. United Nations Dev. Programme, Incentives To Reduce the 10/90 Gap (2002), http://www.undp.org/ods/monterrey-sideevent/incentive.pdf; *see* GLOBAL FORUM FOR HEALTH RESEARCH, THE 10/90 REPORT ON HEALTH RESEARCH 2003-2004 (2004).

those who are treated with it.[32] New drugs are desperately needed for this and many other diseases, but have not been forthcoming. Of the many reasons for this, the most important is that our current drug development system primarily depends on patents (and their corresponding market-based incentives) to draw private companies to fund clinical trials and commercialization activities. Predictably, firms have little interest in developing products for developing countries because these markets are so small: the branded pharmaceutical industry in the United States derives only five to seven percent of its profits from all LMI countries.[33] Indeed, Latin America, Africa, Asia (excluding Japan), and Australia combined comprise only twelve percent of the *total* worldwide market for pharmaceuticals, including generic drugs.[34]

---

32. *See* Drugs for Neglected Diseases Initiative, Sleeping Sickness (Human African Trypanosomiasis), *at* http://www.dndi.org/cms/public_html/insidearticleListing.asp?CategoryId=89&SubCategoryId=147&ArticleId=201&TemplateId=1 (last visited Apr. 8, 2005).

33. These country classifications are made by the World Bank. *See* World Bank Group, Data and Statistics: Country Classification, *at* http://www.worldbank.org/data/countryclass/countryclass.html (last visited Apr. 28, 2005); World Bank Group, Data and Statistics: Country Groups, *at* http://www.worldbank.org/data/countryclass/classgroups.htm (last visited Apr. 28, 2005). In 2002, "94.9% of the global sales of the U.S.-based brand-name pharmaceutical industry came from the U.S., Canada, Europe (including Eastern Europe and Russia), Japan, Australia and New Zealand." William W. Fisher & Talha Syed, Patent Law, Drugs and the Health Crisis in the Developing World 76-77 (Feb. 24, 2005) (unpublished manuscript, on file with authors). The most recent report from the pharmaceutical industry's trade association, PhRMA, offers data that support the conclusion that LMI markets contribute five to seven percent of sales. *See* PHARM. RESEARCH & MFRS. OF AM., PHARMACEUTICAL INDUSTRY PROFILE 2005—FROM LABORATORY TO PATIENT: PATHWAYS TO BIOPHARMACEUTICAL INNOVATION 40 (2005) (including Latin America, Asia-Pacific—except Japan, India, and Pakistan—Central and Eastern Europe, Russia, and the Middle East in the estimation of LMI markets). PhRMA members represent a very large proportion of the patent-based industry, and of U.S. firms engaged in R&D. Its data is thus well-tailored for the purposes of this Article, and we are grateful to Talha Syed for directing us to it.

34. This percentage was steady between 2002 and 2004, according to IMS Health. *See* Press Release, IMS Health, IMS Reports 2004 Global Pharmaceutical Sales Grew 7 Percent to $550 Billion (Mar. 9, 2005) (indicating the percent did not change from 2003 to 2004), http://www.imshealth.com/ims/portal/front/articleC/0,2777,6599_3665_71496463,00.html; Press Release, IMS Health, IMS Reports 8 Percent Constant Dollar Growth in 2002 Audited Global Pharmaceutical Sales to $400.6 Billion (Feb. 25, 2003), http://www.imshealth.com/ims/portal/front/articleC/0,2777,6599_3665_41336931,00.html. Australia, which is of course not an LMI country, is likely a somewhat significant share of this percentage, but IMS Health does not publicly provide these percentages disaggregated by country. The main distinction between this and PhRMA data is due to IMS Health's inclusion of generic sales.

While patents—and the promise of exclusivity—alone cannot stimulate research where there is no attractive market for a medicine, they can create barriers to such research, and thus play a role in perpetuating the R&D gap. Public sector institutions are beginning to address both the need to stimulate research and to ensure patents do not block research, for example, by participating in public-private partnerships to develop medicines for neglected diseases and by seeking to reserve rights to use one another's research tools.

This Article aims to draw upon these examples to demonstrate the potential of public sector institutions, particularly U.S. universities,[35] to address the access and R&D gaps by changing their licensing practices. It also aims to propose a strategy that will allow these institutions to settle on a standard practice that will best use their collective contribution to innovation. We demonstrate that without any changes in the current statutory or regulatory environment, these institutions can use private, contractual instruments to foster commons-based remedies for the problems of our global drug development and distribution system. This Article outlines the structure and characteristics of two such instruments to: (1) eliminate the access barriers exclusive rights pose to patients in LMI countries, and (2) remove patent barriers that might impede research on neglected diseases.

These strategies will only be effective if they are rooted in an understanding of the role that IPRs play in the access and R&D gaps. Part II provides such an analysis and argues that patents and other exclusive rights regimes are now an essential subject for anyone concerned with global health disparities. We also review existing proposals to eliminate the burdens patents can impose on the global poor, and demonstrate the acute need for new approaches—particularly ones that can route around the inaction of governments and firms, and have a demonstration effect that will prompt systemic change.

Part III seeks to define the space from which such an approach can emerge. We discuss a class of commons-based initiatives that provide a model for action that depends neither on top-down governmental action, nor on private market motives and signals.[36] Commons and common-property regimes in material resources have been the subject of substantial

35. While our proposals could be adopted by nonprofits, universities, and even private firms both within and outside of the United States, we concentrate our discussion on U.S. universities.

36. *See* Yochai Benkler, *Commons-Based Strategies and the Problems of Patents*, 305 SCIENCE 1110 (2004).

scholarship over the past two decades.[37] The emergence of free and open source software development has led to increased interest in defining the conditions for sustainable and successful nonproprietary production strategies—for software[38] and more generally for networked information production[39] and some classes of physical resources.[40] These approaches rely on mechanisms other than proprietary exclusion to motivate and to organize production, and they frequently rely upon innovative contractual provisions to create a self-perpetuating commons. In this Part, we discuss recent, exploratory public-sector projects that implement commons-based approaches within the biomedical domain, as well as the models and lessons these projects can take from other commons-based, contractually structured initiatives.

In Part IV, we analyze the current structure of university research and technology commercialization, demonstrating that U.S. research universities are well-positioned to adopt open licensing policies[41] that could meaningfully benefit the global poor. This Part explains the role of universities in the overall biomedical innovation system and discusses the problems with their standard approach to patenting and licensing biomedical innovations. We also map the political economy of a move towards open licensing within universities, demonstrating that such licensing is not contrary to the financial interests of universities, and may in fact provide substantial gains for universities as well as the global poor. The success of this pro-

37. *See, e.g.*, ELINOR OSTROM, GOVERNING THE COMMONS: THE EVOLUTION OF INSTITUTIONS FOR COLLECTIVE ACTION (1990); Carol Rose, *The Comedy of the Commons: Custom, Commerce, and Inherently Public Property*, 53 U. CHI. L. REV. 711 (1986).

38. Open source or free software innovation has attracted significant academic attention, as has peer production of other types of information, knowledge, and culture more generally. *See* STEVEN WEBER, THE SUCCESS OF OPEN SOURCE (2004); Josh Lerner & Jean Tirole, *The Scope of Open Source Licensing*, 21 J.L. ECON. & ORG. 20 (2005); Eric von Hippel & Georg von Krogh, *Open Source Software and the Private-Collective Innovation Model: Issues for Organization Science*, 14 ORG. SCI. 209 (2003).

39. *See, e.g.*, Yochai Benkler, *Coase's Penguin, or, Linux and* The Nature of the Firm, 112 YALE L.J. 369 (2002).

40. *See, e.g.*, Yochai Benkler, *Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production*, 114 YALE L.J. 273 (2004).

41. We define an "open" licensing provision as one that is available to everyone on the same terms. In this sense open licensing is not the same as dedication to the public domain. A self-reinforcing licensing approach that employs patent and other rights—rather than simply dedicating innovations to the public domain—may be necessary to sustain a commons where key institutional players, including national governments and private-sector firms, are intent on promoting the expansion and utilization of exclusive rights.

posal will depend on its adoption by a critical mass of research universities.

In Part V, we propose two open licensing models that universities (and other institutions) can adopt to improve access to biomedical innovations in LMI countries. We call the first approach Equitable Access ("EA") Licensing. The approach relies on including EA clauses in the technology transfer licenses universities negotiate with drug companies engaged in commercializing the universities' academic discoveries. The EA provisions we propose give third parties—for example, manufacturers of generic medicines—freedom to operate in LMI countries with regard to the licensed technology or any derivative products, by adapting the so-called "copyleft" characteristics of some open source licenses.[42] EA clauses also establish a self-enforcing open licensing regime that minimizes transaction costs and is insulated from the vicissitudes of internal university politics and market relationships.

We refer to the second open licensing approach as Neglected Disease ("ND") Licensing. Like EA licensing, ND licensing is a commons-based strategy. ND clauses are designed to provide those engaged in neglected disease research the freedom to experiment on and with proprietary university technologies. Furthermore, ND clauses allow researchers to freely market, in LMI countries, any innovations without hindrance from exclusive rights held by the university.

One of the lessons of our analysis of various licensing provisions is that there is no one-size-fits-all commons-based strategy. Different strategies to create and sustain commons-based production in different contexts may be required by: different economic characteristics of research areas; different industrial structures and relative roles of market-based, governmental, and nonprofit enterprises; and different types of exclusive rights regimes.

Our proposal is deliberately modest. We suggest an intervention in the existing industrial structure of the research, development, manufacture, and distribution of curative and preventive treatments. But the intervention we advocate is not aimed at fundamentally restructuring these fields. Instead, we suggest taking advantage of the existing distribution of firms and

42. EA licensing is not truly an "open source" strategy—a term that describes software for which source code is made freely available to independent software developers. Nonetheless, it mimics open source software's approach to IPRs by ensuring that the licensed technology and subsequent developments remain freely available to all potential users under an EA license. While other open licensing models typically offer the freedom to operate in all markets, including high income markets, our proposal is restricted to low- and middle-income settings.

business models, the relatively large role of public sector institutions, and the distribution of needs, wealth, and markets. Our proposal is intended to complement, rather than displace, current proposals to reorganize the market for drug development through top-down legislative change. The shift we describe provides a way for organizations and firms to take immediate steps to positively affect the lives of patients in LMI countries, and perhaps to catalyze broader action to promote global health.

II. HOW PATENTS AND OTHER EXCLUSIVE RIGHTS AFFECT THE GLOBAL ACCESS AND R&D GAPS

A. Innovation Theory and the Second Enclosure Movement

The past two decades have witnessed a steady, global trend toward ever more restrictive patent and related exclusive rights regimes, dubbed the "second enclosure movement."[43] In the United States, for example, the scope of patentability has expanded to include bioengineered organisms and purified genetic material,[44] and "early 'upstream' inventions that explain disease pathways and mechanisms and identify potential drug targets are increasingly likely to be patented."[45] Patents have also been increasingly supplemented with exclusivity offered at the drug regulatory interface.[46] Over the same period, the United States, the European Union, and Japan have used trade agreements to impose high levels of substantive and procedural protection for IP on countries around the world. The World

43. *See* James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 LAW & CONTEMP. PROBS. 33, 39 (2003).

44. In 1980, the Supreme Court held that genetically engineered microorganisms could be patented. Diamond v. Chakrabarty, 447 U.S. 303 (1980). In 1988, the Patent and Trademark Office (PTO) granted its first patent on a four-legged animal, Harvard University's OncoMouse. *See* U.S. Patent No. 4,736,866 (issued Apr. 12, 1988). Currently, the PTO regularly grants patents on isolated and purified versions of naturally occurring DNA fragments and other biological compounds. *See* Linda J. Demaine & Aaron Xavier Fellmeth, *Reinventing the Double Helix: A Novel and Nonobvious Reconceptualization of the Biotechnology Patent*, 55 STAN. L. REV. 303, 304 (2002).

45. Rebecca S. Eisenberg, *Patents, Product Exclusivity, and Information Dissemination: How Law Directs Biopharmaceutical Research and Development*, 72 FORDHAM L. REV. 477, 481 (2003).

46. Scholars have referred to such rights as a second line of patent protection. *See, e.g.*, *id.* at 482-83. In the United States, for example, data associated with new drugs receive five years of exclusive protection, while data associated with a new indication of an existing drug receive three years of exclusive protection. 21 U.S.C. § 355(c)(3)(E)(ii)-(iii) (2000). This trend has been exported through provisions in trade agreements that require strict protection of pharmaceutical test data. *See* Susan Scafidi, *The "Good Old Days" of TRIPS: The U.S. Trade Agenda and the Extension of Pharmaceutical Test Data Protection*, 4 YALE J. HEALTH POL'Y L. & ETHICS 341 (2004).

Trade Organization's (WTO) Trade Related Aspects of Intellectual Property (TRIPS) Agreement is the foundation of this treaty architecture,[47] but regional and bilateral agreements increasingly impose even higher protections upon countries.[48] This shift towards stronger IP protection—driven by the lobbying power of Hollywood, the recording industry, prepackaged software companies, book publishers, and pharmaceutical companies[49]—represents a massive and unprecedented experiment in innovation policy. This is particularly true in the area of medicine: at the time the Uruguay

47. *See* Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C art. 27.1, LEGAL INSTRUMENTS—RESULTS OF THE URUGUAY ROUND vol. 31, 33 I.L.M. 81 (1994) [hereinafter TRIPS Agreement]. Least-developed countries have until January 1, 2006, to comply with TRIPS and have the right to defer patents and data exclusivity rights on pharmaceuticals until 2016. World Trade Org., Doha WTO Ministerial 2001, *Declaration on the TRIPS Agreement and Public Health*, WT/MIN(01)/DEC/2, ¶ 7 (Nov. 20, 2001) [hereinafter *Doha Declaration*] (amending the timeline for implementation); Press Release, World Trade Org., Council Approves LDC Decision with Additional Waiver (June 28, 2002), http://www.wto.org/english/news_e/pres02_e/pr301_e.htm. For a discussion of the TRIPS Agreement and the actors behind it, see PETER DRAHOS WITH JOHN BRAITHWAITE, INFORMATION FEUDALISM: WHO OWNS THE KNOWLEDGE ECONOMY (2002), and SUSAN K. SELL, PRIVATE POWER, PUBLIC LAW: THE GLOBALIZATION OF INTELLECTUAL PROPERTY RIGHTS (2003). For a guide to the provisions of TRIPS, see MICHAEL BLAKENEY, TRADE RELATED ASPECTS OF INTELLECTUAL PROPERTY RIGHTS: A CONCISE GUIDE TO THE TRIPS AGREEMENT (1996). For a consideration of the particular implications of the TRIPS Agreement for developing countries, see CARLOS CORREA, THE TRIPS AGREEMENT: A GUIDE FOR THE SOUTH (1997).

48. *See, e.g.*, CARSTEN FINK & PATRICK REICHENMILLER, TIGHTENING TRIPS: THE INTELLECTUAL PROPERTY PROVISIONS OF RECENT US FREE TRADE AGREEMENTS (World Bank Group, Int'l Trade Dep't, Trade Note 20, 2005). The European Union has also used free trade agreements to impose TRIPS-plus requirements upon countries. *See* PRADEEP S. MEHTA ET AL., "TRIPS-PLUS": ENHANCING RIGHT HOLDERS' PROTECTION, ERODING TRIPS' FLEXIBILITIES (CUTS Centre for Int'l Trade, Econ., and Env't., Briefing Paper No. 2, 2004). Regional agreements also sometimes impose standards higher than those in the TRIPS Agreement, as is the case for the Bangui Agreement among the African Intellectual Property Organization (OAPI) countries of West Africa. *See* [GR. BRIT.] COMM'N ON INTELLECTUAL PROP. RIGHTS, INTEGRATING INTELLECTUAL PROPERTY RIGHTS AND DEVELOPMENT POLICY 8 (2002) [hereinafter COMM'N ON IPR], http://www.ipr commission.org/papers/pdfs/final_report/CIPRfullfinal.pdf; Ellen 't Hoen, *TRIPS, Pharmaceutical Patents, and Access to Essential Medicines: A Long Way from Seattle to Doha*, 3 CHI. J. INT'L L. 27, 45 (2002).

49. *See* DRAHOS WITH BRAITHWAITE, *supra* note 47, at 85-149 (discussing the history of TRIPS); SELL, *supra* note 47, at 96 (noting that through TRIPS, "[i]n effect, twelve corporations made public law for the world").

Round of trade negotiations was launched, more than fifty countries did not provide patent protection on medicines.[50]

Proponents of this IP expansion contend that it will spur innovation and therefore increase aggregate social welfare. This reflects the dominant justification for patents and other forms of IP. In wealthy economies (even where the copyright tradition is premised upon moral rights), these rights are consistently cast in utilitarian terms: the rights are considered first and foremost a tool to encourage private investment in information goods.[51]

Economists, however, are ambivalent about the effect of strong exclusive rights on innovation and welfare.[52] This stems from the fact that information is both nonrival and a critical input for further innovation. Once produced, information—such as a scientific formula—is most efficiently accessible at its marginal cost of zero. If priced at zero, however, firms will not invest in research. Patents are one solution to this; they incentivize innovation by granting firms a temporary monopoly period in which to reap supra-marginal profits. But they also create deadweight loss by raising the marginal cost of consumption above zero. Such exclusive rights also have the potential to stymie innovation because information is a component in its own production. Patents thus raise the costs of innovation, even as they increase its potential value. As a result, even in a dynamic analysis, an overly expansive set of rights leads to too little innovation.[53] Strong patents—particularly in the aggregate—have been shown, both theoretically and empirically, to reduce both innovation and welfare.[54]

---

50. *See* CARLOS CORREA, INTEGRATING PUBLIC HEALTH CONCERNS INTO PATENT LEGISLATION IN DEVELOPING COUNTRIES 11 (2000) (citing UNITED NATIONS CONFERENCE ON TRADE AND DEVELOPMENT, THE TRIPS AGREEMENT AND DEVELOPING COUNTRIES, NEW YORK AND GENEVA (1996)), http://www.southcentre.org/publications/publichealth/publichealth-04.htm.

51. *See, e.g.*, Rovira, *supra* note 8, at 401 n.3; *see also* John H. Barton, *TRIPS and the Global Pharmaceutical Market*, 23 HEALTH AFF. 146, 148-49 (2004) ("[T]he logic of the patent system is to permit an elevated price to allow recovery of research and development (R&D) costs.").

52. For a brief review of some of the relevant views in economics, see FREDERICK M. ABBOTT, THE TRIPS AGREEMENT, ACCESS TO MEDICINES AND THE WTO DOHA MINISTERIAL CONFERENCE 6 (Quaker U.N. Office, Occasional Paper 7, 2001).

53. Kenneth Arrow articulated this basic tradeoff between rights, innovation, and welfare over forty years ago. *See* Kenneth J. Arrow, *Economic Welfare and the Allocation of Resources for Invention*, *in* THE RATE AND DIRECTION OF INVENTIVE ACTIVITY: ECONOMIC AND SOCIAL FACTORS 609, 614-15 (Nat'l Bureau of Econ. Research ed., 1962).

54. *See, e.g.*, JOSH LERNER, PATENT PROTECTION AND INNOVATION OVER 150 YEARS (Nat'l Bureau Econ. Research, Working Paper No. 8977, 2002); *see also* ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS: HOW OUR BROKEN PAT-

Many who accept these premises nonetheless consider the pharmaceutical sector an exception.[55] They argue that the industry is distinguished by its relatively high cost of R&D and relatively low cost of reverse engineering, and they point to survey data that suggest that patents are central to pharmaceutical firms' appropriation strategies.[56] But all that these facts show, respectively, is that some mechanism is necessary to promote innovation in this sector, and that those firms that dominate under the current system are dependent upon the tools that brought them to dominance. Economists have long debated whether direct government funding or prize systems would have better welfare effects than patents.[57] Calls for alternative strategies to incentivize pharmaceutical development have grown more marked recently, supported by claims that the current pharmaceutical market misdirects innovation and marketing resources, leads to inefficiently high prices, and promotes both counterfeiting and price controls.[58]

Importantly, all of these general conclusions are premised (if only implicitly) on the experiences of wealthy countries and on a one-country model of the market for innovation. When we consider the particular con-

---

ENT SYSTEM IS ENDANGERING INNOVATION AND PROGRESS, AND WHAT TO DO ABOUT IT 16-18 (2004); Arrow, *supra* note 53, at 616-17.

55. *See, e.g.*, Joshua D. Sarnoff, *Abolishing the Doctrine of Equivalents and Claiming the Future After* Festo, 19 BERKELEY TECH. L.J. 1157, 1209 (2004) (noting that although "[s]ubstantial evidence points to the increasingly weak incentives that patents provide relative to other mechanisms for protecting innovations and investments, . . . patent protection may be important to particular technology sectors (such as the pharmaceutical and software industries)").

56. *See, e.g.*, Richard C. Levin et al., *Appropriating the Returns from Industrial Research and Development*, 1987 BROOKINGS PAPERS ON ECON. ACTIVITY 783, 796.

57. *See, e.g.*, Nancy Gallini & Suzanne Scotchmer, *Intellectual Property: When Is It the Best Incentive System?*, *in* 2 INNOVATION POLICY AND THE ECONOMY 51 (Adam B. Jaffe et al. eds., 2001); Brian D. Wright, *The Economics of Invention Incentives: Patents, Prizes, and Research Contracts*, 73 AM. ECON. REV. 691 (1983).

58. *See, e.g.*, Aidan Hollis, An Efficient Reward System for Pharmaceutical Innovation 4-9 (Jan. 17, 2005), http://econ.ucalgary.ca/fac-files/ah/drugprizes.pdf. Hollis summarizes the problem with the traditional consensus in support of patent-driven pharmaceutical R&D this way: "Because pharmaceutical markets function poorly, the patent system does not effectively stimulate drug research and development. Instead, it induces large amounts of research into drugs with relatively little incremental therapeutic value, while providing inadequate incentives to innovate in some areas of great therapeutic value." *Id.* at 1. He proposes, instead, a prize-based system that would reward inventors according to the incremental therapeutic benefit offered by their inventions. *Id.*; *see also* Medical Innovation Prize Act of 2005, H.R. 417, 109th Cong. (2005) (proposing, with regard to medical products, to replace the patent system with a prize fund); Michael Kremer, *Pharmaceuticals and the Developing World*, J. ECON. PERSPS., Fall 2002, at 67, 82 (advocating, with regard to "products needed primarily by developing countries," advance purchase commitments to reward research outputs).

text of developing countries, we find that patents will cost them significantly more, and benefit them significantly less. In the Section that follows, we make this case and relate it to the existing global crises around access to medicines and R&D for neglected diseases. We demonstrate that exclusive rights can be an important cause of unaffordable pricing of existing medicines in LMI countries and can also create impediments to R&D for neglected diseases.

B. The Access Gap and Exclusive Rights

According to the World Health Organization (WHO), roughly ten million lives around the world could be saved every year by improving access to essential medicines and vaccines that already exist.[59] Approximately thirty percent of people around the world do not have regular access to essential medicines, and "in the poorest parts of Africa and Asia this figure rises to over 50%."[60] This is what we term the "access gap," and it has many determinants.[61]

---

59. WORLD HEALTH ORG., EQUITABLE ACCESS TO ESSENTIAL MEDICINES: A FRAMEWORK FOR COLLECTIVE ACTION 1 (2004) [hereinafter WHO, FRAMEWORK FOR ACTION]. The WHO defines essential medicines as "those that satisfy the priority health care needs of the population." World Health Org., Essential Drugs and Medicines Policy, *at* http://www.who.int/medicines (last updated Mar. 3, 2005). The WHO's essential drugs list (EDL) is compiled by an Expert Committee, on the basis of a variety of factors including "the disease burden and sound and adequate data on the efficacy, safety and comparative cost-effectiveness of available treatments." World Health Org., Procedure to Update and Disseminate the WHO Model List of Essential Medicines, Document EB109/8 (Annex), Dec. 7, 2001, http://www.who.int/medicines/organization/par/edl/procedures.shtml (last updated July 28, 2004). While the EDL is useful for many purposes, it is important to note that it is not a list of all life-saving medicines, much less all medicines that would provide medical benefit to individuals in developing countries.

60. *See* WHO, FRAMEWORK FOR ACTION, *supra* note 59, at 1.

61. Jonathan Quick of the WHO's Essential Medicines Division identifies four: "(1) irrational use of medicines, (2) unfair financing for healthcare, including medicines, (3) unreliable delivery systems and (4) high medicines prices." Jonathan D. Quick, Editorial, *Essential Medicines Twenty-Five Years On: Closing the Access Gap*, 18 HEALTH POL'Y & PLAN. 1, 1 (2003); *see* HANNAH E. KETTLER & CHRIS COLLINS, USING INNOVATIVE ACTION TO MEET GLOBAL HEALTH NEEDS THROUGH EXISTING INTELLECTUAL PROPERTY REGIMES 40 (Comm'n on Intellectual Prop. Rights, Study Paper 2b, 2004) (identifying "[f]inancial resources, health care infrastructure, and political will" as some of the pivotal factors), http://www.iprcommission.org/papers/pdfs/study_papers/sp2b_kettler_study.pdf.

One important determinant is price.[62] Unsurprisingly, there is "considerable evidence that consumption of medicines is sensitive to price."[63] In particular, price has disproportionately severe effects on patients in LMI countries.[64] Not only are consumers in these countries poorer on average,

62. Many existing drugs are unaffordable for patients around the world. *See, e.g.*, MSF, UNTANGLING THE WEB, *supra* note 6, at 4 (noting that "[t]he high price of HIV/AIDS medicines continue[s] to represent one of the main barriers to their availability in developing countries," citing in particular the high cost of second-line therapies for drug resistant HIV). Price is not just a problem for people living with HIV/AIDS. The high cost of interferon/ribavrin combination therapy for Hepatitis C is "unquestionably beyond the reach of developing countries." MÉDECINS SANS FRONTIÈRES, DOHA DERAILED: A PROGRESS REPORT ON TRIPS AND ACCESS TO MEDICINES 6 (2003), http://www.accessmed-msf.org/documents/cancunbriefing.pdf. Access to other drugs, from certain classes of antibiotics to anti-cancer drugs, has also been limited by price. *See, e.g.*, *id.*; Nadia Aït-Khaled et al., *Chronic Respiratory Diseases in Developing Countries: The Burden and Strategies for Prevention and Management*, 79 BULL. WORLD HEALTH ORG. 971 (2001) (describing need for low-cost generic alternatives to treat asthma in developing countries); Mogha Kamal Smith, *Why Developing Countries Need Access to Cheap Treatments for Diabetes*, DIABETES VOICE, July 2003, at 31, 32 (noting that only three percent of people with diabetes in developing countries get treatment "partly because the majority of these people have to pay for their drugs out of their own pockets"); *Thousands Denied Anti-Cancer Drugs*, BBC NEWS, Feb 14, 2003 (citing price as a major barrier to access to cancer drugs in developing countries), *at* http://news.bbc.co.uk/2/hi/health/2761277.stm. High prices also constitute a barrier to the drugs that do exist for neglected diseases. *See* Rachel Cohen, *An Epidemic of Neglect*, MULTINATIONAL MONITOR, June 2002, http://multinationalmonitor.org/mm2002/02june/june02corp1.html; Médecins Sans Frontières, The Campaign: Target Diseases, Leishmaniasis, *at* http://www.accessmed-msf.org/campaign/lsh01.shtm (last visited Mar. 9, 2005) (describing lack of access to treatment in countries where there is no generic available); Médecins Sans Frontières, The Campaign: Target Diseases, Sleeping Sickness, *at* http://www.accessmed-msf.org/campaign/slp01.shtm (last visited Mar. 9, 2004) (describing severe lack of access to diagnostics and treatment for African sleeping sickness).

63. COMM'N ON IPR, *supra* note 48, at 37 (citing several studies about the specific and very positive effects that price cuts on ARVs would have upon consumption in countries like Uganda). This is especially the case for the poor. *See* ADAM WAGSTAFF & MARIAM CLAESON, WORLD BANK, THE MILLENNIUM DEVELOPMENT GOALS FOR HEALTH: RISING TO THE CHALLENGES 9 (2004) ("Higher money prices tend to reduce demand—especially among the poor—unless accompanied by improvements in service quality."); WAGSTAFF & CLAESON, *supra*, at 75 ("Affordability—the price paid relative to discretionary income—is undoubtedly one important barrier preventing the use of health services.").

64. *See* Quick, *supra* note 61, at 2-3 ("[G]overnments, other health care providers, and households in developing countries are each highly sensitive to medicines prices."). In LMI countries, high drug prices have been shown to have devastating results for the poor. For example, in Vietnam in 1993, just one visit by an individual in a household in the poorest fifth of the population to a local health center "resulted in a bill for drugs equal to 11 percent of the household's annual nonfood consumption." WAGSTAFF & CLAESON, *supra* note 63, at 119 box 7.9. As many as three million Vietnamese have been

but they also tend to pay a greater proportion of their own medical costs than consumers in wealthy countries. While patients in wealthy countries are often insulated from the high cost of medicines by third party payers (for example, insurance companies or government funded programs), in LMI countries, "public medicine expenditure does not cover the basic medicine needs of the majority of the population"[65] and private health insurance is rare.[66] In both low- and middle-income countries, the public sector pays less than thirty percent of drug costs.[67]

Price, in turn, is affected by patent status. Empirical studies focused on developing countries predict, for example, that "the introduction of patent regimes . . . has, or is predicted to have, the effect of raising prices. The estimates range widely depending on the drugs and countries being considered—from 12% to over 200%, but even the lower estimates imply very substantial costs for consumers."[68] Development and aid agencies working in the field confirm these theoretical predictions. MSF has concluded that "[t]he most significant factor in lowering prices [is] the introduction of generic sources in a country,"[69] and Oxfam International has

---

"pushed into poverty as a result of high out-of-pocket payments for healthcare, much of it attributable to high drug costs." WAGSTAFF & CLAESON, *supra* note 63, at 119 box 7.9. Households also "appear to have been deterred from using health services because of high drug costs." WAGSTAFF & CLAESON, *supra* note 63, at 119 box 7.9.

65. WHO, FRAMEWORK FOR ACTION, *supra* note 59, at 1; *see* Jonathan D. Quick, *Ensuring Access to Essential Medicines in the Developing Countries: A Framework for Action*, 73 CLINICAL PHARMACOLOGY & THERAPEUTICS 279, 282 (2003) ("Private out-of-pocket spending on medicines is the largest household health expenditure in many [developing] countries . . . ."). By comparison, "in many high income countries, over 70% of pharmaceuticals are publicly funded." WHO, FRAMEWORK FOR ACTION, *supra* note 59, at 1.

66. WHO, FRAMEWORK FOR ACTION, *supra* note 59, at 5 (noting that median insurance coverage "is 35% in Latin America, 10% in Asia, and less than 8% in Africa" and that "the inclusion of medicine reimbursement in health insurance varies greatly").

67. *See* WORLD HEALTH ORG., THE WORLD MEDICINES SITUATION 46 tbl.5.3 (2004) [hereinafter WHO, WORLD MEDICINES SITUATION].

68. COMM'N ON IPR, *supra* note 48, at 37. Developing countries newly introducing patents also are disadvantaged by the fact that the resulting profits are likely to accrue mostly to companies outside the country. *See* JEAN O. LANJOUW, THE INTRODUCTION OF PHARMACEUTICAL PRODUCT PATENTS IN INDIA: "HEARTLESS EXPLOITATION OF THE POOR AND SUFFERING?" 5-6 (Nat'l Bureau Econ. Research, Working Paper No. 6366, 1998).

69. MÉDECINS SANS FRONTIÈRES ET AL., SURMOUNTING CHALLENGES: PROCUREMENT OF ANTIRETROVIRAL MEDICINES IN LOW- AND MIDDLE-INCOME COUNTRIES 46 (2003), http://www.accessmed-msf.org/documents/procurementreport.pdf.

called generic competition the single most important tool to remedy the access gap.[70]

Some have argued that pharmaceutical companies are unlikely to patent in LMI countries, and therefore that we ought not focus on patents as a barrier to access.[71] This position has been widely discredited based on evidence of patenting, particularly in key supplier markets.[72] Pharmaceutical companies have been willing to patent widely, and cling to the exclusivity that their patents provide, even where the public health implications are dire.[73] And although, as we might expect, gross national income, market size, and relative income inequality are generally important determinants of patenting strategy,[74] patenting still occurs in low-income countries.[75]

Furthermore, the absence of patents in a given country is not the *sine qua non* of effective access to generics. A supply of medicines must also exist, but "[d]eveloping countries differ substantially in terms of their existing pharmaceutical production capacity."[76] In the poorest countries,

---

70. *See* MOHGA K. SMITH, GENERIC COMPETITION, PRICE, AND ACCESS TO MEDICINES: THE CASE OF ANTIRETROVIRALS IN UGANDA 2 (Oxfam Briefing Paper No. 26, 2002).

71. *See* Amir Attaran, *How Do Patents and Economic Policies Affect Access to Essential Medicines in Developing Countries?*, 23 HEALTH AFF. 155 (2004); *see also* Amir Attaran & Lee Gillespie-White, *Do Patents for Antiretroviral Drugs Constrain Access to AIDS Treatment in Africa?*, 286 JAMA 1886, 1888 tbl.1 (2001).

72. *See, e.g.*, COMM'N ON IPR, *supra* note 48, at 20-26, 29-51; Kevin Outterson, *Pharmaceutical Arbitrage: Balancing Access and Innovation in International Prescription Drug Markets*, 5 YALE J. HEALTH POL'Y L. & ETHICS 193, 255-58 (2005).

73. Many of the most important ARVs, for example, are widely patented in Africa. *See* Marleen Boelaert et al., Letter to the Editor, *Do Patents Prevent Access to Drugs for HIV in Developing Countries?*, 287 JAMA 840 (2002); Consumer Project on Technology et al., Comment on the Attaran/Gillespie-White and PhRMA Surveys of Patents on Antiretroviral Drugs in Africa (Oct. 16, 2001), *at* http://www.cptech.org/ip/health/africa/dopatentsmatterinafrica.html; *see also infra* note 160 (noting GlaxoSmithKline's attempt to prevent generic companies from selling cheaper versions of their ARV products in Ghana and Uganda).

74. *See* Attaran, *supra* note 71, at 158.

75. *See id.* at Supplemental Exhibit, *available at* http://content.healthaffairs.org/cgi/content/full/23/3/155/DC1. One example is Malawi, which has a per capita gross national income of less than $200 per year. *Id.*

76. Frederick M. Abbott, *The WTO Medicines Decision: The Political Economy of World Pharmaceutical Trade and the Protection of Public Health*, 99 AM. J. INT'L L. (forthcoming 2005) (manuscript of Mar. 31, 2005 at 28, on file with authors). Few developing countries can produce the essential active pharmaceutical ingredients (APIs), although if they are able to acquire APIs cheaply, many can formulate finished products. *Id.* (manuscript at 28 n.147) (noting that the APIs that make up ARV medicines are complex, and made only by a few companies in the world); *see also* WHO, WORLD MEDI-

even when medicines are locally formulated, they may be unaffordable because of inefficiencies in production and limited market size.[77] As a result, patents in a variety of countries can matter a great deal to the shape of the supply curve. Patents may obstruct production and export from certain countries, namely middle-income supplier countries—such as India—which play a critical role in the global market for generics. They may also limit the available markets to those that are too small to justify the costs of reverse engineering specific medicines, retooling production facilities to make them, or establishing distribution networks. For example, while many poor and low-prevalence countries in Africa have few or no patents on ARV medicines, it was not until late 2003 that the first African company began to locally produce ARVs.[78] As one would expect, that company is based in South Africa (where, not incidentally, patents first had to be overcome).[79]

While patents are not the only factor blocking access to medicines, exclusive rights in one LMI country can create serious, preliminary obstacles to access in that country and prevent the emergence of a competitive market to supply medicines to another country that has no such barriers. Finally, the aspects of this problem that are visible today are only the tip of the iceberg. It is easy, but shortsighted, to ignore the value of medicines that have not yet been developed.[80] Obviously we expect—and need—new medicines. As they come into being, as TRIPS takes hold in supplier coun-

---

CINES SITUATION, *supra* note 67, at 6 (reporting that only thirteen countries in the world make both formulations and APIs).

77. *See, e.g.*, ROBERT LEWIS-LETTINGTON & CHIKOSA BANDA, A SURVEY OF POLICY AND PRACTICE ON THE USE OF ACCESS TO MEDICINES-RELATED TRIPS FLEXIBILITIES IN MALAWI 14 (2004) (noting that although Malawi has some capacity to make finished products, it imports APIs from India or China, and that "generic pharmaceutical products manufactured in Malawi are generally more expensive than those imported from elsewhere, for example, from India" because, inter alia, of high transportation costs, high communications costs, and limited markets), http://www.dfidhealthrc.org/Shared/publications/Issues_papers/ATM/Lettington.pdf; *see also* ROBERT LEWIS-LETTINGTON & PETER MUNYI, WILLINGNESS AND ABILITY TO USE TRIPS FLEXIBILITIES: KENYA CASE STUDY 12-13 (2004) (reporting the same dynamics in Kenya), *available at* http://www.dfid.gov.uk/pubs/files/dfidkenyareport.pdf.

78. *See S. Africa's Aspen To Launch First Local AIDS Drug*, REUTERS NEWSMEDIA, Aug. 5, 2003, http://www.aegis.com/news/re/2003/RE030806.html; *see also* Otterson, *supra* note 72, at 257.

79. *See S. Africa's Aspen To Launch First Local AIDS Drug*, *supra* note 78.

80. Some argue that patents are not a significant concern in developing countries because the WHO's EDL is mostly comprised of drugs that are off patent. *See* Attaran, *supra* note 71, 159-60. This argument ignores medicines not yet invented and the fact that medicines only appear on the EDL after an assessment that includes their "cost-effectiveness." *See supra* note 59.

tries such as India, and as TRIPS-plus provisions take effect in more and more countries, the role of exclusive rights in the access crisis will grow more important. Though sobering, this is only half of the problem.

C. The R&D Gap and Exclusive Rights

1. The 10/90 Gap

Significant morbidity and mortality in developing countries result from diseases for which there are currently no effective, easy-to-use medicines.[81] Unfortunately, our patent-based R&D system does not adequately address this problem. A mere ten percent of the world's expenditure on R&D is devoted to conditions that cause ninety percent of the global disease burden—a situation that has been termed the "10/90 gap."[82] Only one percent of medications introduced between 1975 and 1999—thirteen out of an estimated 1393—targeted tuberculosis and tropical diseases (including malaria and infectious diarrhoeal diseases) which cause 11.4% of the global disease burden, including a substantial proportion of the disease burden in developing countries.[83]

Beyond the gap in development, the current system fails to optimize existing medicines and medical technologies for use in developing countries. For instance, heat stable formulations—essential in countries with warm climates and little refrigeration—do not exist for several essential medicines, such as insulin and oxytocin.[84] Many desirable fixed-dose

---

81. *See generally* MÉDECINS SANS FRONTIÈRES & DRUGS FOR NEGLECTED DISEASES WORKING GROUP, FATAL IMBALANCE—THE CRISIS IN RESEARCH AND DEVELOPMENT FOR DRUGS FOR NEGLECTED DISEASES (2002); Carlos M. Morel, *Neglected Diseases: Under-funded Research and Inadequate Health Interventions*, 4 EMBO REP. S35 (2004); Ellen F.M. 't Hoen, *The Responsibility of Research Universities To Promote Access to Essential Medicines*, 3 YALE J. HEALTH POL'Y L. & ETHICS 293 (2003); Patrice Trouiller et al., *Drug Development for Neglected Diseases: A Deficient Market and a Public-Health Policy Failure*, 359 LANCET 2188 (2002).

82. *See* GLOBAL FORUM FOR HEALTH RESEARCH, *supra* note 31.

83. Trouiller et al., *supra* note 81, at 2189-90; *see* Press Release, Médecins Sans Frontières, Drugs for Neglected Diseases Initiative: Teaming Up To Address Neglect (Mar. 12, 2003), http://www.accessmed-msf.org/prod/publications.asp?scntid=12320031354463&contenttype=PARA&. Public-sector-based research, particularly R&D sponsored by the military, has been an important source of drugs for diseases that have primary incidence in LMI countries. *See, e.g.*, Donald G. McNeil Jr., *Herbal Drug Widely Embraced in Treating Resistant Malaria*, N.Y. TIMES, May 10, 2004, at A1 (discussing artemisinin, a treatment for malaria first isolated by Chinese military researchers, as well as mefloquine, an antimalarial drug developed at the Walter Reed Army Institute of Research in the 1960s).

84. WARREN KAPLAN & RICHARD LAING, WORLD HEALTH ORG., PRIORITY MEDICINES FOR EUROPE AND THE WORLD 62 (2004), http://mednet3.who.int/priority

combinations, which combine several medicines into a single pill and make prescribing and adhering to complex medical regimens much simpler, do not exist.[85] Diagnostic and monitoring tools developed for high-income markets are often inappropriate for use in developing countries and may cost more to use than the medicines involved in treating the underlying illness.[86] We also lack formulations for small patient populations with special needs, such as children, particularly where most patients live in developing countries.[87]

As Juan Rovira, a former Senior Health Economist at the World Bank, has observed, "the patent system leads R&D toward profitable diseases and conditions, rather than toward diseases that cause the most morbidity and mortality."[88] Thus, just as the static costs imposed by patents cannot be understood in a hypothetical one-country model, neither can the potential dynamic effects of patents. The dynamic benefits of patents for poor countries are likely to be much smaller than the one-country model predicts, because their markets are small compared to those markets that already offer patent protection.[89] Under the circumstances, it is not surprising that pharmaceutical companies do not direct their research towards these markets.[90]

---

meds/report/index.htm. Oxytocin is used to treat post-partum hemorrhage in women, which is a major cause of disability and death in developing countries. *Id.* at 47.

85. These would be especially useful for second-line ARVs and multidrug resistant tuberculosis. *Id.* at 124.

86. *See, e.g.*, Renuka Rayasam, *Austin-Based Company Will Build Device To Improve Treatment in Developing Countries*, AUSTIN AM. STATESMAN, July 9, 2004.

87. Few drug companies have tailored treatments to suit children with AIDS, in part because there are declining numbers of children born with HIV/AIDS in wealthy countries. *See* Editorial, *Children and AIDS*, N.Y. TIMES, Feb. 22, 2005, at A16. The market is apparently too small to attract even the modest investment needed to create low-dose, breakable, or chewable tablets. *See* Médecins Sans Frontières, Children and AIDS: Neglected Patients (July 15, 2004), http://www.msf.org/content/page.cfm?articleid=C35A2DA2-D4E3-425A-879860086416E313.

88. Rovira, *supra* note 8, at 405; *see also* Jean O. Lanjouw, *Intellectual Property, and the Availability of Pharmaceuticals in Poor Countries*, *in* 3 INNOVATION POLICY AND THE ECONOMY 91, 100 (Adam B. Jaffe et al. eds., 2003).

89. *See, e.g.*, LANJOUW, *supra* note 68, at 7-8 (presenting this argument but also offering reasons that it may "paint[] too gloomy a picture").

90. This cannot reasonably be attributed to a lack of patent protection or enforcement. *See generally* Lanjouw, *supra* note 88. Moreover, when research is oriented toward conditions affecting LMI populations, it tends to target those affecting the upper classes. *See* Emmanuel Combe et al., *Pharmaceutical Patents, Developing Countries, and HIV/AIDS Research*, *in* ECONOMICS OF AIDS AND ACCESS TO HIV/AIDS CARE IN DEVELOPING COUNTRIES 151, 160 (2003).

2. Patents As Barriers to R&D, Particularly for Low-Commercial-Value Research

The R&D gap is perhaps the most obvious manifestation of the dynamic failures of patents for people living in LMI countries. Simply put, patents do not help the poorest of the poor because a monopoly in such a market is worth very little. But as noted above, patents can also create a drag on the innovation process itself; this can be particularly problematic where the research in question has low commercial value.

As the number of patents and patent holders associated with a given biomedical innovation increases,[91] so do the transaction costs associated with conducting research. These costs are at the center of recent concerns about the growth of an "anticommons"[92] or "patent thickets."[93] The need to negotiate permission to use or litigate disagreements about research tools slows research and increases its cost.[94] While transaction costs only rarely completely prohibit commercially valuable research,[95] they may hinder research at universities or nonprofit institutions concerned with developing world diseases where commercial pay-offs are at best uncertain.[96] Indeed, several of the concrete examples we have of patent thickets

---

91. *See* John P. Walsh et al., *Research Tool Patenting and Licensing and Biomedical Innovation*, *in* PATENTS IN THE KNOWLEDGE-BASED ECONOMY 285, 331 (Wesley M. Cohen & Stephen A. Merrill eds., 2003) (confirming that such patenting is increasing).

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

93. Carl Shapiro, *Navigating the Patent Thicket: Cross-Licenses, Patent Pools, and Standard Setting*, *in* 1 INNOVATION POLICY AND THE ECONOMY 119 (Adam B. Jaffe et al. eds., 2000). Shapiro describes a patent thicket as the "overlapping set of patent rights requiring that those seeking to commercialize new technology obtain licenses from multiple patentees." *Id.* at 119.

94. *See* Walsh et al., *supra* note 91, at 314 (noting that more than one-third of respondents in the authors' survey of scientists, IP attorneys, and business managers reported that patents on research tools caused delays and added to the costs of research); *see also* John P. Walsh et al., *Working Through the Patent Problem*, 299 SCIENCE 1021, 1021 (2003) (noting that assertions of IP rights may hinder science and that policy makers should take steps to ensure continued protection of science intended for the public domain). *But see* David E. Adelman, *A Fallacy of the Commons in Biotech Patent Policy*, 20 BERKELEY TECH. L.J. 985 (2005) (arguing that the potential adverse effects of biotech patenting are less significant than many have predicted).

95. *See* Walsh et al., *supra* note 91, at 286.

96. Arti K. Rai, *Proprietary Rights and Collective Action: The Case of Biotechnology Research with Low Commercial Value*, *in* INTERNATIONAL PUBLIC GOODS AND TRANSFER OF TECHNOLOGY UNDER A GLOBALIZED INTELLECTUAL PROPERTY REGIME 288, 289 (Keith E. Maskus & J.H. Reichman eds., forthcoming 2005); *see also* Walsh et al., *supra* note 91, at 304 (noting that transaction costs were only relevant when projects had questionable commercial viability).

that have caused lengthy delays, or of broad and exclusively licensed research tool patents that have obstructed research initiatives, relate to products intended for developing countries.[97]

There is also evidence that patents cause scientists to redirect their research efforts towards "areas with more intellectual property freedom."[98] Such redirected research may be less efficient or successful, particularly if the areas most crowded with patents are also those that scientists deem most promising. Patenting practices may also dampen scientific exchange. Recent data suggest that university-based geneticists who engage in commercial research are more likely than their peers to withhold data from fellow academic scientists.[99]

More broadly, patents give their owners the right to block research outright.[100] A few patents on an important gene target, for example, have

97. *See* Rai, *supra* note 96, at 295-96 & nn.38-44 (discussing these problems in the context of a malaria vaccine and transgenic agricultural products relevant to developing countries). We might hope that companies would be more amenable to granting research licenses to low-commercial-value projects since these do not threaten the product markets that the company cares about. *Cf. id*. at 299-300 (suggesting that collective rights management has a better chance of success with low-commercial-value research). However, direct competition is only one of the concerns that such research may pose to the profits of a patent-holding firm. Another concern is the potential for follow-on research that might raise safety questions about a therapeutic compound. *See infra* note 103 and accompanying text.

98. *See* Walsh et al., *supra* note 91, at 286. Patents on compounds seem to trigger this response more often than do patents on research tools, but this is little comfort for those concerned with R&D for new medicines. *Id.* at 303 (reporting that "[o]f the 11 industry respondents who did mention IP as a cause for redirecting their research, seven . . . were primarily concerned with IP on compounds, not on research tools").

99. *See* Eric G. Campbell et al., *Data Withholding in Academic Genetics: Evidence from a National Survey*, 287 JAMA 473, 479 (2002) (concluding that "[t]he commercial applications of genetics research, along with increasing dependence on industry funding and the rise of commercial norms in the academy may be partially responsible" for this withholding). Campbell et al.'s survey showed that over a three-year period, about half of geneticists polled had been unable to obtain information or materials from another university-based geneticist, and twenty-one percent had therefore abandoned a promising line of research. *Id.* at 478. In about twenty percent of the cases, one important reason cited for refusing to grant access to others was the need to abide by an agreement with an industrial sponsor or preserve confidentiality for patenting purposes. Although the most common reason given for such refusals was the "effort required," this category "probably also includes costs associated with difficulties in concluding complex negotiations over [Material Transfer Agreements]." Rai, *supra* note 96, at 294 (discussing Campbell et al.'s results).

100. *See* 35 U.S.C. § 271 (2000).

the potential to slow research for a generation.[101] A firm may want to block other researchers for a variety of reasons, including preventing competitors from gaining an advantage and retaining all of the potential value of improvements for itself.[102] Pharmaceutical firms are particularly reluctant to allow research on therapeutic compounds, citing two concerns: (1) the possibility of being excluded from future developments of their products, and (2) the possibility that the researcher will "generat[e] and disclos[e] data that could create problems for the firm in seeking FDA approval."[103]

Unlike companies, universities may be willing to license the research tools they develop freely to other public sector institutions. In practice, though, they sometimes grant exclusive licenses to companies that then refuse to sublicense any rights or that impose onerous terms on sublicensees.[104] Recent research suggests that public institutions may issue such exclusive licenses with alarming frequency, even where the tools are useful primarily for diseases prevalent in developing countries. For example, a recent map of patents relevant to the development of a malaria vaccine found that only eight of the twenty-seven "moderate to high priority" pat-

101. Myriad Genetics has used its patents on genes that appear to trigger breast cancer to force medical schools to abandon research programs. *See* JAFFE & LERNER, *supra* note 54, at 16-17. Walsh and colleagues report "widespread complaints" about patent holders asserting exclusive rights over potential drug targets. Walsh et al., *supra* note 91, at 310, 312-14 (discussing several important targets that firms have sought to exclude others from using, including targets related to HIV, cancer, and hepatitis C).

102. *See, e.g.*, Rebecca Eisenberg, *Patents and the Progress of Science: Exclusive Rights and Experimental Use*, 56 U. CHI. L. REV. 1017, 1079-84 (1989) (discussing a case where a patent holder sued a competitor to prevent it from making a preferable, synthetic version of the blood clotting compound Factor VIII); *see also* David P. Hamilton, *Silent Treatment How Genentech, Novartis Stifled a Promising Drug*, WALL ST. J., Apr. 5, 2005, at A1.

103. NAT'L INSTS. OF HEALTH, REPORT OF THE NATIONAL INSTITUTES OF HEALTH (NIH) WORKING GROUP ON RESEARCH TOOLS (1998), http://www.nih.gov/news/researchtools. NIH reports that firms may seek to either block such research outright, or permit it only if accompanied by a grant-back of a nonexclusive, royalty-free license to any improvements or new uses. *Id.*

104. Harvard's exclusive license of the transgenic OncoMouse to DuPont is a well-known example. *See* Sasha Blaug et al., *Managing Innovation: University-Industry Partnerships and the Licensing of the Harvard Mouse*, 22 NATURE BIOTECHNOLOGY 761, 762 (2004); Walsh et al., *supra* note 91, at 307-08; Victoria Slind-Flor, *Can These Mice Be Saved?; Fenwick Lawyers Say That DuPont's Licensing Terms Are Preventing Researchers from Using the Harvard Mouse*, IP L. & BUS., Sept. 30, 2004, at 11.

ent families that were originally filed by public entities remain available for licensing from that entity.[105]

In response to these types of concerns, researchers have developed strategies to avoid these barriers. Academic scientists report regularly ignoring patents, and companies have rarely sought to prosecute them for infringement.[106] A recent ruling from the Federal Circuit has, however, made it clear that the experimental use exemption that many academics invoke does not protect them.[107] Infringement actions against universities, though rare, are not unprecedented,[108] and a few high-profile actions could quickly shift the tentative balance.[109] Moreover, there is still cause for concern if individuals are altering their research agendas or expending significant time and money trying to negotiate rights before deciding to infringe.

As we discuss in Part III, some universities have adopted the new strategy of negotiating formal research exemptions for themselves and other academic institutions. However, these exemptions may not extend to commercially-sponsored or -oriented research, limiting their efficacy. Outsourcing of research to jurisdictions where there are fewer patents[110] or

105. *See* Malaria Vaccine Initiative at PATH, Malaria Antigen Patent Access Project Background Information 2 (Mar. 2005) (unpublished manuscript, on file with authors).

106. *See* Walsh et al., *supra* note 91, at 324-26.

107. *Id.* at 235. The Federal Circuit in *Madey v. Duke University* reiterated that the common law research exemption applies only to research conducted "for amusement, to satisfy idle curiosity, or for strictly philosophical inquiry," and further held that the exemption "does not immunize use that is in any way commercial in nature," even if that research occurs at a nonprofit institution. 307 F.3d 1351, 1362 (Fed. Cir. 2002), *cert. denied*, 539 U.S. 958 (2003). Congress has created a statutory exemption for research "reasonably related to the development and submission of information" under federal drug regulations. 35 U.S.C. § 271(e)(1) (2000). This exemption has been used to aid companies preparing, just prior to patent expiration, to launch generic products. The limits of this exception are currently under review at the Supreme Court. Merck KGaA v. Integra Lifesciences I, Ltd., 331 F.3d 860 (2003), *cert. granted*, 125 S. Ct. 823 (2005).

108. *See* Jon F. Merz et al., *Diagnostic Testing Fails the Test*, 415 NATURE 577 (2002) (discussing a series of actions brought by companies to stop academic labs from using patented diagnostic tests).

109. Rebecca S. Eisenberg, *Patent Swords and Shields*, 299 SCIENCE 1018 (2003) ("With their large endowments and habits of documenting their activities in scientific publications, universities would make easy targets."); *see also* Rai, *supra* note 96, at 295.

110. Little is known about how widespread research tool patenting has become outside the United States and other wealthy countries. In the agricultural context, some have argued that concerns about IPRs impeding "research oriented toward food crops for the developing world" are overblown because there are few patent barriers in developing countries. *See, e.g.*, Eran Binenbaum et al., *South-North Trade, Intellectual Property Jurisdictions, and Freedom To Operate in Agricultural Research on Staple Crops*, 51 ECO. DEV. & CULTURAL CHANGE 309, 310, 317 (2003). Others have contended that patents on

where robust research exemptions exist is another possible strategy,[111] but scientific research facilities and expertise will not always be mobile and transferring facilities abroad may entail significant costs.

Finally, neither formal nor informal research exemptions, nor outsourcing, will overcome the problem of blocking patents.[112] The right to research without the ability to commercialize an end product is of little value if we are concerned with improving worldwide health. Indeed, the anticipation of this problem may well be a more important research barrier than the costs and uncertainty associated with anticommons effects for public sector scientists.

D. Recent Proposals and Initiatives To Address the Access and R&D Gaps

The global access and R&D gaps have attracted substantial attention from scholars,[113] NGOs,[114] international bodies,[115] and various national

---

research tools in developing countries may, in fact, have posed barriers to the development and commercialization of GoldenRice™. *See Golden Rice and Trojan Trade Reps: A Case Study in the Public's Mismanagement of Intellectual Property*, RAFI COMMUNIQUE, Sept./Oct. 2000, at 1 (finding a significant number of patents in developing countries but concluding that these patents should not have been considered "insurmountable obstacles"), http://www.etcgroup.org/documents/com_goldenrice.pdf.

111. *See* Walsh et al., *supra* note 91, at 328.

112. Blocking patents arise when a subsequent inventor patents something novel but still within the scope of the original patent. As a result, each party can block the other from making, using, or distributing the follow-on invention. *See* DONALD S. CHISUM, 1 CHISUM ON PATENTS Glossary (2004). Bargaining breakdowns may be likely in such situations. *See* Robert Merges, *Intellectual Property Rights and Bargaining Breakdown: The Case of Blocking Patents*, 62. TENN. L. REV. 75, 75 (1994). Many countries provide for compulsory licensing of blocking patents, with no demonstrably negative effects on investment in research. Merges, *supra*, at 103-05.

113. *See, e.g.*, MICHAEL KREMER & RACHEL GLENNERSTER, STRONG MEDICINE: CREATING INCENTIVES FOR PHARMACEUTICAL RESEARCH ON NEGLECTED DISEASES (2004); Carlos M. Correa, *Public Health and Patent Legislation in Developing Countries*, 3 TUL. J. TECH. & INTELL. PROP. 1 (2001); Patricia M. Danzon & Adrian Towse, *Differential Prices for Pharmaceuticals: Reconciling Access, R&D and Patents*, 3 INT'L J. HEALTH CARE FIN. & ECON. 183 (2003); Lanjouw, *supra* note 88; Susan K. Sell, *TRIPS and the Access to Medicines Campaign*, 20 WIS. INT'L L.J. 481 (2002).

114. *See, e.g.*, Editorial, *The Plagues of Poverty*, N.Y. TIMES, Mar. 19, 2002, at A22 (mentioning the work of the Gates Foundation and Médecins Sans Frontières); Drugs for Neglected Diseases Initiative, *at* http://www.dndi.org (last visited Mar. 30, 2005); HealthGAP, Health Global Access Project (GAP), *at* http://www.healthgap.org (last visited Mar. 30, 2005).

115. Beginning in 2001, the WTO's attention turned to the issue of access leading to the adoption of the Doha Declaration. *Doha Declaration*, *supra* note 47; *see also* World Trade Org., TRIPS and Public Health, *at* http://www.wto.org/english/tratop_e/trips_e/pharmpatent_e.htm (last visited Mar. 11, 2005). In 2000, the WHO and UNAIDS devel-

governments.[116] The increasing attention has generated a number of proposals and initiatives to ameliorate these problems. We focus on those proposals that address the static costs that patents can cause in developing countries, and that seek to stimulate R&D for neglected diseases, and/or target the problems of thickets and other barriers that patents pose to research. These solutions fall into two categories: top-down solutions, which require increased government funding and/or interventions in domestic or international legal regimes, and private sector action that relies on the voluntary initiative of firms.

1. Top-Down Change To Address the Access Gap

Proposals to reduce the difference between patent-based pricing and marginal cost pricing such as compulsory licensing schemes,[117] price controls,[118] changes to the TRIPS Agreement,[119] and alterations in national

---

oped the Accelerating Access Initiative, *see* UNAIDS & WORLD HEALTH ORG., ACCELERATING ACCESS INITIATIVE 1 (2002), http://www.who.int/hiv/pub/prev_care/en/isbn9241210125.pdf, and in late 2003 launched the 3 x 5 Initiative, *see* World Health Org., Fact Sheet 274: The 3 x 5 Initiative (Dec. 2003), *at* http://www.who.int/mediacentre/factsheets/2003/fs274/en.

116. Over one hundred countries have developed national drug policies. Quick, *supra* note 61, at 1.

117. The Consumer Project on Technology has advocated compulsory licensing and recently created a new nonprofit, Essential Inventions, that plans to request compulsory licenses for AIDS drugs in LMI countries. *See* Essential Inventions, *at* http://www.essentialinventions.org (last visited Feb. 24, 2005).

118. *See, e.g.*, Sanjay Kumar, *India To Extend Price Controls on Drugs*, 329 BMJ 368 (2004); Andrew Quinn, *S. Africa Rules Aim To Cut Drug Prices up to 70 Pct*, REUTERS NEWSMEDIA, Jan. 15, 2004 (discussing price controls in South Africa), http://www.aegis.com/news/re/2004/RE040113.html. Numerous countries have implemented mechanisms to control or influence pharmaceutical prices. *See, e.g.*, Austl. Gov't Dep't of Health & Ageing, About the PBS, *at* http://www.health.gov.au/pbs/general/aboutus.htm (last modified Dec. 24, 2003) (describing the Australian Pharmaceutical Benefits Scheme); Can., Patented Med. Prices Review Bd., http://www.pmprb-cepmb.gc.ca (last visited May 6, 2005). However, for a variety of reasons, price controls are an "unsatisfactory policy instrument," particularly for developing countries. Robert Weissman, *A Long, Strange TRIPS: The Pharmaceutical Industry Drive To Harmonize Global Intellectual Property Rules, and the Remaining WTO Legal Alternatives Available to Third World Countries*, 17 U. PA. J. INT'L ECON. L. 1069, 1115 (1996) (noting that price controls are difficult for governments to administer and may produce sub-optimal reductions in price due to uncertain data); *see* Outterson, *supra* note 72, at 239-40 (explaining that price controls fail to take a number of important considerations into account); Jean O. Lanjouw, Patents, Price Controls, and Access to New Drugs: How Policy Affects Global Market Entry 2 (Apr. 19, 2005) (unpublished manuscript prepared for the WHO Comm'n on Intellectual Prop. Rights, on file with authors) (finding that price controls may delay market entry of new drugs in poor countries).

patent laws in either rich[120] or poor countries[121] rely on concerted governmental action. Among these proposals, attempts to encourage developing countries to utilize the flexibilities available to them under international agreements has received the most sustained attention. The TRIPS Agreement, for example, affords member countries complete freedom to determine the grounds for compulsory licenses,[122] although it imposes restrictions on the process for granting them.[123] The recent Doha Declaration on TRIPS and Public Health has also given least-developed countries (LDCs) the right to refuse to offer product patents on pharmaceuticals until 2016.[124]

Unfortunately, such strategies are under attack. The United States is currently using free trade agreements to impose TRIPS-plus standards on dozens of countries around the world.[125] Several of these agreements limit compulsory licensing to situations of emergency, public noncommercial use or to remedies for antitrust violations.[126] Almost all of the agreements

---

119. *See, e.g.*, T.N. Srinivasan, *The TRIPS Agreement*, *in* THE POLITICAL ECONOMY OF INTERNATIONAL TRADE LAW 343 (Daniel L.M. Kennedy & James D. Southwick eds., 2002).

120. *See, e.g.*, Jean O. Lanjouw, *A New Global Patent Regime for Diseases: U.S. and International Legal Issues*, 16 HARV. J.L. & TECH. 85 (2002) (discussing Lanjouw's Foreign Filing License proposal).

121. *See, e.g.*, CARLOS M. CORREA, INTELLECTUAL PROPERTY RIGHTS, THE WTO AND DEVELOPING COUNTRIES: THE TRIPS AGREEMENT AND POLICY OPTIONS (2000) (describing ways in which developing countries can implement TRIPS while still maintaining maximal flexibility and public health benefits).

122. *See* TRIPS Agreement, *supra* note 47, art. 31; *Doha Declaration*, *supra* note 47, ¶ 5(b) ("Each Member has the right to grant compulsory licenses and the freedom to determine the grounds upon which such licenses are granted.").

123. *See* TRIPS Agreement, *supra* note 47, art. 31.

124. *Doha Declaration*, *supra* note 47, ¶ 7. Despite this flexibility, patent protection for pharmaceuticals has already been established in all but a few African LDCs. *See* CARLOS M. CORREA, IMPLICATIONS OF THE DOHA DECLARATION ON THE TRIPS AGREEMENT AND PUBLIC HEALTH 38 (World Health Org., EDM Series No. 12, 2002); PHIL THORPE, STUDY ON THE IMPLEMENTATION OF THE TRIPS AGREEMENT BY DEVELOPING COUNTRIES 1 (Comm'n on Intellectual Prop. Rights, Study Paper 7, 2004).

125. *See* FINK & REICHENMILLER, *supra* note 48, at 1 tbl.1. The majority of these are developing countries. Congress has approved agreements with Vietnam, Jordan, Singapore, Chile, Morocco, and Australia. The CAFTA Agreement, which includes the Dominican Republic, Costa Rica, El Salvador, Guatemala, Honduras, and Nicaragua, and the agreement with Bahrain have been signed but not yet approved by Congress. Agreements are currently under negotiation with three Andean countries (Columbia, Ecuador, and Peru), Thailand, Panama, the countries of the Southern African Customs Union (South Africa, Namibia, Botswana, Lesotho, and Swaziland), and the group of countries involved in the Free Trade Area of the Americas. *Id.*

126. *Id.* at 2.

have data exclusivity provisions that may sharply limit or even eliminate the signatories' abilities to use the flexibilities provided in TRIPS and reaffirmed by the Doha Declaration.[127]

These trade agreements are subject to divergent interpretations, and countries could insist that the agreements be interpreted in ways that are consistent with the Doha Declaration. Furthermore, many countries have not yet signed such agreements and need only meet the minimum standards established by TRIPS. Nevertheless, the trend is clear. The most powerful governments on the international stage remain committed to an expansionist IP policy. In recent years, the United States initiated a dispute resolution aimed at Brazil's patent law,[128] threatened the South African government with trade sanctions because it sought to authorize the importation of cheaper medicines,[129] and repeatedly put countries such as Brazil, Thailand, India, and Argentina on the Special 301 watch list[130] because their patent laws did not meet with the approval of the U.S. pharmaceutical industry.[131] The access campaign has certainly drawn attention to this issue and achieved some victories in the short term.[132] However, both the United States and Europe have made it clear—through their positions in

127. *Id.* Note that the U.S. Trade Representative (USTR) is arguably exceeding its mandate in these negotiations, which requires it to "respect the Declaration of the TRIPS Agreement and Public Health adopted at Doha." Bipartisan Trade Promotion Authority Act of 2002 § 2102(b)(4)(C), 19 U.S.C.A. § 3802(b)(4)(C) (West 2004).

128. *See* Outterson, *supra* note 72, at 225. The United States withdrew its request for a WTO panel only after substantial international pressure. *Id.*

129. *See* Ravi Nessman, *South Africa Fights over AIDS Drugs*, ASSOCIATED PRESS, Mar. 5, 2001 (recalling threats by the United States following the passage in 1997 of an amendment to permit compulsory licensing in South Africa), http://www.aegis.com/news/ap/2001/AP010302.html.

130. The Special 301 watch list identifies countries that, in the judgment of the USTR, do not provide adequate protection for U.S. intellectual property. *See* 19 U.S.C. § 2241 (2000). Section 301 of the Trade Act authorizes the executive to impose trade sanctions against such states. *See id.* §§ 2411, 2414.

131. *E.g.*, OFFICE OF U.S. TRADE REPRESENTATIVE, 2004 SPECIAL 301 REPORT (2004), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2004/2004_Special_301/asset_upload_file16_5995.pdf; OFFICE OF U.S. TRADE REPRESENTATIVE, 2003 SPECIAL 301 REPORT (2003), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2003/2003_Special_301_Report/asset_upload_file665_6124.pdf; OFFICE OF U.S. TRADE REPRESENTATIVE, 2002 SPECIAL 301 REPORT (2002), http://www.ustr.gov/assets/Document_Library/Reports_Publications/2002/2002_Special_301_Report/asset_upload_file567_6367.pdf; *see* DRAHOS WITH BRAITHWAITE, *supra* note 47, at 93-95 (discussing the role intellectual-property-based industries play in the 301 process).

132. *See, e.g.*, Zita Lazzarini, *Making Access to Pharmaceuticals A Reality: Legal Options Under TRIPS and the Case of Brazil*, 6 YALE HUM. RTS. & DEV. L.J. 103, 132 (2003); Nessman, *supra* note 129.

bilateral and regional free trade negotiations and at the WTO[133]—that their policies are fundamentally unchanged and still aggressively favor strong IPRs.

The formal and informal pressures exerted by such nations circumscribes the willingness and ability of LMI country governments to use the flexibilities technically open to them. Until last year, Brazil was the only developing country that had successfully used the threat of compulsory licensing to obtain lower-price ARVs.[134] (Brazil, as one of the world's ten largest economies,[135] is in an unusually strong position for a developing country, and has substantial indigenous capacity to reverse engineer and produce medicines.) Only recently, and quietly, have other LMI countries began to issue compulsory licenses covering ARVs.[136]

Failures in accountability and leadership also contribute to the problem in some countries. Many governments are not committed to addressing the needs of the destitute sick within their borders.[137] For example, over the last decade, South Africa, India, and China have come under fire for denying the scope, or even existence, of the HIV/AIDS problem in their countries.[138] Finally, even LMI countries that have some will to address the ac-

133. *See* Duncan Matthews, *WTO Decision on Implementation of Paragraph 6 of the Doha Declaration on the TRIPS Agreement and Public Health: A Solution to the Access to Essential Medicines Problem?*, 7 J. INT'L ECON. L. 73, 86-89, 93 (2004) (describing the United States and European Union positions during recent WTO negotiations over countries' ability to export under a compulsory license).

134. Brazil has repeatedly used the credible threat of compulsory licensing to effectively obtain discounts. *See Brazil's National STD/AIDS Programme Announces Largest Drug Price Reduction Deals in Five Years*, KAISER DAILY HIV/AIDS REP., Jan. 20, 2004, http://www.kaisernetwork.org/daily_reports/rep_index.cfm?hint=1&DR_ID=21751.

135. *See* World Bank, PPP GDP 2003, *at* http://www.worldbank.org/data/databytopic/GDP_PPP.pdf; World Facts Index, History of Brazil, *at* http://worldfacts.us/Brazil-history.htm (last visited Feb. 26, 2005).

136. *See, e.g.*, *Cipla Gets Malaysian Nod for AIDS Drugs*, BUS. STANDARD (India), Feb. 25, 2004 (reporting Malaysia's recent compulsory license); Martin Khor, *Patents vs. Access to Medicines at AIDS Conference*, DAILY NEWS (Sri Lanka), Aug. 10, 2004 (reporting Mozambique's recent compulsory license), http://dailynews.lk/2004/08/10/fea11.html.

137. This may be particularly true where such individuals are ill with a disease as stigmatized as HIV/AIDS.

138. *See, e.g.*, Sara Davis, Opinion, *Hold Beijing To Account for Its AIDS Coverup Before the 2008 Olympics*, INT'L HERALD TRIB., Aug. 25, 2004, at 8; Diddier Fassin & Helen Schneider, *The Politics of AIDS in South Africa: Beyond the Controversies*, 326 BMJ 495 (2003); Michael Specter, *India's Plague*, NEW YORKER, Dec. 17, 2001, at 74. There has been some progress on this front for HIV/AIDS. *See, e.g.*, Lawrence K. Altman, *South Africa Says It Will Fight AIDS with a Drug Plan*, N.Y. TIMES, Aug. 9, 2003, at A1.

cess problem may be derailed by the inauspicious state of their intellectual property laws[139] and lack of expertise in applying these laws, as well as by the burdensome administrative conditions that TRIPS imposes. Because TRIPS requires a case-by-case determination of any compulsory license, developing country governments must establish administrative capacity to take consistent, rapid action wherever patents pose pricing barriers.[140] Similarly, TRIPS may create particular obstacles for countries without their own manufacturing capacity.[141] Some of these administrative costs can be lowered if countries gain experience using these channels. But given the likelihood that rich and poor countries alike will prioritize the wishes of corporate interest groups over the needs of the poor, no strategy involving top-down change alone is likely to remedy the static costs of the global IP regime.

2. Top-Down Change To Address the R&D Gap

Some of the most creative and promising proposals for addressing the R&D gap also follow a top-down model. Several academics have proposed prize systems to compete with or displace the patent system by rewarding inventors according to the therapeutic value their product ulti-

---

139. *See* SISULE F. MUSUNGU ET AL., UTILIZING TRIPS FLEXIBILITIES FOR PUBLIC HEALTH PROTECTION THROUGH SOUTH-SOUTH REGIONAL FRAMEWORKS 24-25 (2004).

140. *See* TRIPS Agreement, *supra* note 47, art. 31(a) (requiring licenses to be decided on the basis of their individual merits). The requirement that in most instances the applicant first make "reasonable" efforts to obtain a license, *see id.* art. 31(b), can also generate substantial delay if strict parameters for reasonableness are not imposed. Finally, countries must afford right holders "adequate remuneration," *id.* art. 31(h), and a form of "judicial review or other independent review." *Id.* art. 31(i). Establishing procedures to meet these requirements can be burdensome for countries with limited resources.

141. That is because TRIPS requires that any use without the authorization of the patent holder "be authorized predominantly for the supply of the domestic market of the Member authorizing such use." *Id.* art. 31(f). This provision has been the subject of intense focus recently, because it could prevent countries without the ability to produce medicines domestically from being able to purchase generics from a country that can. At Doha, the Ministerial agreed to address the issue, and after several years of negotiation, a temporary solution was adopted just prior to the Cancún meeting in 2003. *See* World Trade Org., General Council, *Implementation of Paragraph 6 of the Doha Declaration on the TRIPS Agreement and Public Health*, WT/L/540 (Sept. 1, 2003) [hereinafter WTO, *Paragraph 6*]. The decision provides that, under limited circumstances, and subject to strict and potentially onerous reporting requirements, countries are free to export generic products for the sole benefits of countries lacking manufacturing capacity. The fix has been heavily criticized, *see* Elizabeth Becker, *Cheaper Medicines for the World's Poor; Trade Rules Altered on Patented Drugs*, INT'L HERALD TRIB., Sept. 2, 2003, at 1, but there is little indication that countries are willing to reopen their acrimonious negotiations. For an in-depth discussion of these issues, see Abbott, *supra* note 76.

mately offers.[142] Internationally, advocates have recently proposed an R&D treaty or convention that would set minimum levels of contribution to R&D and weigh national contributions to facilitate investment into neglected public goods.[143]

Most of the scholarly solutions proposed to address anticommons and thicket problems in the United States involve top-down change as well. For example, Richard Epstein has suggested that patent doctrine be interpreted to preclude the patenting of genome fragments (known as expressed sequence tags or ESTs) that have more blocking value than use value.[144] The Federal Circuit's decision in *Madey* has spurred new interest in a statutory research exemption.[145] Professors Arti Rai and Rebecca Eisenberg have suggested that the Bayh-Dole Act, which allows recipients of federal funds to patent and exclusively license federally-funded research,[146] should be revised to give federal agencies more power to require

142. *See* Hollis, *supra* note 58, at 18 (proposing a prize system that sets awards according to the number of "disability adjusted life-years"—a common measurement of morbidity and mortality—that a product avoids).

143. *See* Tim Hubbard & James Love, *A New Trade Framework for Global Healthcare R&D*, 2 PLoS BIOLOGY 147, 148 (2004), *available at* http://biology.plosjournals.org/archive/1545-7885/2/2/pdf/10.1371_journal.pbio.0020052-S.pdf; Letter from James Love et al. to the WHO Executive Board and the WHO Commission on Intellectual Property, Innovation, and Health (CIPIH) (Feb. 24, 2005) (copy on file with authors) (requesting that the WHO evaluate a proposal for a new global medical R&D treaty). Alternatively, others have suggested that countries (or international agencies or foundation) should fund such research through "pull programs," such as advanced purchase commitments. *See, e.g.*, Kremer, *supra* note 58, at 82.

144. Richard A. Epstein, *Steady the Course: Property Rights in Genetic Material*, *in* PERSPECTIVES ON PROPERTIES OF THE HUMAN GENOME PROJECT 153, 188-193 (F. Scott Kieff ed., 2003), *available at* http://www.wulaw.wustl.edu/Academics/Faculty/Bios/Kieff/HGPIP/Final/GEN_50_CH8.pdf.

145. For example, the American Intellectual Property Law Association (AIPLA) recently endorsed legislation to establish a statutory research exemption for efforts to understand and evaluate the validity of the patent, to find other methods of making or using the patented subject matter, or to find substitutes for the patented subject matter. AM. INTELLECTUAL PROP. LAW ASS'N, AIPLA RESPONSE TO THE NATIONAL ACADEMIES REPORT ENTITLED "A PATENT SYSTEM FOR THE 21ST CENTURY" 25-26 (2004), http://www.aipla.org/Content/ContentGroups/Issues_and_Advocacy/Comments2/Patent_and_Trademark_Office/2004/NAS092304.pdf. The American Academy of Arts and Sciences has also convened a working group to consider options for a domestic and international research exemption. *See* Science & Intellectual Prop. in the Public Interest, Research Exemption Working Group, *at* http://sippi.aaas.org/rschexemption.shtml (last visited Mar. 5, 2005).

146. Bayh-Dole University and Small Business Patent Procedures Act, Pub. L. No. 96-517 § 6(a), 1980 U.S.C.C.A.N. (94 Stat.) 3015, 3018-29 (1980) (codified at 35 U.S.C. §§ 200-212 (2000)). The goals of the Act are to, inter alia, "promote the utilization of inventions arising from federally supported research," "promote the commercialization

grantees to dedicate their research outputs to the public domain.[147] But, so far, all of these proposals have fallen on deaf ears.

The fact that the trend on the domestic and international stage has been towards stronger, rather than weaker, IPRs[148] bodes ill for the most ambitious of these top-down proposals. This suggests such strategies will need to be supplemented or catalyzed by solutions from another arena that can circumvent blockages within international and national political systems.

3. Private Sector Voluntary Concessions Regarding Access and R&D

Systemic change could also be initiated by the private, for-profit sector.[149] Unfortunately, history suggests that although the private sector can be pushed, it will not lead. Patent-based drug companies agreed to major price reductions for first-line AIDS therapies, but only after prolonged public outcry.[150] The experience of the Accelerating Access Initiative, a joint effort between U.N. agencies and five major pharmaceutical companies to achieve discounts for AIDS medicines for developing countries, is instructive here. A reporter for *The Washington Post* who interviewed most of those involved in creating the program has detailed its many failings and concluded that:

---

and public availability of inventions," and "ensure that the Government obtains sufficient rights in federally supported inventions to meet the needs of the Government and protect the public against nonuse or unreasonable use of inventions." *Id.* § 200.

147. Arti K. Rai & Rebecca S. Eisenberg, *Bayh-Dole Reform and the Progress of Biomedicine*, 66 LAW & CONTEMP. PROBS. 289, 291 (2003).

148. *See* Sonny Bono Copyright Term Extension Act, Pub. L. No. 105-298, 112 Stat. 2827 (1998) (describing lengthening of the copyright term codified at 17 U.S.C. §§ 108, 203, 301-304); *supra* note 44 (describing expansion of patentable subject matter); *supra* note 46 (describing regulatory exclusivity in the United States and internationally).

149. Some have proposed, for example, that the market- and patent-based pharmaceutical sector proactively change its licensing practices or establish discounts and donation programs to address access concerns. *See* Attaran, *supra* note 71, at 163. *But cf.* Michael A. Friedman, Henk den Besten & Amir Attaran, *Out-Licensing: A Practical Approach for Improvement of Access to Medicines in Poor Countries*, 361 LANCET 341, 341 (2003) (admitting that "donations or discounts offer only limited, often imperfect, solutions").

150. *See, e.g.*, 't Hoen, *supra* note 81, at 294 ("[U]ntil January 2003, . . . one pharmaceutical company was charging $2000 a year more in Guatemala than in Switzerland for its AIDS drug. Only after months of public pressure did the price of the drug come down in Guatemala." (citing *Roche Cuts Price of AIDS Drug to Nations*, ASSOCIATED PRESS (Feb. 13, 2003); *Drug Company Cuts AIDS Drug Prices in S. Africa*, *supra* note 27. *But cf.* Boelaert et al., *supra* note 73, at 840 ("This impressive discount . . . was not merely due to public outcry, but mostly as a response to competition by generic drugs.").

> The drug firms sought to maintain prices in most markets by offering selective discounts that would remain under their control . . . [, i]n the long term, . . . building demand while limiting the duration and scope of the discounts. Most of all, the drug companies wanted to squelch an increasingly damaging debate on prices and patents that the U.N. agencies had helped touch off.[151]

Even today, voluntary discounts have resulted in prices that typically remain above the lowest price for generic versions.[152] They are also often limited by territory or sector in ways that sharply undermine their impact.[153] In sum, they have been applied as grease to squeaky wheels—sporadically and no more liberally than is required to quiet the noise. As a result, voluntary discounts have proven neither efficient nor sufficient.[154]

---

151. Gellman, *supra* note 23. For example, Pfizer refused to join at the inception of the program, insisting that the U.S. prices of Pfizer's medicines were "good value" and worrying that any discussion of differential pricing would threaten its "'core markets.'" *Id.* (recounting an anonymous source's account of statements made by Pfizer's Senior Vice President Ian C. Read). Some participating companies wanted beneficiary countries to explicitly renounce any use of compulsory licensing or parallel importing in exchange for the price concessions. Eight months after the initiative was announced with great fanfare, four of the five companies still refused to reveal the discounts being offered. Discounts were offered on a country-by-country basis and were only made available to the public sector, which in most countries was not providing treatment. A U.N. official who attempted to obtain the same discounted prices for large private sector firms that wanted to provide ARVs to their employees—a plan which could have financed treatment for one million patients in five years—recounts the pharmaceutical companies' response: "'They laughed at us.'" *Id.*

152. For example, although BMS asserted it was selling stavudine below cost in Africa, generic companies have been able to undercut its prices by almost seventy percent. *See* Meldrum, *supra* note 29. Similarly, the latest summary of best available worldwide prices for ARVs shows that generics are cheaper than proprietary products for seventeen out of the twenty formulations for which there are both generic and proprietary suppliers. MSF, UNTANGLING THE WEB, *supra* note 6, at 9-11 tbl.1a.

153. MSF, UNTANGLING THE WEB, *supra* note 6, at 15-19 tbl.2 (reporting diverse restrictions according to geography and the purchasing entity). Note also that discounts on some second-line therapies are still far from adequate, threatening a looming fiscal crisis in developing country ARV programs as first-line drugs fail and need replacement. *See* Médecins Sans Frontières, A Guide to the Post-2005 World: TRIPS, R&D, and Access to Medicines (Feb. 25, 2005) (showing that second-line therapies currently cost twenty-six times the amount that first-line therapies cost), *at* http://www.msf.org/countries/page.cfm?articleid=88694E5B-0FED-434A-A21EDA1006002653.

154. *See* OXFAM INT'L, SAVE THE CHILDREN & VSO, BEYOND PHILANTHROPY: THE PHARMACEUTICAL INDUSTRY, CORPORATE SOCIAL RESPONSIBILITY, AND THE DEVELOPING WORLD (2002); SMITH, *supra* note 70 (finding price discounts less effective than generic competition); 't Hoen, *supra* note 81, at 294 (complaining that ad hoc execution of differential pricing schemes and donation programs has resulted in "efforts [that] have been neither systematic nor sufficient"); Letter from Eugene Schiff, Caribbean Coordina-

The same holds true for voluntary licensing agreements. Until a few years ago, pharmaceutical companies routinely rebuffed requests for voluntary licenses on ARVs.[155] Following the Yale/BMS concession, several major firms offered licenses to South African generic companies, but limited them to the public sector.[156] Unfortunately, these licenses have remained few and far between.[157] Still today, MSF is unable to obtain voluntary licenses from patent-holding companies to use fixed-dose combinations of ARVs in South Africa and China.[158]

Additionally, while companies currently may look the other way when their AIDS drug patents are infringed in poor, heavily affected countries,[159] it is not clear that such forbearance will persist. It was not the

---

tor, Agua Buena Human Rights Association et al., to Mr. Andy Schmeltz & Ms. Konji Sebati, HIV Program, Pfizer (Sept. 17, 2004) (on file with authors) (enumerating the flaws with Pfizer's fluconazole donation program in the Dominican Republic).

155. *See, e.g.*, Letter from Cipla to the South African Registrar of Patents (Mar. 7, 2001) (requesting a compulsory license on several AIDS medicines for the South African market and noting that their requests for voluntary licenses had been rebuffed), *available at* http://www.cptech.org/ip/health/sa/ciplanetsh03072001.html.

156. *See, e.g.*, Aspen Pharmacare Press Release, *supra* note 28.

157. GlaxoSmithKline and Boehrenger Ingelheim refused to extend their licenses to the private sector until 2003, when they were faced with an impending judgment by the South African Competition Commission in a suit charging them with unfair trade practices, including excess pricing of their antiretroviral medicines, and seeking a compulsory license to produce the drugs. *Reducing the Price of Antiretroviral Medicines*, TAC NEWSLETTER (Treatment Action Campaign, S. Afr.), Oct. 27, 2003, http://www.tac.org.za/newsletter/2003/ns28_10_2003.htm. To avoid setting this precedent and becoming subject to compulsory licenses, the companies settled, and as a result have since entered into additional licenses that permit private sector sales. *See, e.g.*, *GlaxoSmithKline Issues Voluntary License for Lamivudine, Zidovudine to South African Generic Drug Company*, KAISER DAILY HIV/AIDS REP., July 1, 2004, *at* http://www.kaisernetwork.org/daily_reports/rep_index.cfm?hint=1&DR_ID=24507. Merck followed suit by granting a voluntary license to its AIDS drug to the main South African generic manufacturer. *See* Press Release, Merck & Co., Inc., Grants License for HIV/AIDS Drug Efavirenz to South African Company, Thembalami Pharmaceuticals (July 13, 2004), http://www.pressmethod.com/releasestorage/5003645.htm. There has also been one such license in Kenya. *See* Press Release, GlaxoSmithKline, GlaxoSmithKline Grants a Fourth Voluntary License for the Manufacture and Sale of HIV/AIDS Medicines in Africa (Sept. 22, 2004), http://www.gsk.com//press_archive/press2004/press_09222004.pdf.

158. E-mail from Ellen 't Hoen, Acting Director, Campaign for Access to Essential Medicines, Médecins Sans Frontières, to Amy Kapczynski (Jan. 20, 2005).

159. For example, a search of the Tanzanian Food and Drugs Authority website, http://www.tfda.or.tz, reveals that several generic forms of AZT (zidovudine) are registered in the country, despite the fact that AZT is patented there, *see* Tanzanian Patent No. 2429 (issued Sept. 30, 1991). The Clinton Foundation HIV/AIDS Initiative has also publicly stated that it intended to supply generics, including AZT, to the Tanzanian market, *see* Lawrence K. Altman, *Clinton Group Gets Discount for AIDS Drugs*, N.Y. TIMES,

norm with regard to ARVs before the political tide turned,[160] and it will likely not extend to diseases that garner less political attention. Indications are that patent-based companies are still quite willing to use exclusive rights to extract rents in even the poorest countries.[161] Having advocated vigorously for maximum IP protections around the world, proprietary companies will presumably exploit the protections when they perceive it to be in their interests. It is fair to conclude, therefore, that the for-profit drug sector will not take positive action to address the static costs of IPRs in LMI countries unless others take the initiative and raise the costs of inaction for the for-profit drug sector. Furthermore, experience suggests that any such initiative must be carefully crafted to make commitments easy to enforce and to minimize withdrawal opportunities for companies.

There are only slightly more encouraging signs from the private sector in the R&D domain. Companies have done little on their own to develop drugs for neglected diseases, but some have been persuaded to contribute to the public-private partnerships we describe below. There are more promising signs that the private sector—or at least some parts of it—will help address the barriers that upstream patents can pose for researchers. Just as information technology companies are beginning to see the business sense in free and open source software, some companies that invest heavily in biomedical research are beginning to see the logic of investing in the public domain. Merck recently invested millions of dollars in a public genomics database because it "sees gene sequences as inputs, rather

---

Oct. 24, 2003, at A8, but has not announced an intention to obtain a license there. Despite this, there have been no reports of infringement actions.

160. In 2000, GlaxoSmithKline sent cease-and-desist letters to Cipla regarding the generic company's activities in Uganda and in Ghana. *See* Gellman, *supra* note 23 (noting that Glaxo sent Cipla a cease-and-desist letter in Uganda in November 2000); Mark Schoofs, *Glaxo Enters Fight in Ghana on AIDS Drug*, WALL ST. J., Dec. 1, 2000, at A3 (reporting that Glaxo had issued a cease-and-desist letter to Cipla, causing it to stop importing ARVs, and, remarkably, that Glaxo appeared not to hold the cited patents in Ghana).

161. Recently, for example, the Swiss pharmaceutical company Novartis obtained exclusive marketing rights (EMR) in India for its drug Gleevec, which treats chronic myeloid leukemia and has no therapeutic equivalents. Most of the generic companies producing Gleevec before the EMR issued left the market, and Novartis brought suit to enjoin the others from selling the drug. The company charges more than ten times the generic price. *See* Prati Jatania, *In Search of the Sugar-Coating: The New Product Patents Regime Will Decide the Future of Hundreds of Leukemia Patients*, INDIAN EXPRESS, Dec. 19, 2004. As a result, the Indian government is apparently considering withdrawing the EMR. *See* Priya Ranjan Dash, *Govt Puts Novartis Cancer Drug on Notice*, TIMES INDIA, Feb. 15, 2005, http://timesofindia.indiatimes.com/articleshow/1022035.cms.

than end products."[162] But, of course, companies that operate at the other end of the research spectrum have different incentives and are using their influence to prevent changes that would create more freedom for researchers.[163] Because firms have different interests in this area, and because it is unclear which side, if either, will prevail in a contest between them, we cannot rely entirely on the private sector to solve the problems that patents cause for research.

In conclusion, both theory and experience give us reason to believe that neither governments nor firms will act spontaneously and systemically to close the R&D gap, or to eliminate static costs created by the contemporary global IPR regime.[164] Given the stakes, there is an acute need for new models of IPR management.

III. MODELS AND LESSONS OF COMMONS-BASED PRODUCTION

In this Article, we propose two models that avoid roadblocks set up by governments and industry and that have the potential to catalyze wider change. These approaches draw on recent literature and experience with commons-based production modalities. We use the term "commons-based" to signify forms of production and coordination that rely on a mechanism other than proprietary exclusion and that treat all actors symmetrically vis-à-vis the resource in question. Commons-based initiatives offer a model by which a network of independent but interconnected participants can choose to act—not to change the legal system, but to change their practices within it. In so doing, they can circumvent barriers posed by standard applications of exclusive rights, such as patents and copyrights, and by rent-seeking lobbying that blocks statutory and regulatory change. These efforts do not rely on government action or private-sector, price-

162. *See* Robert P. Merges, *A New Dynamism in the Public Domain*, 71 U. CHI. L. REV. 183, 188 (2004).

163. Several companies that supply reagents and research equipment are organizing against the AIPLA-proposed statutory research exemption discussed *supra* note 145. *See* Memorandum from Janet Lynch Lambert of Invitrogen & Paul Grossman of Applied BioSystems to Interested Members of the Life Science Community (Mar. 8, 2005) (on file with authors).

164. Commentators anticipate that the access gap will grow wider in years to come, citing factors including the continued growth and influence of multinational pharmaceutical companies and the strengthening of IP protections through international agreements. *See, e.g.*, OXFAM INT'L, UNDERMINING ACCESS TO MEDICINES: COMPARISON OF FIVE US FTAS (2004), http://www.oxfamamerica.org/pdfs/fta_comparison.pdf; Mary Crewe, *Spectacular Failure—A View from the Epicenter*, 4 YALE J. HEALTH POL'Y L. & ETHICS 157, 160 (2004).

driven, market-mediated solutions, but on collaborative practices buttressed by contractual tools that apply property-like rights to ensure access and distribution rather than control and exclusion.

This Part describes the commons-based projects that have proliferated in recent years in the area of the production and distribution of information. It also discusses the innovative contractual forms that sustain many of these initiatives. Finally, this Part shows that universities and other public sector institutions are already beginning, in fragmentary and preliminary ways, to adapt some of these models to the biomedical domain.

A. Commons-Based Production Models

The wide range of open source and free software created by programmers who freely contribute their time and talent to collaborative efforts confounds the historic presumptions of property law. These presumptions say that property rights, price signals, and managers are necessary to organize and incentivize efficient production.[165] Free and open source projects, ranging in size from projects with merely two or three programmers to large-scale projects like the Linux kernel, use none of these presumptions and yet produce high-quality software that has come to occupy an increasingly prominent place in the information technology economy.[166]

Free and open source software could not have flourished in this way without the legal innovation embodied in the GNU General Public License (GPL).[167] The GPL was developed in the 1980s by Richard Stallman, a programmer from MIT who sought a way to protect the historically collaborative mode of software development[168] from the encroachment of firms that wanted to make software proprietary.[169] The GPL has two key components. First, it gives users the right to copy, alter, and distribute the software source code, as modified or in its original form. Second, it includes a "copyleft" requirement, obliging those who create derivative code to grant the same rights to those who receive the derivative software.[170] Thus, the GPL not only shares but also requires others who benefit from the license to share their own contributions. The GPL turns copyright on its head, by guaranteeing rights to use, learn, freely distribute, and modify,

165. *See* Benkler, *supra* note 39, at 372.

166. *See* WEBER, *supra* note 38, at 5-6.

167. *See* The GNU General Public License (GPL), *at* http://www.opensource.org/licenses/gpl-license.php (last visited Mar. 9, 2005). Although there are now many kinds of open source software licenses, the GPL is by far the most commonly used. *See* Lerner & Tirole, *supra* note 38, at 23 tbl.1.

168. Lerner & Tirole, *supra* note 38, at 4.

169. WEBER, *supra* note 38, at 46-47.

170. *Id.* at 182.

but not the right to exclude. This legal jiu-jitsu is well-suited to the cooperative nature of peer-produced software and its reliance on reciprocal sharing of innovation.[171] It has also been a model for other commons-based initiatives seeking to arm themselves against the rapid expansion of exclusive rights to information and culture over the past few decades.

Creative Commons is one of the most rapidly growing of these initiatives. It offers authors and artists a series of simple licenses that allow them to contract around the default in copyright law that reserves for them "all rights" in their creative works.[172] Using the Creative Commons website, individuals can choose between a menu of eleven licenses. The Attribution License, for example, permits content to be freely shared, modified, and commercially used, as long as the original author is given credit.[173] The Noncommercial License allows the same activities, but only for non-commercial purposes.[174] There is also a Share Alike license, which requires that any derivative works be distributed under the same terms as the original work.[175]

In the academy, commons-based production has become an important model for scientific publishing. The recently-created Public Library of Science (PLoS) offers peer-reviewed Internet-based content free to readers.[176] It covers the production costs of its journals with philanthropic donations and per-page-fees paid by authors and ensures the free distribution of articles by applying the Creative Commons Attribution License to them.[177] The National Institutes of Health (NIH) has recently adopted a policy intended to improve the public's access to publications resulting from NIH-funded research.[178] The policy calls upon scientists to submit the final-version-accepted-for-publication manuscripts to the NIH, and provides that the manuscripts will be made freely available on the Internet through the NIH's digital archive, PubMed Central, within twelve months of their final publication.[179]

---

171. *See* Benkler, *supra* note 39, at 379-80.

172. Creative Commons, About Us, *at* http://creativecommons.org/about/history (last visited Apr. 28, 2005).

173. Creative Commons, Licenses Explained, *at* http://creativecommons.org/about/licenses (last visited Apr. 28, 2005).

174. *Id.*

175. *Id.*

176. Pub. Library of Sci., About PLoS, *at* http://www.publiclibraryofscience.org/about/index.html (last visited Apr. 20, 2005).

177. *Id.*

178. *See* Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research, 70 Fed. Reg. 6891, 6899-900 (Feb. 9, 2005).

179. *Id.*

Genomics research has been another major area for commons-based initiatives. The most prominent of such efforts is the Human Genome Project (HGP), a publicly funded, international research project that committed itself to releasing its data and not claiming patent rights in the mapped genome.[180] Many of the follow-on projects which seek to functionally specify genomic sequences and create maps useful for applied research have also adopted commons-based structures. The Ensembl Genome Browser uses open source software to create free, annotated maps of primarily mammalian genomes.[181] The HapMap project, which seeks to identify haplotypes (shared genetic variations) to help researchers better understand and address diseases with a genetic component, is also commons-based.[182] Like the HGP, HapMap makes its data available for free on the Web. Unlike the HGP, it took the additional step of creating a click-wrap license to prevent those accessing its data from combining it with their own data and patenting the results.[183]

The recently-launched Biological Innovation for Open Society (BIOS) project is perhaps the most self-conscious inheritor of both the lessons and tools of the free software movement.[184] A nonprofit created by the Australian organization CAMBIA, BIOS seeks to catalyze the creation of a new,

---

180. *See* John Sulston, *Intellectual Property and the Human Genome*, *in* GLOBAL INTELLECTUAL PROPERTY RIGHTS: KNOWLEDGE, ACCESS AND DEVELOPMENT 61, 64 (Peter Drahos & Ruth Mayne eds., 2002).

181. *See* Ewan Birney et al., *An Overview of Ensembl*, 14 GENOME RES. 925, 925 (2004); Ensembl Genome Browser, *at* http://www.ensembl.org (last visited Apr. 20, 2005).

182. *See* Int'l HapMap Project, *at* http://www.hapmap.org (last updated Mar. 3, 2005).

183. *See* Int'l HapMap Project, Registration for Access to the HapMap Project Genotype Database, *at* http://www.hapmap.org/cgi-perl/registration (last updated Mar. 3, 2005). The license was "not intended to block the ability of users to file for intellectual property protection on specific haplotypes for which they have identified associated phenotypes, such as disease susceptibility, drug responsiveness, or other biological utility," but merely to preserve public access to HapMap data. *Id.* This requirement likely stems from the conflict between the HGP and Celera, a private company that made use of HGP data but kept its own secret, and sought to patent resulting gene sequences. *See* Sulston, *supra* note 180, at 64. This restriction on so-called "parasitic patenting" has since been dropped, for two reasons. The HapMap consortium felt that the map and surrounding science has advanced to the stage where any haplotypes derived from their released data would be obvious and thus unpatentable. In addition, the leadership was concerned that the license prevented their data from being included in other public genome databases. *See* Press Release, National Institutes of Health, International HapMap Consortium Widens Data Access (Dec. 10, 2004), http://www.nih.gov/news/pr/dec2004/nhgri-10.htm.

184. *See* Biological Innovation for Open Society (BIOS), *at* http://www.bios.net (last updated Apr. 17, 2005).

self-sustaining commons for researchers in the field of agricultural biotechnology.[185] It aims to do this by creating portfolios of essential biotech research tools and licensing them under a GPL-style license. The scientist behind the initiative, Richard Jefferson, has already created two technologies that engineer around proprietary tools critical for biotechnology-based crop improvement.[186] Licensees who want access to these technologies must accept the terms of the BIOS license, which requires them to share and make available to other participants in the initiative any improvements they make to the core licensed technology.[187] Licensees are permitted to patent and license any products they develop—as distinguished from improvements on the tools licensed by BIOS—in whatever way they wish, and uses of the licensed technology are not limited by territory or field. BIOS, like the HGP and HapMap Projects, is betting that certain research tools are shareable, even in wealthy markets and under current IP regimes, because the tools' research value is greatest if they are freely accessible.

B. Lessons from Commons-Based Production Models

The initiatives described in the preceding Section represent a class of solutions to information production problems. They demonstrate that, in response to the new enclosure movement, collective action can successfully coordinate cooperative, open-access initiatives to produce and distribute innovations to target groups of users and researchers. The first lesson, then, is that commons-based modalities can play an important role in information production, including in the biomedical sector. The second lesson is that new contractual regimes are essential to the success of some of these initiatives. Only one of the projects mentioned above—the Human Genome Project—adhered to the classic public domain model and dedicated its outputs to the public without further restriction. Free software projects, the HapMap project (initially) and the BIOS initiative, all operate by conditioning access to their benefits on reciprocal sharing of appropriately defined improvements. They create a self-binding commons rather than an unrestricted public domain.

---

185. CAMBIA CEO Richard Jefferson describes the motivation behind the effort: "'So much of what we want to do is all tied up in somebody's intellectual property . . . . It's a complete sclerotic mess, where nobody has any freedom of movement. Everything that open source has been fighting in software is exactly where we find ourselves now with biotechnology.'" Thomas Goetz, *Open Source Everywhere*, WIRED, Nov. 2003, http://www.wired.com/wired/archive/11.11/opensource_pr.html.

186. *See* Andrew Pollack, *Open-Source Practices for Biotechnology*, N.Y. TIMES, Feb. 10, 2005, at C8 (quoting Gary Toenniessen of the Rockefeller Foundation).

187. The CAMBIA BIOS License Agreement Version 1.1, *at* http://www.bios.net/daisy/license/210 (last updated Feb. 8, 2005).

Because information is nonrivalrous, its sharing and use in a commons raises none of the allocation concerns characteristic of a physical commons. The only economic concerns raised by an information commons are ones of provisioning—that is, how the innovation will be paid for ex ante.[188] As we demonstrated in the preceding Part, where the information goods in question are specific to diseases affecting developing countries, innovation will likely have to be paid for by public or philanthropic sources because of the small size of associated markets. The right to produce drugs solely for use in LMI countries, or sharing rights to do research into diseases that disproportionately affect the poor, will therefore have little effect on incentives to invest more generally in commercial R&D.

It therefore comes as little surprise that biomedical research institutions, particularly in the public sector, are increasingly adopting commons-based strategies to promote production and access to information. Some have begun to utilize sharing principles to address the access gap, relying upon the fact that supra-marginal returns in developing countries are not necessary to the development of many health-related products. These initiatives piggyback on research funded by public or philanthropic institutions or by private investors seeking returns in rich country markets, and adopt contractual terms to ensure that resulting products will be available at low cost in developing countries. The Yale/BMS agreement not to enforce Yale's stavudine patent in South Africa is perhaps the most prominent example of this approach, but several others have recently emerged. For example, for technologies with a worldwide market, the NIH has begun to adopt licensing terms requiring North American and European companies to "provide a marketing plan for making products available to developing countries."[189] The University of California at Berkeley recently signed a Memorandum of Understanding with the government of Samoa for rights to an antiviral compound, which the University hopes to develop into an AIDS drug. In the Memorandum, the parties agree that Berkeley will pay royalties to the government and local communities on sales of any eventual end product, and both parties agree "to license their respective intellectual property rights so that prostratin (if it is approved as

188. In the context of universities, the answer is largely a combination of government grants, tuition, and philanthropic giving. *See infra* notes 212-13 and accompanying text. Strategies that abstain from enforcing IPR exclusivity therefore have little effect on provisioning by the academic sector.

189. Luis A. Salicrup et al., *An Innovative Program To Move Biomedical Health Technologies from the Laboratory to Worldwide Application*, 12 IP STRATEGY TODAY 1, 7 (2005). Because these provisions are new, none have been enforced yet. *Id.*

an anti HIV-AIDS therapy) is made available to developing nations at minimal cost."[190]

Public sector institutions have also begun to adopt commons-based, open licensing approaches to address the R&D gap. In September 2004, the Office of Technology Transfer at the NIH announced its intention to develop U.S.-owned technology for a rotavirus vaccine by offering partially-exclusive, regional licenses to companies in developing countries.[191] This model suggests that by working with a diverse array of partners in LMI countries, innovators can find ways, even under current market conditions and without the injection of additional public or philanthropic funds, to develop technologies for neglected diseases and simultaneously minimize the costs to patients that result from exclusivity.

In recent years, a number of nonprofit initiatives have also been launched to address the R&D gap. Some of them seek to develop medicines or vaccines for global diseases that cause high morbidity in developing countries. Prominent examples include the Global Alliance for TB Drug Development (TB Alliance), the International AIDS Vaccine Initiative (IAVI), and the Medicines for Malaria Venture. Others focus on a broader range of diseases, such as the Institute for OneWorld Health and the Drugs for Neglected Diseases Initiative.

The exact mode of operation of each initiative differs, in part because of the different characteristics of the diseases they target,[192] but they gen-

---

190. Memorandum of Understanding between the Government of Samoa and the Regents of the University of California, Berkeley for Disposition of Future Revenue from Licensing of Prostratin Gene Sequences, an Anti-Viral Molecule § VI (Aug. 13, 2004) (on file with authors); *see also* Press Release, University of California, Berkeley, Landmark Agreement Between Samoa and UC Berkeley Could Help Search for AIDS Cure (Sept. 29, 2004), http://www.berkeley.edu/news/media/releases/2004/09/29_samoa.shtml.

191. *See* Prospective Grant of Partially-Exclusive Licenses: Human-Bovine Reassortant Rotavirus Vaccine, 69 Fed. Reg. 57,335 (Sept. 24, 2004). Two Indian companies were offered co-exclusive licenses, a Brazilian company was offered an exclusive license, and Chinese companies were offered nonexclusive licenses. *Id.*; *see also* Salicrup et al., *supra* note 189, at 9 (noting that "[t]he degree of exclusivity was determined by the needs of prospective licensees in each country"). Note also that the licensed territories exclude the United States, Canada, and Europe. *See* Prospective Grant of Partially-Exclusive Licenses: Human-Bovine Reassortant Rotavirus Vaccine, 69 Fed. Reg. at 57,335.

192. For example, those working on global diseases that are found in rich countries are likely to be able to attract more interest from private industry partners. *See* Drugs for Neglected Diseases Initiative, Questions and Answers, http://www.dndi.org/cms/public_html/insidearticleListing.asp?CategoryId=160&ArticleId=309&TemplateId=2#mostnegdiseases (last visited Apr. 28, 2005) (noting that initiatives targeting HIV/AIDS, malaria, and TB "have relied heavily on market-based incentive mechanisms, including public-

erally support their operations through some combination of public and philanthropic funds and collaborations with private industry.[193] Those that have made their patenting and licensing policies public have indicated that they will either address access concerns by requiring their licensees to make subsequent inventions available, affordable, and accessible in target countries,[194] or by granting only nonexclusive licenses for sales to international agencies such as the WHO.[195] The TB Alliance also seeks to minimize patent barriers to research, and states that it will generally not seek patent protection on research tools "where the sole benefit of such protection is financial returns."[196] It also tries to ensure that licensees will continue to make technologies developed in partnership with the TB Alliance available to other entities conducting tuberculosis (TB) research.[197]

Universities have been active partners in such initiatives. In 2003, Yale University and the University of Washington granted OneWorld Health, a nonprofit drug company, an exclusive license to a novel class of high potency compounds, potentially effective against parasitic diseases common in the developing world.[198] The license allows OneWorld Health to develop the compounds for use against neglected diseases, while Yale and the University of Washington are free to pursue "a pharmaceutical partner to develop the same compounds for fungal infections in industrialized countries."[199] Early in 2004, the University of California at Santa Barbara donated to OneWorld Health "the patent rights to [a] class of cardiovascu-

---

private partnerships," and that these mechanisms are less likely to be effective for the most neglected diseases).

193. *See* Barton, *supra* note 51, at 151 ("These efforts involve public or donor funds and often work in cooperation with the private sector.").

194. This approach is taken by the TB Alliance and IAVI. *See* CTR. FOR MGMT. OF IP IN HEALTH R&D, MIHR: HANDBOOK OF BEST PRACTICES FOR MANAGEMENT OF INTELLECTUAL PROPERTY IN HEALTH RESEARCH AND DEVELOPMENT app.D, at 193-94, 198 (Richard Mahoney ed., 2003). The TB Alliance gives no details about the precise conditions it negotiates. IAVI's policy is to negotiate reasonable pricing requirements for sales to the public sector in LMI countries as defined by the World Bank. *Id.* at 198.

195. This is the approach taken by the International Vaccine Institute. *See id.* at 194-95.

196. *Id.* at 193.

197. *Id.* at 194.

198. Press Release, Yale University, Institute for OneWorld Health Licenses Potent Therapy from Yale and University of Washington To Treat Chagas, One of the Largest Paristic Diseases in the World (July 8, 2003), http://www.yale.edu/opa/newsr/03-07-08-01.all.html.

199. *Id.*

lar medicines [for] their novel use as a potential treatment for schistosomiasis, a parasitic scourge that kills more than 200,000 people a year."[200]

Universities have also become more proactive about reserving rights in licensed technologies for their own research purposes and sometimes also for other academic institutions.[201] The NIH has strongly encouraged them in this direction.[202] The most common approach appears to be to reserve rights only for the licensed technology or materials, and only for noncommercial research.[203] However, two more ambitious examples exist. Stanford University's model exclusive licensing agreement reserves rights for commercial as well as noncommercial research for both itself and other universities.[204] The Wisconsin Alumni Research Foundation (WARF), the holder of the University of Wisconsin's stem cell patents, reserves rights only for noncommercial research, but it captures improvements into the scheme and also retains the right to sublicense to governmental agencies and nonprofit research institutions.[205]

Such individualized initiatives are limited and must each bear the cost of negotiating around barriers to research on their own. Collaborative responses hold more promise because they can pool resources and further

200. Associated Press, *UC Santa Barbara Patent Gift To Aid Parasite Fight*, SAN JOSE MERCURY NEWS, Feb. 24, 2004, *at* http://www.mercurynews.com/mld/mercurynews/news/local/8031289.htm (free subscription site).

201. *See, e.g.*, University of Chicago, Guidelines for Grant and Contract Management, *at* http://researchadmin.uchicago.edu/guidelines/300/312.shtml (last visited Apr. 10, 2005) (noting that the university will "make every effort to reserve rights to the [exclusively] licensed material to the University and other non-profit institutions").

202. *See, e.g.*, Best Practices for the Licensing of Genomic Inventions, 69 Fed. Reg. 67,747, 67,748 (proposed Nov. 19, 2004) (noting that the Public Health Service "believes that it is important for funding recipients and the intramural technology transfer community to reserve in their license agreements the right to use the licensed technologies for their own research and educational uses, and to allow other non-profit institutions to do the same"); *see also* NIH Office of Technology Transfer, Model PHS Patent License Agreement—Exclusive, *at* http://ott.od.nih.gov/pdfs/Exclusiv.pdf (last visited May 5, 2005).

203. *See, e.g.*, Baylor Coll. of Med., Exclusive License Agreement (Therapeutic) § 2.2, *at* http://research.bcm.tmc.edu/BLG/bcmt-models.html (last visited Apr. 10, 2005); Univ. of Iowa, Model License Agreement § 2.3(b), *at* http://research.uiowa.edu/techtransfer/forms/model.pdf (last visited Apr. 10, 2005).

204. Stanford Univ., Exclusive Agreement 2, § 3.4, http://otl.stanford.edu/industry/resources/exclusive.pdf (last visited Mar. 28, 2005) ("Stanford retains the right, on behalf of itself and all other nonprofit academic research institutions, to practice the Licensed Patent and use Technology for any purpose, including sponsored research and collaborations.").

205. *See* Wisconsin Alumni Research Found., Standard Non-Exclusive License Agreement 1, § 2B(i), *at* http://www.warf.ws/uploads/media/20031002132027680_Std_non_exclusive_license_agrmt.pdf (last visited Apr. 28, 2005).

reduce transaction costs. The recently-created Public Intellectual Property Resource for Agriculture (PIPRA) illustrates the promise of this approach.[206] Faced with substantial fragmentation as well as exclusive public-to-private licensing of IP rights, several public sector agricultural research institutions founded PIPRA to improve management of IP resources.[207] PIPRA's members include more than twenty major academic research institutions who have committed themselves to collaboratively facilitate the development and dissemination of crops for developing countries. These institutions are exploring several possible approaches, including a standard research exemption that would preserve their rights to issue licenses for research and distribution of products in developing countries,[208] and a public-sector database that would assist scientists in obtaining information about patent landscapes.[209]

These initiatives are first steps along the path of commons-based production, and as such, many of them only partly protect the interests of patients and researchers. But they show that public sector institutions are willing to explore such initiatives, even though the institutions have yet to agree on a strategy that maximizes their collective potential. Public sector institutions—universities chief among them—can implement the solutions outlined here, if these implementations are informed by a careful assessment of the barriers that IPRs can pose for access to medicines and research. The remainder of this Article applies these lessons to create a collective, standardized strategy that universities and other public sector institutions can adopt to govern innovations that have public health applications in LMI countries.

---

206. One additional example that bears mention is a call for an open source, sharing-based model to promote the development of medicines for neglected diseases. *See* Stephen M. Maurer et al., *Finding Cures for Tropical Diseases: Is Open Source an Answer?*, 1 PLOS MED. 183, 183 (2004). Called the "Tropical Diseases Initiative," it aims to capitalize on the convergence between computation and computational biology by creating a "decentralized, Web-based, community-wide effort" where public and private sector scientists would work together to enhance the research base for specific neglected diseases (for example, by searching for protein targets or molecules that would work against known targets). *Id.* at 183-84. Central to the initative's efforts would be a commitment to some type of open licensing scheme. *Id.* at 183.

207. Pub. Intellectual Prop. Res. for Agric., Background, *at* http://www.pipra.org/background.htm (last visited Mar. 12, 2005).

208. Pub. Intellectual Prop. Res. for Agric., Draft Definition of Humanitarian Use (2005) (on file with authors).

209. *See* Pub. Intellectual Prop. Res. for Agric., Activities, *at* http://www.pipra.org/activities.htm (last visited Mar. 28, 2005).

IV. THE CASE FOR UNIVERSITY ACTION

Although universities played a significant role in the various commons-based initiatives discussed in Part III, they have yet to consolidate their efforts in the biomedical domain. This may be poised to change. First, universities play an important role in the biomedical R&D system in the United States. This gives them the power to act to improve the lives of patients and also to collectively persuade private sector partners of the need for an open licensing approach. Second, key members of university communities, from researchers dedicated to open science to students and faculty committed to social justice, will likely support and even demand such a change. Third, there is no significant economic risk associated with the shift—to the contrary, it has the potential to increase the resources available to universities.

A. The Role of Universities in Biomedical Research

Universities are responsible for more than half of the basic research in the United States.[210] Their relative importance to the R&D system is significant and growing: they conducted 14.5% of all R&D activity (both basic and applied) in 1997, nearly double the proportion they conducted in 1960.[211] The majority of all academic research is still funded by the federal government,[212] and although the importance of private sector funding is growing, it does not provide even as much financial support as academic institutions themselves.[213]

Although universities specialize in basic, upstream research intended to advance scientific understanding and to develop the tools of the research field,[214] "it is a fallacy to think of U.S. university research as tradi-

---

210. *E.g.*, NAT'L SCI. BD., SCIENCE AND ENGINEERING INDICATORS 2004, at 5-5, 5-8; *cf.* Francis Narin et al., *The Increasing Linkage Between U.S. Technology and Public Science*, 26 RES. POL'Y 317, 328 (1997) (showing that 73.3% of all the papers cited in U.S. industry patents from 1993-94 were from public science).

211. 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, 101 (2001).

212. In 2001, the federal government provided fifty-nine percent of all academic research funds. *See* NAT'L SCI. BD., *supra* note 210, at 5-5.

213. Industry funded 6.8% of such research in 2001, up from only 2.8% in 1972. *Id.* Academic institutions themselves accounted for another 20% in 2001, nearly doubling their share of total R&D support since the early 1970s. *Id.*

214. *See, e.g.*, Wesley M. Cohen et al., *Links and Impacts: The Influence of Public Research on Industrial R&D*, 48 MGMT. SCI. 1 (2002) (describing the traditional view of public research); Jerry G. Thursby et al., *Objectives, Characteristics and Outcomes of University Licensing: A Survey of Major U.S. Universities*, 26 J. TECH. TRANSFER 59 (2001).

tionally 'basic' and conducted with no attention to practical objectives."[215] In fact, university researchers frequently create new advances in areas as diverse as medical devices, computer software, and scientific instrumentation.[216]

A host of economic studies have confirmed that public sector research, including research done at universities, is a central contributor to R&D in some industries, and particularly in pharmaceuticals.[217] In a survey published in 1991, companies in the drug industry reported that seventeen percent of their products and eight percent of their processes were very substantially influenced by academic research, and that in the absence of academic research, twenty-seven percent of their new products and twenty-nine percent of new processes would have been substantially delayed.[218] Similar results have been confirmed by other studies.[219]

Universities have long engaged in commercialization and patenting, but the scope and nature of these activities has changed profoundly over

---

215. *See* Mowery et al., *supra* note 211, at 101.

216. *Id.*; *see also* Cohen et al., *supra* note 214, at 1-2 (rejecting the "linear model" that casts universities mainly as incubators of basic research and concluding from a survey of R&D managers from many industries, including the pharmaceutical industry, that government and university research is "used at least as frequently to address existing problems and needs as to suggest new research efforts").

217. *See, e.g.*, Cohen et al., *supra* note 214, at 1, 8-10 (concluding that university and government research labs have a "substantial impact on industrial R&D in a few industries, particularly pharmaceuticals"); Alvin K. Klevorick et al., *On the Sources and Significance of Interindustry Differences in Technological Opportunities*, 24 RES. POL'Y 185, 197 (1995) (noting that "almost all the industries that value the contribution of the biological sciences generically . . . also value university-based contributions in that field"); *see also* Adam B. Jaffe, *Real Effects of Academic Research*, 79 AM. ECON. REV. 957, 967 (1989) (discussing geographic spillover effects from universities to industry, which are "statistically strongest in [d]rugs").

218. *See* Edwin Mansfield, *Academic Research and Industrial Innovation*, 20 RES. POL'Y 1, 2-3 & tbl.1 (1991).

219. *See* IAIN COCKBURN & REBECCA HENDERSON, PUBLIC-PRIVATE INTERACTION AND THE PRODUCTIVITY OF PHARMACEUTICAL RESEARCH 5 (Nat'l Bureau Econ. Research, Working Paper No. 6018, 1997) (noting that without the contribution of universities and other public sector research institutions approximately sixty percent of thirty-two innovative medicines studied "'would not have been discovered or would have had their discoveries markedly delayed'" (quoting ROBERT A. MAXWELL & SHOHREH B. ECKHARDT, DRUG DISCOVERY: A CASE BOOK AND ANALYSIS (1990)); *see also* SENATE JOINT ECONOMIC COMM., THE BENEFITS OF MEDICAL RESEARCH AND THE ROLE OF THE NIH 27 (2000) (reporting that public research funding was instrumental in developing fifteen of the twenty-one drugs considered by experts to have had the highest therapeutic impact on society).

the last twenty-five years.[220] The number of U.S. patents granted annually to U.S. academic institutions grew more than ten-fold between 1970 and 2001.[221] In fact, from 1993 to 2003, the number of patents issued to respondents of a survey of leading research universities more than doubled.[222] Licenses have increased concomitantly: American universities, hospitals, and other nonprofit research centers concluded more than 4,500 license and option agreements in 2003, more than double the license and option agreements executed in 1993.[223] A major share of these university patents are in the biomedical field.[224]

Accompanying this growth in patenting and licensing, the number of university technology transfer offices (TTOs) has increased dramatically.[225] TTOs identify, protect, market, and license university IP for

220. *See* Daniel J. Kevles, *Principles, Property Rights, and Profits: Historical Reflections on University/Industry Tensions*, 8 ACCOUNTABILITY RES. 293, 293 (2001).

221. NAT'L SCI. BD., *supra* note 210, at 5-6; *see* Mowrey et al., *supra* note 211, at 104 tbl.1. The rate of increase in the number of utility patents issued to universities is much faster than the overall rate of growth of patenting during the period; during the same time period the number of utility patents issued to U.S. applicants by the PTO did not even double. *Compare* Mowery et al., *supra* note 211, at 104 tbl.1 (citing number of utility patents issued to universities from 1969-97), *with* U.S. Patent & Trademark Office, U.S. Patent Statistics, Calendar Years 1963-2003, http://www.uspto.gov/web/offices/ac/Fido/oeip/taf/us_stat.pdf. Moreover, this growth in university patenting far outpaced the growth in university R&D spending. From 1975 to 1990, "universities increased their patenting per R&D dollar during a period in which overall patenting per R&D dollar was declining." Mowery et al., *supra* note 211, at 104.

222. *Compare* ASS'N OF UNIV. TECH. MANAGERS, AUTM LICENSING SURVEY: FY 2002, SURVEY SUMMARY, at 1 tbl.S-6 (2003) [hereinafter 2002 AUTM SURVEY] (showing 1603 patents issued to responding universities in 1993), *with* ASS'N OF UNIV. TECH. MANAGERS, AUTM LICENSING SURVEY: FY 2003, INTERIM REPORT, at 1 (2004) [hereinafter 2003 AUTM INTERIM REPORT] (showing 3450 patents issued to responding universities in 2003). Note that AUTM surveys only report data from their membership, which does not include all U.S. universities, and that the membership and number of respondents have grown over time, meaning that responses between years are not strictly comparable. 2002 AUTM SURVEY, *supra*, at 6 tbls.S-1, S-2. The 1993 data included responses from eighty-five percent of the top U.S. research universities by research funding, and the 2002 data included responses from ninety-four percent of the same group. *Id.* at 6 tbl.S-1.

223. *Compare* 2002 AUTM SURVEY, *supra* note 222 (showing 2227 licenses and options executed in 1993), *with* 2003 AUTM INTERIM REPORT, *supra* note 222 (showing 4955 licenses and options executed in 2003).

224. Rai & Eisenberg, *supra* note 147, at 292.

225. Although the number of institutions starting TTOs increased throughout the 1970s, that number grew dramatically in the 1980s in the wake of Bayh-Dole. *See* 2002 AUTM SURVEY, *supra* note 222, at 7 figs.1 & 2.

commercial use. Thus, TTOs are the key institutional player in universities' increasingly focused and proactive approach to securing IPRs.[226]

Turning to current practices, universities frequently patent the research tools they develop, and have been criticized for licensing some very important tools exclusively.[227] Universities have also come under fire for what many perceive as overly aggressive terms in research tools licenses. Typically, when licensing to private firms, universities seek fees or reach-through royalties on resulting products.[228] Furthermore, "[e]ven when they do not seek patents, universities often seek to preserve their expectations for profitable payoffs by imposing restrictions on the dissemination of research materials and reagents that might generate commercial value in subsequent research."[229] A review conducted by the NIH concluded that universities have sought just about every kind of clause in research tool licenses to which they themselves have objected, including publication restrictions, rights in or the option to license future discoveries, and prohibition on transfer to other institutions or scientists.[230] As noted above, it appears that universities are beginning to reserve research rights for themselves and other academic institutions when they issue exclusive licenses, but the reach of these clauses is often limited to noncommercial research.[231]

Where an invention has potential to be developed into a pharmaceutical product, universities will typically patent it in the United States,

---

226. Typically, university scientists are required to report any potentially important innovation—for example, a new molecular entity with pharmacological significance—to their TTO, which then evaluates the invention to determine whether it has commercial potential. *See, e.g.*, Office of Intellectual Prop., Mich. State Univ., Handling Your Invention (2001) ("Researchers are obligated to report any inventions to the [TTO] . . . ."), http://www.msu.edu/unit/oip/handling.html; *see also* Lawrence Berkeley Nat'l Lab., How the Tech Transfer Process Works, http://www.lbl.gov/Tech-Transfer/researchers/how-tt-works.html (last visited Mar. 29, 2005) (describing the invention evaluation process and encouraging researchers to contact the TTO to discuss any research that may have produced an invention).

227. *See* Rai & Eisenberg, *supra* note 147, at 293, 301, 309 (discussing the patenting and exclusive licensing of the University of Wisconsin's stem cell patents); *id.* at 302 (discussing the patenting and exclusive licensing of an important cell signaling pathway by Harvard, MIT, and the Whitehead Institute for Biomedical Research); *supra* text accompanying notes 104-05. Professor Rai and Professor Eisenberg note that many university patents cover research tools and that "one recent study of Columbia University's patent portfolio indicates that more than 50% of its licensed patents represent research tools." *See* Rai & Eisenberg, *supra* note 147, at 292.

228. *See* Rai & Eisenberg, *supra* note 147, at 294.

229. *Id.* at 291.

230. *See* NAT'L INSTS. OF HEALTH, *supra* note 103.

231. *See supra* text accompanying note 203.

Europe, Canada, Australia, and Japan.[232] There is no comprehensive data on the overall ownership position of universities in current pharmaceutical technologies, but the aforementioned trends in R&D and patenting suggest their ownership share is both substantial and increasing. In recent years universities have obtained U.S. patent rights in a number of key pharmaceutical products, including: the cancer drugs cisplatin and carboplatin,[233] pemetrexed (Alimta),[234] and cetuximab (Erbitux);[235] the anemia treatment epoetin alfa (Epogen);[236] the AIDS drugs stavudine (Zerit),[237] 3TC (Epivir),[238] abacavir (Ziagen),[239] and T20 (Fuzeon);[240] and the best-selling glaucoma medicine latanoprost (Xalatan).[241] Universities also hold patents on essential manufacturing processes.[242]

A TTO will also decide in which foreign countries it wishes to patent.[243] This decision appears to be based on a narrow economic calcula-

---

232. *See* YALE UNIV. CTR. FOR INTERDISCIPLINARY RESEARCH ON AIDS, ACCESS TO ESSENTIAL MEDICINES AND UNIVERSITY RESEARCH: BUILDING BEST PRACTICES 4 (2003) [hereinafter WORKSHOP REPORT].

233. Michigan State University held the IPR. *See* Eyal Press & Jennifer Washburn, *The Kept University*, ATL. MONTHLY, Mar. 2000, at 39.

234. Princeton University holds the IPR. U.S. Patent No. 5,344,932 (issued Sept. 6, 1994).

235. University of California at San Diego holds the IPR. U.S. Patent No. 4,943,533 (issued July 24, 1990).

236. Columbia University holds the IPR. *See* DEP'T OF HEALTH & HUMAN SERVS., NAT'L INSTS. OF HEALTH, A PLAN TO ENSURE TAXPAYERS' INTERESTS ARE PROTECTED (2001) [hereinafter DHHS/NIH], http://www.nih.gov/news/070101wyden.htm.

237. Yale University holds the IPR. U.S. Patent No. 4,978,655 (issued Dec. 18, 1990).

238. Emory University holds the IPR. *See* EMORY UNIV., OFFICE OF TECH. TRANSFER, PRODUCT PIPELINE 5 (2004), http://www.ott.emory.edu/shared_web/technologies/Emory_Pipeline.pdf.

239. University of Minnesota holds the IPR. *See* Univ. of Minnesota, Fact Sheet on Glaxo-Wellcome AIDS Discovery Settlement (Oct. 5, 1999), http://www.umn.edu/urelate/newsservice/newsreleases/99_10glaxofacts.html.

240. Duke University holds the IPRs. U.S. Patent No. 5,464,933 (issued Nov. 7, 1995); U.S. Patent No. 6,133,418 (issued Oct. 17, 2000).

241. Columbia University holds the IPR. U.S. Patent No. 4,599,353 (issued July 8, 1986); *see also* Jeff Gerth & Sheryl Gay Stolberg, *Medicine Merchants: Birth of a Blockbuster; Drug Makers Reap Profits on Tax-Backed Research*, N.Y. TIMES, Apr. 23, 2000, at 1-1.

242. For example, Columbia University's co-transformation patent is used to manufacture biotech drugs and has made the university nearly $100 million annually during the patent's life. *See* Bernard Wysocki Jr., *College Try: Columbia's Pursuit of Patent Riches Angers Companies*, WALL ST. J., Dec. 21, 2004, at A1. Florida State University also holds a key patent on the process to make the cancer drug Taxol. *See* DHHS/NIH, *supra* note 236.

243. *See, e.g.*, Lawrence Berkeley Nat'l Lab., *supra* note 226.

tion, measuring the expected net present value of exclusivity against the cost of obtaining and defending a patent, and without factoring in non-economic considerations, such as access to medicines for LMI country residents.[244] While there is no comprehensive data and no easy way to determine patent status in the majority of LMI countries,[245] universities report that few of their inventions are patented in LMI countries because the benefits of exclusivity rarely justify the cost of securing patents.[246] However, the more likely a technology is to have application in a developing country, the more likely it is the economics will weigh in favor of patenting. The calculus shifts further in favor of patenting if a private sector licensee is willing to bear the associated costs. In that case, universities typically permit licensees to decide where to patent—although most universities retain the patents in their own name.[247]

If the innovation is intended as a pharmaceutical or diagnostic end product, it will typically be licensed under a worldwide exclusive license,[248] often to a small start-up company,[249] which will usually develop the product further before sublicensing it to larger firms.[250] In exchange for exclusivity, the university will typically receive royalty payments

---

244. *See* WORKSHOP REPORT, *supra* note 232, at 4.

245. Few LMI countries have online patent databases, making it difficult to obtain reliable information on such patents except by directly asking those who may hold them or by consulting local patent offices.

246. *See* WORKSHOP REPORT, *supra* note 232, at 4; *cf.* Rebecca S. Eisenberg, *Public Research and Private Development: Patents and Technology Transfer in Government-Sponsored Research*, 82 VA. L. REV. 1663, 1666 (1996) ("If anyone sees money to be made through patenting a government-sponsored research discovery, chances are it will be patented . . . [, but] resource constraints prohibit patenting many discoveries that emerge from government-sponsored research.").

247. *See* Lita Nelsen, *The Role of University Technology Transfer Operations in Assuring Access to Medicines and Vaccines in Developing Countries*, 3 YALE J. HEALTH POL'Y L. & ETHICS 301, 304 (2003).

248. *See, e.g.*, OFFICE OF COOPERATIVE RESEARCH, YALE UNIV., FROM BENCH TO BEDSIDE: 1996-1998, at 5 (1999), *available at* http://www.yale.edu/ocr/images/docs/ocr_report_96-98.pdf; Jeannette Colyvas et al., *How Do University Inventions Get Into Practice?*, 48 MGMT. SCI. 61, 67 (2002) (finding, in an empirical study of university technology transfer, that "the ability to issue exclusive licenses is most important in the context of embryonic inventions," such as early-stage potential drug compounds); Nelsen, *supra* note 247, at 303.

249 . Two-thirds of university licensing agreements are made with "newly formed or existing small companies." ASS'N OF UNIV. TECH. MANAGERS, AUTM LICENSING SURVEY: FY 2000, Executive Summary, at 1 (2001) [hereinafter 2000 AUTM SURVEY]; *see* 2002 AUTM SURVEY, *supra* note 222, at 1.

250. *Cf. Big Trouble for Big Pharma*, ECONOMIST (London), Dec. 4, 2003.

and/or equity in the licensee.[251] Due diligence clauses are also commonplace, to ensure that the university technology does not lay fallow. These clauses oblige the licensee to develop the compound—to conduct clinical trials and other developments necessary to market the product—or face revocation of the license.[252]

Because universities license their technologies in order to secure the investment and expertise necessary to further develop and market the technologies, the university's licensed patent will frequently be a key component, but not the entirety, of the rights necessary to generate the end product. The licensee may acquire secondary or improvement patents on subsequent developments such as dosages or delivery systems.[253] The licensee will also generate the safety and efficacy data needed to market the drug, and will be able to exercise exclusive rights over this data.[254]

B. Institutional Principles and the Internal Political Economy of Universities

1. Open Science and the Goals of Technology Transfer

Universities' core institutional principles include the production and dissemination of knowledge, as well as a related and more general dedica-

---

251. *E.g.*, Bruce Berman, *From Tech Transfer to Joint Ventures—Part I*, PATENTCAFE, Mar. 6, 2002, http://2xfr.patentcafe.com/article.asp?id=555; *see also* Maryann Feldman et al., *Equity and the Technology Transfer Strategies of American Research Universities*, 48 MGMT. SCI. 105 (2002) (describing the growth of university equity holdings).

252. *E.g.*, Indiana Univ. Research & Tech. Corp., Inventors & Creators—What IURTC Negotiates for in Licensing Agreements, http://iurtc.iu.edu/tt_marketing-terms.html#diligence (last visited Mar. 31, 2005).

253. *See* WORKSHOP REPORT, *supra* note 232, at 3; Nelsen, *supra* note 247, at 305. While it is not clear how often such improvement patents are filed in LMI countries, they sometimes are. For example, in Thailand, BMS obtained a patent on the pill form of ddI combined with an antacid buffer, although the underlying compound was not under patent in Thailand. *See* Tina Rosenberg, *Look at Brazil*, N.Y. TIMES, Jan. 28, 2001, § 6 (Magazine), at 26. It also seems that such patents will be frequently sought by companies in important source countries like India. There are reportedly over 7000 pharmaceutical patents in India's "mailbox." *See* KG Narendranath, *Patent Mailbox Opens, Pfizer Is Top Applicant*, FIN. EXPRESS (India), Mar. 21, 2005 (noting that the vast majority of the patent applications in India's mailbox belong to foreign filers), *at* http://www.financialexpress.com/fe_full_story.php?content_id=85782. These applications almost certainly include thousands of patents on combinations, formulations, dosages, and other minor improvements. (According to TRIPS, member countries that did not offer patent protection for pharmaceutical and agricultural chemical patents on the date that the Agreement entered into force had to provide patent holders with a means to file such applications, commonly referred to as a "mailbox." TRIPS Agreement, *supra* note 47, art. 60.8(i).)

254. *See, e.g.*, *supra* note 46.

tion to improving human welfare. The centuries-old academic tradition of open scientific practice[255] faces increasing pressure from universities' patenting and commercialization activities. As a result, these policies, and the TTOs that administer them, have come under attack.[256] Because a TTO's performance is generally measured by licensing revenue,[257] TTO professionals have incentives to aggressively seek patents and high licensing fees, which the research community as a whole might rather forego.[258]

There is, however, nothing inherent in the existence of university patents or of TTOs that requires that this be the case. TTO incentives would change if contributions to health, particularly global health, were made a part of the calculus. As access concerns have come to the forefront, leading members of the technology transfer community have shown signs of supporting steps to address health concerns of the developing world.[259] Universities, after all, are different kinds of organizations than pharmaceutical firms: they have different revenue structures, different R&D invest-

255. *See* Nannerl O. Keohane, *The Mission of the Research University*, DAEDALUS, Fall 1993, at 101, 122 ("Proprietary knowledge . . . is in principle antithetical to the openness in sharing knowledge that is at the heart of the university's mission."); *cf.* ROBERT KING MERTON, THE SOCIOLOGY OF SCIENCE: THEORETICAL AND EMPIRICAL INVESTIGATIONS (1973) (describing the "normative structure of science" and finding that one of the institutional mores is collaboration).

256. Rai & Eisenberg, *supra* note 147, at 305.

257. *See, e.g.*, Eisenberg, *supra* note 246, at 1710 (discussing universities' view of royalties as a measure of TTO success); Thursby et al., *supra* note 214, at 65-66 (reporting that surveyed TTOs list generating royalties and license fees as the most important measure of TTO success, followed by the number of licenses or options signed).

258. *See* Eisenberg, *supra* note 246, at 1710, 1714-15.

259. Both individual TTO directors and their national organization have demonstrated interest in these issues. For example, the Director of MIT's Office of Technology Licensing has authored an article encouraging the technology transfer community to learn about its power to promote access to medicines in developing countries and outlining possible strategies in this area. Nelsen, *supra* note 247, at 303-04; *see* Jon Soderstrom, Managing Director Office of Cooperative Research Yale University, The Future of University Technology Transfer: Where Do We Go from Here, Presentation to the Gordon Research Conference on Global Aspects of Technology Transfer (Sept. 21, 2003) (encouraging universities to promote technology transfer to benefit developing countries). In December 2003, the Association of University Technology Managers (AUTM) formed a group to examine global health issues. *New AUTM Special Interest Group Announced: Technology Transfer Professionals for Global Health*, AUTM NEWSLETTER (Ass'n of Univ. Tech. Managers, Northbrook, Ill.), Nov./Dec. 2003, at 9. The 2003 and 2004 AUTM annual meetings included several global-health-related poster presentations and workshops. *See* ASS'N OF UNIV. TECH. MANAGERS, AUTM 2004 ANNUAL MEETING [hereinafter AUTM 2004 MEETING], http://www.autm.net/events/eventFiles/AUTM04FP.pdf; Nelsen, *supra* note 247, at 303.

ment motivations, and different cultural self-perceptions.[260] To the extent universities have managed their patent portfolios as though the universities were for-profit firms, it is a result of a failure to properly define the universities' interests and power as holders of significant patent stakes. A narrow focus on maximizing the amount of revenue generated by university discoveries is difficult to reconcile with the spirit of university patent and licensing policies, which typically declare that the ultimate purpose of technology includes the advancement of the public good.[261]

As Yale's experience with stavudine demonstrates, the conflict between a university's ethos and its patenting practices can erupt into public protests from both students and faculty: students organized to support MSF's request for a patent concession, and one of the scientists who had discovered that stavudine could be used to treat HIV voiced his disapproval of the university's practices in the *New York Times*.[262] A student-led group, Universities Allied for Essential Medicines (UAEM) continues to challenge closed licensing practices at universities,[263] while the academic community is giving increased attention to the access and R&D gaps. The campaign for divestiture from South Africa, and the more recent anti-sweatshop movement targeted at university apparel,[264] demonstrate that student-driven protests can produce changes in university policy. For institutions dependent on philanthropy and government funding, the goodwill gained by acting to alleviate the access gap is potentially signifi-

---

260. *Cf.* Goldie Blumenstyk, *A Contrarian Approach to Technology Transfer*, CHRON. HIGHER EDUC., Mar. 12, 2004, at 27. *But cf.* DEREK BOK, UNIVERSITIES IN THE MARKETPLACE: THE COMMERCIALIZATION OF HIGHER EDUCATION (2003) (lamenting the increasing commercialization of universities).

261. *See, e.g.*, Office of Tech. Licensing, Stanford Univ., OTL and the Inventor: Roles in Technology Transfer, *at* http://otl.stanford.edu/inventors/resources/otlandinvent.html (last updated Aug. 8, 2003) ("OTL is responsible for managing the intellectual property assets of the University for the public good."); Office of Tech. Transfer, Univ. of Cal., University of California Patent Policy, http://www.ucop.edu/ott/patentpolicy/patentpo.html#pol (Oct. 1, 1997) ("The following University of California Patent Policy is adopted to encourage the practical application of University research for the broad public benefit . . . ."); Tech. Licensing Office, MIT, Mission Statement, http://web.mit.edu/tlo/www/mission.html (last visited Mar. 31, 2005) ("[Our] mission . . . is to benefit the public by moving results of M.I.T. research into societal use via technology licensing . . . .").

262. Donald G. McNeil Jr., *Yale Pressed To Help Cut Drug Costs in Africa*, N.Y. TIMES, Mar. 12, 2001, at A3.

263. Univs. Allied for Essential Meds., *at* http://www.essentialmedicine.org (last visited Apr. 1, 2005).

264. *See, e.g.*, Peter Dreier & Richard Appelbaum, *The Campus Anti-Sweatshop Movement*, AM. PROSPECT, Sept.-Oct. 1999, at 71, http://www.prospect.org/print/V10/46/dreier-p.html.

cant, while the negative publicity from exposure of internal fissures over this issue may be damaging. In addition, to the extent the behavior of their TTOs deviates from the public interest, universities may face other, potentially less-expected risks.[265]

Finally, it is worth noting that the NIH will unlikely step in to resolve these tensions. Although the Bayh-Dole Act gives the agency the authority to "march in" on patents to ensure that federally-funded inventions are accessible to the public,[266] the NIH has so far rejected every request that it use these powers to make medicines more accessible.[267] The NIH also "has no authority under the Bayh-Dole Act to issue broadly applicable substantive regulations concerning the licensing of inventions (as distinguished from making specific determinations regarding march-in rights in the context of particular grants)."[268] This substantially limits the NIH's ability to deal with barriers caused by patents on federally-funded research

---

265. *See, e.g.*, Peter D. Blumberg, Comment, *From "Publish or Perish" to "Profit or Perish": Revenues from University Technology Transfer and the 501(c)(3) Tax Exemption*, 145 U. PA. L. REV. 89 (1996) (arguing that university income from technology transfer should be subject to the unrelated business income tax to the extent TTO practices stray from universities' educational and scientific mission, such as when TTOs license a technology exclusively). The result in the *Madey* decision also turned in part on the perception that universities are fundamentally analogous to businesses, although formally they are structured as nonprofits. *See* Madey v. Duke Univ., 307 F.3d 1351 (Fed. Cir. 2002), *cert. denied*, 539 U.S. 958 (2003).

266. Bayh-Dole reserves for the government a nonexclusive, nontransferable, irrevocable, paid-up license to practice federally-funded inventions, 35 U.S.C. § 202(c)(4) (2000), and the right to force patentees to license government-funded patented inventions to third parties on reasonable terms to ensure the invention's public availability, *id.* § 203. The government may require compulsory licensing if there is inappropriate delay in achieving practical application of the invention, *id.* § 203(1)(a), which includes making the invention available to the public on reasonable terms, *id.* § 201(f), or if licensing is necessary to alleviate health or safety needs, *id.* § 203(1)(b). The NIH also seems to construe these provisions to apply to patents held in other countries. *See generally* Letter from Harold Varmus, Director, NIH, to Ralph Nader, James Love, and Robert Weissman (Oct. 19, 1999) (discussing but declining to exercise the federal government's power to license patent rights, including foreign patent rights, to the WHO), *available at* http://www.cptech.org/ip/health/sa/varmusletteroct19.html.

267. *See* Barbara M. McGarey & Annette C. Levey, *Patents, Products, and Public Health: An Analysis of the CellPro March-In Petition*, 14 BERKELEY TECH. L.J. 1095 (1999); Nat'l Insts. of Health, Office of the Director, Opinion in the Case of Norvir (Jul. 29, 2004), *available at* http://ott.od.nih.gov/Reports/March-In-Norvir.pdf; Nat'l Insts. of Health, Office of the Director, Opinion in the Case of Xalatan (Sept. 17, 2004), *available at* http://ott.od.nih.gov/Reports/March-In-Xalatan.pdf. Scholars have criticized the NIH's inaction and recommended substantial reforms to the Bayh-Dole regime. *See* McGarey & Levey, *supra*; Rai & Eisenberg, *supra* note 147, at 303-04, 310-13.

268. Rai & Eisenberg, *supra* note 147, at 308.

tools.[269] In other words, if change in university practice is to occur, it will likely have to be initiated by universities themselves.

2. The Economics of Technology Transfer and the Access Gap

Fortunately, adopting the proposals made in this Article is financially viable for universities. To begin, universities do not rely substantially on technology transfer revenues. Although TTOs have managed to obtain tens of thousands of patents, they tend to remain money-losing endeavors.[270] The number of schools that make money from technology transfer is small, and those that profit tend to do so from a limited number of highly successful patents.[271] Licensing revenues are typically equivalent to just four percent of a university's research funds, and this figure decreases significantly when the costs of patent and license management, as well as the inventors' share of royalty income, are subtracted.[272] When patent royalties are compared to total university revenue, they appear quite small, constituting only 0.5 to 2% of revenues, even for the subset of universities that are patent-productive.[273]

Most significantly for our purposes, the proportion of revenue that a university would obtain from developing countries, even on a blockbuster drug, will be vanishingly small—only a few percent of those few percent of total revenues that PhRMA companies make in LMI countries.[274] Yale reported no lost revenue as a result of the stavudine patent concession in

---

269. *Id.* at 309.

270. William Brody, President of Johns Hopkins University, has observed, "[t]he dirty secret is that for many universities—perhaps most—they are not breaking even, much less making money on the proposition." William R. Brody, From Minds to Minefields: Negotiating the Demilitarized Zone Between Industry and Academia, Remarks at Biomedical Engineering Lecture Series (Apr. 6, 1999), *available at* http://www.jhu.edu/~president/speech/biomlec.html; *see also* Bhaven N. Sampat, The Effects of Bayh-Dole on Technology Transfer and the Academic Enterprise: A Survey of the Empirical Literature 12-13 (2004) (unpublished manuscript, on file with authors), http://www.vannevar.gatech.edu/papers/bdsurvey.pdf.

271. *See* 2002 AUTM SURVEY, *supra* note 222, at 19 fig.25 (showing the distribution of total licensing income received by U.S. universities in 2002); Sampat, *supra* note 270, at 11-12 (observing that a small fraction of universities realize significant income from licensing, with the only a few schools owning truly lucrative patents).

272. *See* Nelsen, *supra* note 247, at 302 (giving figures of two percent to four percent) (citing 2002 AUTM SURVEY, *supra* note 222)).

273. *See* Benkler, *supra* note 36, at 1110.

274. *See* Jerry G. Thursby & Marie C. Thursby, University Licensing Under Bayh-Dole: What Are the Issues and Evidence? 4 (May 2003) (unpublished manuscript, on file with authors) ("For all university technologies, an average royalty rate of 2% is common. For pharmaceuticals the maximum rate one typically encounters for university technologies is 5%; however, the rates are usually closer to 1.5%.").

South Africa, and Yale's Dean of Public Health Michael Merson stated that "[t]his change was made at Yale without any negative consequences to the University—financial or otherwise."[275] Universities could even financially benefit from adopting the policies we propose, if at the margins it helped them to attract scientists, students, or funding for research.

Because of the small size of the LMI market, patent-based pharmaceutical firms can promote access at minimal cost, and without sacrificing profits to any substantial degree, simply by allowing generics to enter LMI markets. The same is true of efforts to free up research on neglected diseases or developing country indications for existing medicines; because such companies do not currently seek revenues from such research, allowing others to do it will not affect their profits.

The pharmaceutical industry's increasing dependence on external research, including university research, to fill its R&D pipelines and provide it with research tools[276] further suggests that universities can promote research and access without material risk of losing deals, reducing income, or jeopardizing the viability of technology transfer operations. This is particularly true if universities act collectively and in a standardized fashion rather than trying to promote access on a deal-by-deal basis. While pharmaceutical and biotechnology companies will likely resist any changes to the status quo, if major research institutions act together to implement new practices—and thereby redefine the norms—pharmaceutical and biotech-

---

275. Michael Merson, *Preface* to WORKSHOP REPORT, *supra* note 232, at v. It should be noted, however, that universities often prize even limited income from technology transfer because these funds can in large part be used for any purpose, as opposed to research grants and even donations, which typically may only be used for predefined purposes. *See, e.g.*, Thomas A. Massaro, *Innovation, Technology Transfer, and Patent Policy: The University Contribution*, 82 VA. L. REV. 1729, 1735 (1996). Furthermore, a portion of technology transfer monies flow to inventors (typically thirty to forty percent), *see* 35 U.S.C. § 202(c)(7)(B) (2000) (mandating revenue sharing under Bayh-Dole); Robert L. Barchi, IP and Technology Transfer from the Academic Perspective 19, http://www7.nationalacademies.org/step/Barchi_ppt.ppt (last visited Mar. 31, 2005); Office of Cooperative Research, Yale Univ., Yale Univ. Patent Policy § 4(d) (Feb. 1998), http://www.yale.edu/ocr/invent_policies/patents.html, which provides direct incentives for university researchers. But given the highly uncertain returns on research, the expected value of patent royalties is small, and the loss of royalties from LMI country sales particularly insignificant, even to those who value unrestrained funds and personal financial incentives the most.

276. *Cf. Big Trouble for Big Pharma*, *supra* note 250 ("Many big drug firms have begun to license more of their technology and products from outside companies, especially biotechnology start-ups."); Jerry G. Thursby & Marie C. Thursby, *Who Is Selling the Ivory Tower? Sources of Growth in University Licensing*, 48 MGMT. SCI. 90, 90 (2002) (noting pharmaceutical companies' "increased business reliance on external R&D").

nology companies will have little choice. While an individual university may be dispensable to the pharmaceutical industry, universities in aggregate are not.[277]

V. WHAT UNIVERSITIES CAN DO

If universities are to harness the potential of their technologies to close the access and R&D gaps, they must formulate a strategy that will help them achieve these goals. This strategy must be easy to use, to allow even those universities with the fewest resources to implement it. It must also be highly standardized, to capitalize on universities' collective influence and bargaining power. Below, we elaborate two commons-based approaches that meet these criteria and are configured to serve the needs of people living in developing countries.

A. Addressing the Access Gap: Equitable Access Licensing

Part II demonstrated that the price differential between exclusivity-based pricing and marginal-cost pricing can constitute a serious barrier to access to medicines for people living in the developing world. Here we describe an approach called Equitable Access ("EA") licensing as a means of removing that differential. This mode of licensing, like the licensing practices that govern free software, uses proprietary rights to secure freedom for an open class of potential users, rather than to secure exclusivity for a closed class of licensees. Like the GPL, it uses IPRs not to exclude and monopolize, but rather to ensure the right of third parties to access and distribute the innovation and its derivative products. Finally, the EA license is commons-based because it seeks to use the university's rights to create a self-binding commons—a universe of information resources necessary to produce the end product—that is open for all to use. This free-

277. We expect that successful collective action by universities is likely if the dominant players in the technology transfer field take the lead. More than 3000 institutions received NIH funding in FY 2003; twenty-three of the top twenty-five recipients were universities, and they alone garnered thirty-four percent of the total funding awarded in the United States. *See* Office of Extramural Res., Nat'l Insts. of Health, Award Trends, Rankings: All Institutions, FY 2003; By State, FY 2003, *at* http://grants.nih.gov/grants/award/awardtr.htm#c (last visited Apr. 5, 2005). Of the 112 medical schools earning NIH funding, the top twenty-five received fifty-six percent of the monies. *See* Office of Extramural Res., Nat'l Insts. of Health, Award Trends, Rankings: Medical Schools, FY 2003, *at* http://grants.nih.gov/grants/award/awardtr.htm#c (last visited Apr. 5, 2005). Successful technology transfer deals are similarly concentrated among a small group of elite universities. *See* 2002 AUTM SURVEY, *supra* note 222, at 20 (showing that less than one percent of active licenses generated more than one million dollars in 2002, and that three institutions reported granting ten or more of such licenses).

dom, we predict, will entice other actors to provide the end product at a competitive price.

Simply stated, an Equitable Access License is one that seeks: (1) to ensure freedom to operate for any party that manufactures and distributes the licensed technology and any derivative products in LMI countries, and (2) to minimize administrative overhead and political contingency by initiating a self-enforcing open licensing regime.

1. The Choice and Definition of the "Freedom To Operate" Approach

Theoretically, if a university developed a drug, vaccine, or diagnostic tool from its lab bench to the pharmacy shelf without any partners, it could eliminate supra-marginal cost pricing in developing countries by simply not patenting or seeking other exclusive rights in these territories, and allowing anyone to export the university's development.[278] However, this is not generally how R&D happens. Universities operate in a universe where they are not the only holders of IPRs, and they frequently contribute only at one stage in the value chain. Non-patenting alone will not, therefore, ensure that generics will be available in LMI countries, just as releasing copyrighted works into the public domain will not ensure that derivative works will remain open for anyone to use, modify, and distribute. To resolve this problem, EA clauses must adopt the strategy used by the GPL—they must leverage the exclusive rights associated with a patent to ensure accessibility of derivative products.

An initial choice faces any innovator who wishes to do this: whether to adopt a fair pricing approach or implement a freedom to operate strategy. Under the first option, the licensor would oblige its licensee to distribute the end product in the selected territories quickly, in sufficient quantities, and at the marginal cost of manufacture. Under the second option, what we call the "freedom to operate" approach, the innovator uses open licensing to achieve the goal of marginal cost pricing. Ensuring freedom to operate here means guaranteeing third parties the right to compete in a market without being blocked by patents or other forms of exclusive rights. It does not mean guaranteeing third parties the active transfer of materials or know-how to assist their production of a generic alternative—or what we term "enablement." The choice between these options is essentially a choice between heightened regulation of the licensee's behavior and contractual deregulation of the end-product market.

278. Even if possible, this approach would have disadvantages. Patents can be useful for defensive purposes, for example, because they can give the university a bargaining chip in cross-licensing negotiations. *Cf. supra* note 41.

The fair pricing approach might appear, at first blush, to be the most direct and efficient means of achieving marginal cost pricing. After all, the licensee need not incur the cost of reverse engineering, will have existing production capacity, and may also be able to take advantage of economies of scale. Some licensees may also prefer this option, as it offers them more control than does the freedom to operate approach. But a strategy that relies on freedom to operate will generally produce better results for both patients in LMI countries and for universities, for several reasons.

To begin, the freedom to operate approach is preferable for universities because it does not require them to take an active role in monitoring or enforcement. This approach avoids placing any ongoing demands on universities or their licensees by introducing a third set of players—typically generic companies—with market incentives to narrow the access gap by offering low-priced, but still profitable, products. A university that signs a fair pricing clause, on the other hand, must be willing to monitor the clause's implementation and make a credible threat to bring legal action against a defaulting licensee, or to deem the licensee in breach and revoke the overall license itself.

In principle, the empirical challenges of the monitoring role can be overcome.[279] But monitoring all of these issues would require universities to devote substantial resources to the task, and enforcement would constitute an even more costly endeavor.[280] Universities are not all equally able to invest in monitoring or enforcement of licenses, and even the best-situated universities will have limited resources to devote to such activities. Moreover, the fact that universities are repeat players in a game

279. There are well-established accounting standards to define marginal cost of manufacture, and the university could reserve for itself the right to audit a licensee's books. This would be necessary because patent-based companies have proven generally unwilling to reveal cost in a transparent fashion. *See, e.g.*, Outterson, *supra* note 72, at 222, 253 n.255 ("Given the endemic opacity of all PhRMA data on costs, perhaps the best way to calculate marginal cost is through compulsory licensure. . . . Absent the patent monopoly, generic companies in a competitive environment will certainly price much closer to marginal cost than PhRMA companies."). In addition, it would be relatively easy to monitor whether a company was meeting its deadline for drug registration in particular countries, so long as these deadlines are clearly established. Determining whether a company is meeting all existing market need would be more difficult, possibly requiring an investigation into in-country conditions, but may nonetheless be manageable.

280. A reasonable pricing effort would presumably be plagued with the same kinds of delays and inefficiencies that have adversely affected existing donation and discount programs.

where reputations often travel quickly is likely to discourage aggressive monitoring and enforcement.[281]

Second, the freedom to operate approach can be expected to provide patients in LMI countries with cheaper medicines than the fair pricing approach would. Experience indicates that generic companies will almost always be able to undercut the "at cost" prices of proprietary firms. It is not clear whether proprietary products have higher marginal costs, whether companies calculate marginal cost in different ways, or whether proprietary companies are simply being dishonest when they claim to be selling at cost. Regardless of the reason, it is clear from available evidence that competition has been more reliable as a method of lowering prices than voluntary "at cost" pricing. Again, because patients and even governments in LMI countries are extremely sensitive to even small differentials in price, the freedom to operate approach has a substantial advantage here.

The prevailing legal environment gives the freedom to operate model a third advantage over the fair pricing strategy: it can reduce the risk of both physical and price arbitrage. Differentially priced products sold by the originator company may be susceptible to parallel trade.[282] The freedom to operate approach sidesteps this issue by relying upon generic provisioning to reach marginal cost pricing. Due to patent barriers, generic versions are not susceptible to parallel trade in the same way as originator products may be. Licensees may also express concerns about the generic products illegally finding their way into high-income countries. There is no empirical evidence of any substantial flows of medicines from LMI to rich countries;[283] but insofar as this is a concern, an EA clause can address it in the

281. Universities are, in this regard, differently situated from the single-issue public-private partnerships (PPPs). This fact, and the lack of an articulated alternative, may explain why such PPPs have relied heavily upon reasonable pricing requirements to address access concerns.

282. Parallel trade is a form of arbitrage that puts pressure on companies that seek to price discriminate. *See generally* Outterson, *supra* note 72. While TRIPS allows parallel trade of originator products, many countries (such as the United States) prohibit it—either as a matter of patent exhaustion law or as a result of regulatory barriers. *See id.* at 209-15.

283. *See id.* at 257-60 (discussing two alleged instances of dysfunctional arbitrage and determining that the claims were inappropriate or unsubstantiated); *id.* at 262 ("As of April 2002, both the European Commission and the pharmaceutical companies acknowledged that pharmaceutical arbitrage from poor countries into the high income was 'still largely theoretical.'" (citing DG TRADE, EUROPEAN UNION, TIERED PRICING FOR MEDICINES EXPORTED TO DEVELOPING COUNTRIES, MEASURES TO PREVENT THEIR RE-IMPORTATION INTO THE EC MARKET AND TARIFFS IN DEVELOPING COUNTRIES § 3.3 (EU Working Document, 2002))). For detailed responses to pharmaceutical industry concerns

same manner that the WTO has treated the issue—by requiring use of different packaging, pill color, and pill shape to facilitate identification of illegal importations where this is feasible and does not significantly increase the price of the product.[284] In theory, generic provisioning also ought to assuage some of the licensee's concerns—whether justified or not—about what we might call "politically mediated arbitrage," where discounted prices in one country fuel public demand for lower prices in another.[285]

Finally, the freedom to operate approach is preferable because it will tend to generate a more sustainable and appropriate supply of low cost medications in LMI countries. This approach puts a thumb on the scale of technology transfer by presenting a small—but, for generic companies, meaningful—market to attract the investment necessary to reverse engineer and scale-up production. The long-term health of the generics industry requires a diffusion of technical knowledge and markets sufficient to sustain what is widely acknowledged to be a very low margin business. Encouraging competitive provisioning in LMI countries will foster and sustain the development of diverse nodes of technological capability necessary to reverse engineer and manufacture medicines.

2. Transactional Flow

In order to achieve freedom to operate and to minimize administrative overhead, we propose adapting commons-based approaches to create a self-enforcing open licensing regime for biomedical R&D. Under this approach, when a university licenses a health-related technology to a firm, the university obtains all of the necessary rights to ensure freedom to operate in LMI countries for any resulting products. This requires that any rights in an end product which belong to the licensee must be transferred to the university via a grant-back and cross-licensing structure.[286] The

about diversion, see SANJAY BASU, PHARMACEUTICAL PRODUCT DIVERSION: DIVERTING ATTENTION AWAY FROM THE REAL PROBLEM? (Oxfam Briefing Paper No. 35, forthcoming 2005) (manuscript at 3-4, on file with authors) (reporting that "the scope of product diversion and the difficulty of controlling it have been exaggerated by the pharmaceutical industry," and noting that generic drugs have been produced in India for decades without undermining Western markets).

284. *See* WTO, *Paragraph 6*, *supra* note 141, § 2(b)(ii) (requiring product differentiation); *see also* BASU, *supra* note 283 (manuscript at 7-8).

285. Pharmaceutical companies may be particularly concerned about the effects of differential pricing on their negotiations with high-income markets where prices are set by national regulators. *See, e.g.*, Lanjouw, *supra* note 118, at 2 (describing international pricing externalities).

286. "A grant-back clause in a patent license requires the licensee to grant back to the licensor patent rights which the licensee may develop or acquire." 6 DONALD S. CHISUM,

transferred rights only allow the university to grant licenses to third parties who wish to supply the end product in LMI countries. To take advantage of these licensing terms, the third party licensee must simply notify the university and the university's licensee of its intent to operate under the protection of the EA clause.

a) Identifying Appropriate Technologies and Beneficiaries

i) Target Technologies

The first step of EA licensing is to identify an appropriate technology. Generally speaking, EA licenses will be most appropriate and feasible where the value of a technology is clear and the university controls a good deal of it. The EA approach will be more difficult to apply to technologies that are inchoate or where technologies clearly have small potential commercial value. Much depends, in other words, on a university's bargaining power—which, as noted above, can be substantially increased if universities adopt a standardized, collective approach.

Because the EA approach seeks to share aspects of an innovation that have little commercial value, it should be possible to use the approach when licensing a wide variety of technologies. The most obvious candidates are potential pharmaceutical products, both "small molecule" drugs (for example, aspirin, cisplatin, and stavudine) and "biologic" therapies (for example, insulin, Epogen, and Herceptin).[287] Small molecule compounds are readily reverse engineered, and thus are ideal candidates for EA licensing. Biologics—which include a wide array of therapeutic protein products, from vaccines to monoclonal antibodies—present a potentially more complicated situation. This is due to the increased complexity associated with the production of biologics.[288] While there is no reason to

---

CHISUM ON PATENTS § 19.04[3][j] (2003). Technically, an EA license would utilize both a grant-back (for newly developed or acquired rights) and a cross-licensing mechanism (for any existing licensee rights that could be used to block production of the end product).

287. Biologics are referred to by a variety of names—including biologicals, macromolecules, and biopharmaceuticals. For an assessment of the biologic medicines and genomic technologies most likely to be of use in developing countries, see Abdallah S. Daar et al., *Top Ten Biotechnologies for Improving Health in Developing Countries*, 32 NATURE GENETICS 229, 229-30 (2002).

288. Biologics are structurally more complex and difficult to characterize than small molecules. Correspondingly, the manufacturing processes are both more complicated and challenging to reproduce—biologics are typically derived from living cells, rather than synthesized through chemical processes. *See* Shawn Glidden, *The Generic Industry Going Biologic*, 20 BIOTECHNOLOGY L. REP. 172 (2001); Michael Kleinberg & Kristen Wilkinson Mosdell, *Current and Future Considerations for the New Classes of Biologicals*,

categorically exclude biologics from EA licenses, ensuring freedom to operate in this context may require additional steps.

Although often neglected in discussions focused around access to medicines, diagnostic technologies—for example, those that may help more accurately diagnose cervical cancer or determine whether people with HIV have tuberculosis—should not be ignored. They are essential to the doctor's arsenal, and may be highly amenable to an EA approach.

There is no reason why universities could not also assert EA requirements when licensing manufacturing technologies or even upstream research tools like gene targets. In the past, some universities have attempted to obtain reach-through royalties on upstream innovations such as cell lines or drug screening tools.[289] It should therefore be possible for universities to seek access provisions with a similar reach-through structure.

---

61 AM. J. HEALTH-SYS. PHARMACY 695, 695-97, 701-02 (2004) (including a description of the added costs associated with the manufacture of biologics). Although biologic therapies are mostly under patent because they are relatively new, it is estimated that roughly $10 billion worth of these products will be off patent by the end of 2006. Arman H. Nadershahi & Joseph M. Reisman, *Generic Biotech Products: Provisions in Patent and Drug Development Law*, BIOPROCESS INT'L, Oct. 2003, at 26. Several companies now focus on producing generic biologics, and a number of products have been developed or are in development. *See, e.g.*, Enrico T. Polastro, *The Future of Biogenerics: When Will We See Legal Generics of Top Biopharmaceuticals?*, CONTRACT PHARMA, Oct. 2001 (describing development taking place outside of the principal high-income markets), http://www.contractpharma.com/Oct013.htm. However, to date, none have been approved in the United States or Europe. In fact, there is not yet an established regulatory framework in the United States to assess and approve generic biologics. *See* Kleinberg & Mosdell, *supra*, at 702-03. This regulatory uncertainty stems from the ongoing debate about whether an abbreviated regulatory process—as we have for small molecule generics—is scientifically viable for biologics. *See* Glidden, *supra*, at 176-77 (comparing and contrasting the FDA regulatory challenges of biogenerics with two other similar situations); Selena Class, *Biogenerics: Waiting for the Green Light*, IMS HEALTH, Oct. 28, 2004 (focusing on arguments by makers of biogenerics), *at* http://www.ims-global.com/insight/news_story/0410/news_story_041027a.htm; *FDA Looks at Biogeneric Issue, But Action Unlikely in the Near Term*, SPECIALTY PHARMACY NEWS, Nov. 10, 2004 [hereinafter *FDA Looks at Biogeneric Issue*] (summarizing the debate and describing the European Union's framework for case by case assessment of biosimilar comparability), http://www.aishealth.com/DrugCosts/specialty/SPNFDABiogeneric.html. The FDA has said that it will issue regulatory guidance this year, but regardless of the outcome, generic biologics may be found in other markets. Mike Faden, *Biogenerics Hang at the Starting Gate*, PHARM. BUS. STRATEGIES, Mar. 2005, http://www.pbsmag.com/Article.cfm?ID=169; *see also FDA Looks at Biogeneric Issue*, *supra* (noting that Australia has approved one biogeneric and that various biogenerics are being sold in Asian and South American markets with relatively lax regulatory systems).

289. *See* NAT'L INSTS. OF HEALTH, *supra* note 103.

Finally, a technology appropriate for EA licensing ought to be health-related.[290] As long as this standard is met, an EA clause should be applied, regardless of the type of health condition the product addresses. Universities should resist the pervasive tendency to presume that access concerns in developing countries are limited to drugs for diseases such as HIV/AIDS, TB, and malaria. This tendency is encouraged by pharmaceutical companies, and fuelled by the dangerous misconception that chronic, noncommunicable diseases do not affect developing countries, only affect the elderly, or cannot be effectively treated and prevented.[291]

The majority of the global burden of chronic, noncommunicable diseases such as diabetes, cancer, cardiovascular disease, and chronic respiratory disease—life-threatening conditions for which a significant and growing array of medicines is available in high-income countries—is borne by those living in developing countries.[292] As their prevalence increases, such diseases become an even more pressing public health concern.[293] Cardiovascular diseases, malignant neoplasms, and chronic respiratory diseases each cause more deaths in developing countries than does HIV/AIDS.[294] These conditions are not only common in developing countries, their implications are also more severe. Individuals in developing countries tend to die sooner and at a higher rate from chronic diseases than do individuals in

290. Note that there is no legal or moral reason that universities should limit EA licensing to technologies that come from their pharmacology departments, medical schools, and molecular biology programs. Innovations in fields such as engineering or agriculture can also have a vital health impact, and we expect that universities will be concerned about LMI country access with regard to any technology that has a health-related—or more broadly, a human welfare—benefit. Our focus, however, is in biomedical technologies. We leave an assessment of the value of EA licensing to these other fields for those who are expert in them. By "health-related," we mean any technology with a demonstrated medical benefit. We are less concerned with EA licensing for so-called "lifestyle" drugs.

291. Derek Yach et al., *The Global Burden of Chronic Diseases*, 291 JAMA 2616, 2620 (2004); *see also id.* (noting that cardiovascular disease accounts for as many deaths in young and middle-aged adults as HIV/AIDS in developing countries).

292. *Id.* at 2616. This is of course partly due to the fact that eighty percent of the world's population resides in less developed regions—as do ninety-five percent of new persons added to the world each year. *See* UNITED NATIONS SECRETARIAT, THE WORLD AT SIX BILLION 3 (1999), ESA/P/WP.154 (citing the U.N. Population Division), http://www.un.org/esa/population/publications/sixbillion/sixbilpart1.pdf.

293. *See* Yach et al., *supra* note 291, at 2616 (noting, for example, that approximately 298 million people in developing countries are expected to suffer from diabetes by 2030).

294. *Id.* at 2618 fig.2. Consider also that "[i]n South Africa, infectious diseases account for 28% of years of lives lost, while chronic diseases account for 25%." *Id.* at 2617.

high-income countries.[295] Although the treatment of communicable diseases generates a distinct set of positive externalities,[296] from both a health and human standpoint, there is no reason to distinguish between types of diseases or medicines.

Universities should therefore apply EA licensing to technologies relevant to all diseases, including medications for cancer and heart disease, interventions related to diabetes, and so forth. Indeed, EA licensing may be more effective to alleviate the disease burden of "global diseases" like cancer and diabetes than it will be for neglected diseases. Where a university technology only has an application in developing countries, the innovation is unlikely to be developed without a partner, such as the Drugs for Neglected Diseases Initiative. Such partners will themselves be both motivated and well-suited to address access concerns, meaning that it may either be unnecessary or superfluous to insert EA terms into these licenses.

ii) Identifying Beneficiary Countries and Sectors

An EA license must also identify beneficiary countries and beneficiary sectors within these countries. We contend that, in order to meet the health needs of patients in developing countries, EA provisions must include middle-income countries, as well as the right to supply the private sector in LMI countries. Excluding these markets would substantially undermine the university's attempt to address the access gap.

It is true that some middle-income countries have rapidly growing economies, and may come to represent a larger percentage of the pharmaceutical market over the years. Of course, those that grow sufficiently to be recognized as high-income countries will no longer be beneficiaries of the license. In the meantime, middle-income countries are characterized by highly unequal income distributions.[297] Although some residents in

295. *See* DEREK YACH & CORINNA HAWKES, TOWARDS A WHO LONG-TERM STRATEGY FOR PREVENTION AND CONTROL OF LEADING CHRONIC DISEASES 11 (2004) (noting that "72% of deaths from all chronic diseases occur in low- and middle-income countries, and death rates are higher among all age groups").

296. *Cf.* WAGSTAFF & CLAESON, *supra* note 63, at 118-19.

297. The Gini index is the most popular measure of income or resource inequality. A score of 100 on the Gini index would represent absolute inequality (where one person held all the wealth of a society), and a score of zero would represent absolute equality. *See* Statistics Div.—Advisory Comm. on Indicators, United Nations, *at* http://unstats.un.org/unsd/indicatorfoc/indsearchpage.asp?cid=87 (last visited Apr. 5, 2005). Many of the middle-income countries that companies might most like to exclude from an EA provision are very high on the Gini index. *See* UNITED NATIONS DEV. PROGRAMME, HUMAN DEVELOPMENT REPORT 188-91 (2004). For example, Brazil, Mexico, and South Africa all have Gini index values of more than fifty. In Brazil, the richest ten percent consume 46.7% of the country's income, while the richest ten percent in South Africa consumes

middle-income countries are wealthy, a large portion of them are destitute.[298] And, they, along with the poor in low-income countries, typically must obtain their own care in the private sector.[299] If EA licenses limit low-cost generics to the public sector in LMI countries, or exclusively to low-income countries, they will leave out many individuals who universities aim to benefit.[300]

Additionally, excluding middle-income countries threatens the potential effect of EA provisions in the place they might otherwise work best. EA provisions are most likely to be applied to medicines that will be developed for wealthy country markets, such as those addressing chronic, noncommunicable diseases. As between low-income and middle-income countries, it is in fact middle-income countries that are in more acute need of such medicines.[301]

Moreover, both middle-income countries and the private sector generally are critical to ensure that there are sufficient incentives to sustain the generic companies providing the medicines in question. As profit-seeking enterprises, they must evaluate whether the available markets justify their investment in reverse engineering and scaling-up production; these relatively larger markets figure prominently in this determination.

Finally, any line dividing markets within or between nations is in some sense arbitrary. In theory, some combination of measures that would more finely track income, disease prevalence and distribution, and the purchasing power of the public system would likely be more satisfying intellectu-

---

46.9% of that country's income. China's Gini index value is 44.7, and Thailand's is 43.2. For comparison purposes, the Gini index value of the United States is 40.8. *Id.*

298. *See* UNITED NATIONS DEV. PROGRAMME, *supra* note 297, at 147-49. In South Africa, 23.8% of individuals live on less than $2 per day; the comparable numbers for Thailand and China are 32.5% and 46.7% respectively. *Id.*

299. *See supra* notes 65-67 and accompanying text.

300. Some might object to the inclusion of middle-income countries on the basis that international transfers of wealth should not be directed at countries that have a reasonably high capacity to address access concerns but are failing to do so. However, the only alternative, in this case, is to punish individuals for the inaction of their governments. Furthermore, we are not persuaded that the sharing strategies adopted here are best thought of as "transfers of wealth," since the good being shared is nonrival. Indeed, one might consider that imposing rules requiring limits on market competition and thus permitting rent-extraction from economies whose demand pull has no positive incentive effect on R&D, by firms located in the rich countries and that orient their research towards demand from rich countries, is a form of regressive tax on access to the international trade system.

301. *See* YACH & HAWKES, *supra* note 295, at 12 (noting that "large middle income, low mortality developing countries"—countries such as China are exactly the ones that companies are most likely to seek to exclude—are particularly heavily affected by chronic diseases).

ally than a simple geographic/income based division. Universities could try to set up a process to evaluate the effect upon patients of excluding particular middle-income countries or the private sector in select LMI countries, but we suspect that this would not be worth the effort. Because profits from such countries and sectors are unnecessary to stimulate product development, and given the extreme difficulty of defining and implementing distinctions that are more closely tailored to poor patients' ability to pay, the optimal distinction will likely be the one drawn between high-income countries and LMI countries.

b) Flow of Rights

EA licensing involves limited cross-licensing between the university and its licensees, structured to create freedom to operate for third parties for the benefit of LMI country distribution.[302] In exchange for permission to use the university's exclusive rights in high-income countries, the licensee and its sublicensees cross-license exclusive rights they own in the end product to the university. This cross-license is limited and available only for the purpose of an automatic sublicense flowing from the university to any third party who notifies the university and licensee of its intent to supply an LMI market and fulfills some additional requirements (we refer to this third party as "the notifier," or, where relevant, as "the improver"). The university need not hold a patent in the LMI countries where the drug is to be distributed. It is only necessary that the university own technology that the licensee wishes to use, in exchange for which the licensee agrees to the limited cross-license. After the initial agreement is reached, to obtain freedom to operate, all the generic manufacturer need do is notify the university and its licensee. It thereby receives a limited license to all of the patents that belong to the university or its licensees that are necessary to produce the ultimate end product for distribution solely in LMI countries.

i) Cross-License and Grant Back

The first transactional element of an EA license is an exchange of licenses. The university grants to the licensee rights to a particular technology or innovation, and sets the parameters of the license. The license will likely include, at a minimum, rights to practice the university's technology in some or all high-income countries. In exchange, the licensee will cross-license to the university its "associated rights." These rights must include

302. In order to ensure that the EA structure is self-perpetuating, the license should require that any sublicenses carry the terms of the EA license with them. This is important because the initial license may be with a biotech company, which will sublicense the university technology to a pharmaceutical company only after further development.

all of the potentially exclusive rights it holds that could prevent a third party from producing or delivering an eventual end product, including rights in any patents and data possessed by the licensee during the term of the license that are necessary to make, use, sell, import, or export the end product. This right does not reach know-how or any other secret or material property possessed by the licensee. It would, however, cover associated rights that the licensee possessed or developed that do not rely directly upon the university technology but are nonetheless necessary to the production or sale of the end product.

In coming years, rights to clinical trial data are likely to become an increasingly important tool of exclusion in developing countries.[303] EA clauses must therefore include such data in the bundle of rights received from the licensee and openly sublicensed to the notifier. Within the EA model, the license to use data means only that no exclusive rights will prevent the generic company from relying in its application on publicly available data generated by the licensee or the fact that the drug has been registered in another country.[304] The license removes the formal right to exclusivity. It does not give a notifier the authority to obtain otherwise nonpublic data from the university's original licensee. The generic producer will, of course, still have to meet other regulatory requirements related to bioequivalence and manufacturing standards, to the extent these requirements exist in the notified country.

The university obtains these rights for the sole purposes of providing freedom to operate in LMI countries. Although in some circumstances grant-back arrangements for open source projects may implicate the patent misuse doctrine,[305] this does not appear to be a concern in this case.[306] Figure 1 illustrates this initial transactional flow.

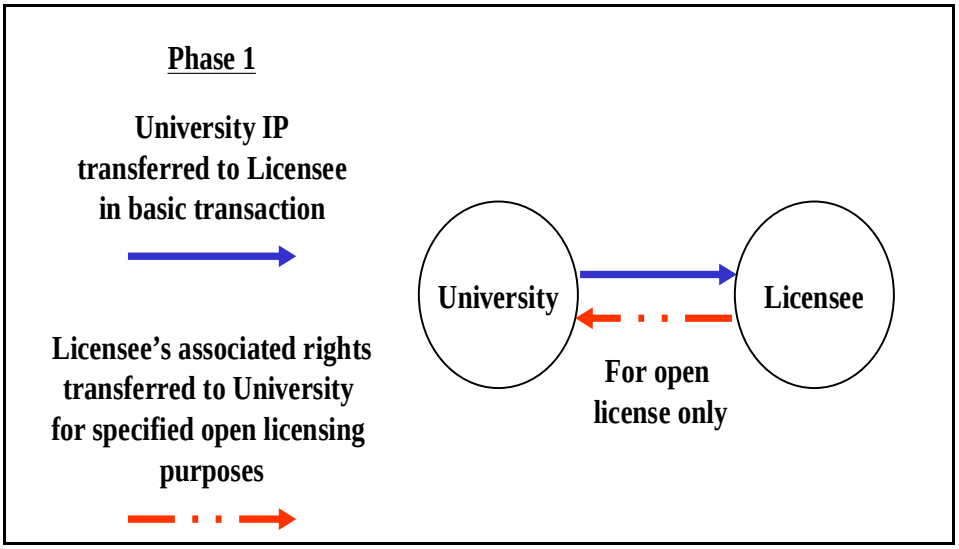
---

303. This is due to provisions in Article 39.3 of the TRIPS Agreement and the increasingly stringent data exclusivity provisions in free trade agreements. *See* Scafidi, *supra* note 46.

304. National drug regulatory agencies differ widely in the data they require and the processes they follow. Some allow generic companies to rely on the fact that a drug has been registered in another country, or on data that was submitted to regulatory agencies in another country. *See* FINK & REICHENMILLER, *supra* note 48, at 2-3.

305. According to a federal district court, for a grant-back license agreement to constitute patent misuse, the licensee generally "must provide specific evidence that the clause actually stifled innovation." Robin Feldman, *The Open Source Biotechnology Movement: Is It Patent Misuse?*, 6 MINN. J.L. SCI. & TECH. 117, 155 (2004) (discussing *Transparent-Wrap Machine Co. v. Stokes & Amith Co.*, 166 F. Supp. 551 (S.D.N.Y. 1958), and other case law). Thus, exclusive grant-backs may raise concerns, as may agreements that cover products invented using a research tool rather than incorporating a patented invention, but nonexclusive grant-backs are typically acceptable. *Id.* at 156-59.

Figure 1: EAL Transaction Flow Phase 1



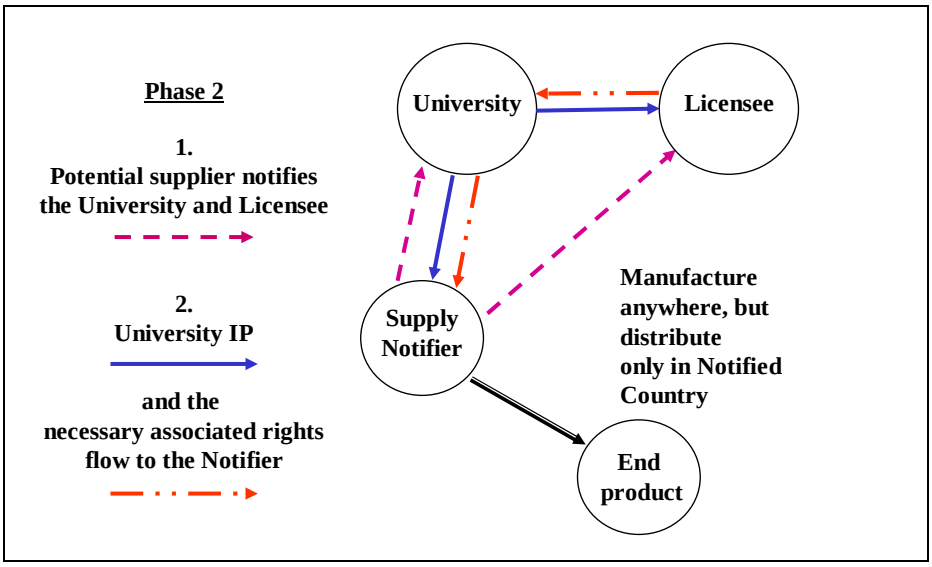
ii) Notification

The second transactional component of an EA license is an automatic open licensing structure organized around a simple notification procedure. This component's core attribute places power to act in the hands of a third party, typically a generic company. Any disputes about the applicability of the freedom to operate are left to the licensee and the party seeking to enter the market, permitting the university to remain largely out of the picture. Figure 2 illustrates this transaction.

306. Courts typically use two methods to determine whether a patentee's actions constitute misuse: the antitrust rule of reason and the patent policy inquiry. *See id.* at 167. An antitrust inquiry focuses on whether the licensing agreement has anticompetitive effects and, if so, whether those effects outweigh the agreement's pro-competitive benefits. A patent policy inquiry examines whether the agreement is consistent with patent policy. *Cf. id.* at 163-65.

Because the EA license promotes competition and ultimately seeks to increase rather than restrict competition, it is difficult to argue that any anticompetitive effects exist, or that those effects could outweigh pro-competitive effects. *See id.* at 163-65. In contrast, one might argue that the EA license conflicts with patent policy because it reduces incentives to innovate, forcing competition for LMI-country markets and thereby reducing licensee profits. But the limited nature of monopoly rents available from LMI markets suggest instead that an EA clause is highly unlikely to harm the patent-based pharmaceutical industry's incentives to innovate, and that it might stimulate innovation by nonprofit and generic pharmaceutical companies. *Cf. id.* at 159-63.

Figure 2: EAL Transaction Flow Phase 2



EA provisions are triggered when a third party notifies the university and licensee that it intends to make, use, or sell the end product in, or import the end product into, an LMI market. The notifier can be any entity, but we anticipate three primary users of the notification procedure: (1) generic companies that wish to produce or sell in an LMI country; (2) a government agency such as a ministry of health, or NGO such as MSF, that wishes to import generics from a third party; or (3) a researcher who wishes to adapt the end product for developing country use.

In order to foster a competitive environment, the EA model presumes that multiple entities may notify for a particular market. Although it would be possible to arbitrarily limit the number of firms that could notify, and thereby shelter the first notifiers from a fully competitive environment, this has obvious risks. Over time some form of limited exclusivity might be required to induce generics to introduce a product to market. But at this time, there is no clear evidence of this need, and generic manufacturers have entered LMI markets when patents did not present a barrier without promise of exclusivity. Should practice indicate that some stronger inducement is necessary, the standard approach could be revised to offer a limited period of exclusivity for the first notifier that brings a product to market in a particular country.[307]

307. *Cf.* 21 U.S.C. § 355(j)(5)(B)(iv) (2000) (authorizing a 180-day period of exclusivity for the first generic to enter the market in the United States).

Upon notification, the open licensing provisions of the EA license are engaged. The university's licensed rights, including associated rights from its licensee, flow to the notifier for the sole purpose of manufacturing for distribution and distributing in the notified country. Patent, regulatory, and manufacturing barriers are lifted for the notifying entity by this flow of rights. This can be achieved by a statement in the EA license that a notifier shall receive from the university an open license permitting the making, using, selling, offering to sell, importing, and exporting of the end product in the notified country. A royalty payment could be required in consideration for the open licenses. For low-income countries, the license could specify a rate within the range recommended by UNDP of zero to six percent.[308] Because middle-income countries can afford more, on average, sales in these countries could be subject to a slightly higher flat rate.[309] Finally, the license will also have to establish an equitable division of royalties between the university and the licensee.

It is important to note that this model permits production of the end product in any country (including a high-income country), as long as manufacture is for the sole purpose of exporting to and supplying the end products in the notified country. This will increase the likelihood of finding a generic supplier for more complicated drugs. It will also create maximal competition in the market to supply LMI countries, which in turn will drive prices towards marginal cost.

---

308. *See* UNITED NATIONS DEV. PROGRAMME, HUMAN DEVELOPMENT REPORT 108 (2001) (citing a normal rate of four percent, and providing for adjustments up and down), http://www.undp.org/hdr2001/chapterfive.pdf. Other novel approaches to determining royalty rates merit consideration: in May 2004, Canada passed legislation to give effect to the WTO's August 30th decision, which allows Members with manufacturing capacity to export compulsorily licensed drugs for the benefit of Members without such capacity. *See* WTO, *Paragraph 6*, *supra* note 141. Subsequently, the Department of Industry published a draft implementing regulation for public comment. The draft regulation sets forth a novel formula for calculating the royalties to be paid by the developing nation licensees to the patentees. Currently, calculations using the formula produce a potential range of royalties from .02% to 3.5%. *See Use of Patented Products for International Humanitarian Purposes Regulations: Regulatory Impact Analysis Statement and Draft Regulation*, CANADA GAZETTE, Oct. 2, 2004, http://gazetteducanada.gc.ca/partI/2004/20041002/html/regle9-e.html.

309. If complexity and inequality within middle-income countries were less of a concern, EA licenses could also seek to implement proposed royalty models that better reflected a country's ability to pay. *See, e.g.*, William Jack & Jean O. Lanjouw, *Financing Pharmaceutical Innovation: How Much Should Poor Countries Contribute?*, 19 WORLD BANK ECON. REV. (forthcoming 2005), http://www.georgetown.edu/faculty/wgj/jack-lanjouw-draft.pdf.

iii) Notifier Improvements

If the EA license defines the terms "end product" and "open license" appropriately, it can also operate to permit notifiers in any country to engage in research to improve the end product. This could substantially benefit patients because it would allow companies and academic researchers in those countries to adapt the technology to local circumstances in a way that a proprietary company might be unwilling or unable to do. For example, the first three-in-one pill for AIDS patients was developed not in the United States or Europe, but in India—and it was created without any guarantee of exclusivity.[310] As described in Part II, many products must be altered in specific ways to meet the needs of patients in developing countries. Some of these modifications, such as pediatric dosing, require minimal investment and are currently being undertaken by generic companies.[311] These examples suggest that if potential innovators are ensured freedom to experiment and sell improved versions of products in LMI countries, we may see not only cheaper products in these countries, but better ones as well.

To meet these goals, however, any such improvements should be licensed back to the university for the sole purpose of sublicensing them under EA terms to subsequent notifiers in LMI countries. The notifier's improvements would themselves be subject to the EA terms. The notifier would be paid royalties for the use of its improvements in LMI country markets, but the notifiers could not prevent others from exploiting the improvements.

Some might advocate allowing the improver to patent its own improvements in high-income countries and then to negotiate the necessary cross-licenses with the university and/or licensee. The opportunity for a potentially lucrative cross-license might offer the notifier an incentive to make innovative improvements (although it is not immediately clear how well these would align with the improvement needs of LMI markets). The alternative is to include in the EA license a requirement that the improver grant the university and/or licensee an option to license any improvement. The EA license could specify the terms of this option, namely a reasonable royalty rate for licensing the improvement.

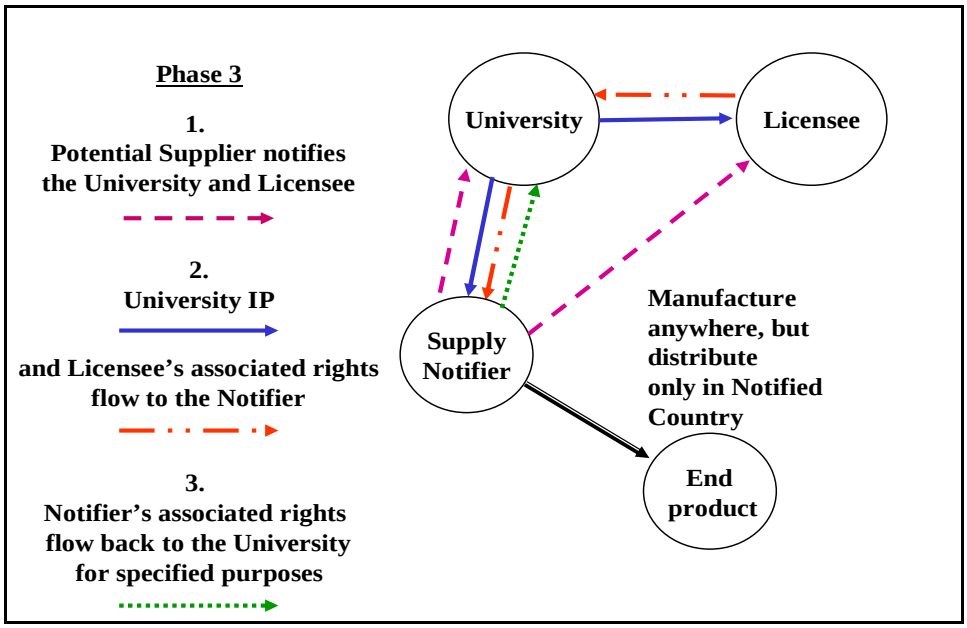
---

310. *See* Outterson, *supra* note 72, at 254.

311. *See, e.g.*, Ranbaxy Pharmaceuticals Inc., Ranbaxy Pharmaceuticals Announces Nationwide Availability of DipserMox™ (First-Ever Amoxicillin Tablets for Oral Suspension) in Time for the Respiratory Season (Nov. 3, 2003), http://www.ranbaxyusa.com/newsroom/03-11-03.htm.

We have now described the complete flow of rights associated with an EA license. Figure 3 illustrates this transactional flow.

Figure 3: EAL Transaction Flow Phase 3



iv) Resolving Disputes

Under an EA license a notifier is automatically deemed to have an open license; therefore, it may immediately and lawfully begin to sell the end product in the specified LMI country without infringing upon any rights held by the licensee or sublicensee.[312] An EA license is itself the legal protection provided to any entity making use of the license's provisions. While structured as a license, it operates at a minimum as a covenant by the university and its licensee not to sue entities that rely on technology to which they have rights solely for the EA license purposes. There is some legal risk involved for the entities that seek to rely on the EA license, because of the potential for variations between jurisdictions with regard to such third-party reliance on the provisions of a license to which they were not a party. Nonetheless, because of the relatively widespread use of covenants not to sue and the small value of the markets covered, that risk is likely to be manageable. Generics always have the option of

312. Of course, separate from the matter of rights provided by the EA license, the generic entrant will have to comply with any regulatory requirements in the LMI country (for example, registration) before it will be able to sell its product.

also seeking a more direct license. If they do so, the existence of the threat to operate under the EA license as a fallback is likely to improve the entrants' negotiating position.[313]

If the licensee or the university wishes to contest the applicability of the license to the product or patents included in the notification, they may of course do so, by challenging the actions of the notifying party and/or taking legal action. A notifying party operating inconsistently with the terms of the EA provisions (for example, by seeking to sell in a high-income country or seeking to sell products that are not covered by the license) will infringe the underlying patent rights and will be subject to the usual remedies for patent infringement.

v) Additional Concerns

EA licensing is appealing because it provides simple, clear freedom to operate with one stroke of the pen—the signing of the original license between the university and its licensee. Of course, additional provisions could be added to an EA license to meet specific concerns of the university and licensee. For example, universities could ensure that only manufacturers with a certain demonstrated capacity to produce quality products can legally notify, by requiring notifiers to have a certification of Good Manufacturing Practices (GMP)[314] or other guarantee of quality.[315] This would involve the university in a broader effort to police the actions of generic companies, creating obstacles where there are no established quality assessment standards. The same objective might be achieved more easily through an indemnification and insurance requirement, which universities may well wish to have in any case.

---

313. A secondary assurance might also be built into an EA license in the form, for example, of a statement that the acceptance of the notifier's royalty payment represents a covenant not to sue, guaranteeing additional legal protection from any later claims that the patents and products notified for are not covered by the underlying license. The utility of such a mechanism is doubtful, however, given that the rates involved and value of sales will likely be small. High-income market licensees may choose to reject the royalty in order to preserve their options and increase uncertainty.

314. For a description of GMP, see WHO, WORLD MEDICINES SITUATION, *supra* note 67, at 98 fig.9.1.

315. A license could turn on registration with the WHO's Prequalification Project, which provides governments and pharmaceutical manufacturers with information about how to ensure the quality, safety, efficacy, and rational use of pharmaceutical products. The Project focuses on a small number of priority medicines, which are, to date, those related to TB, HIV/AIDS, and malaria. World Health Org., Essential Medicines and Policy Dept. (EDM), Prequalification Project, http://mednet3.who.int/prequal/about.htm (last visited Mar. 15, 2005).

With some technologies, know-how and materials, which cannot be transferred under the freedom to operate model, may be an essential aspect of the rights a licensee uses to control production.[316] Any attempt to require affirmative transfer of materials or information from the licensee to a third party could raise some of the same enforcement challenges that a fair pricing approach would. There is no reason, in principle, that an EA license could not seek to bind a licensee to provide enabling know-how and associated materials reasonably necessary to the production of the end product. However, if a licensee violated the agreement, concerns over privity, in particular with respect to claims brought by an open and undefined class, suggest that it would often be left to the university to bring an enforcement action.[317]

Rather than providing a stark choice between enablement and litigation, an EA license might instead specify an intermediate step, requiring negotiations between all parties if enablement and transfer of materials were requested. In those cases where a university is concerned that products will only become available if the licensee itself produces them, it may not be able to avoid becoming involved in enforcement. It could then require transfers of know-how to third parties or seek to regulate the licensee's distribution diligence and pricing directly. This degree of individually negotiated requirements, monitoring, and enforcement may be beyond the resources and negotiating power of any individual university. If biologics become a much more important component of the pharmaceutical market, as some predict,[318] and generic companies are unable to readily reverse engineer them, effective pursuit of an EA strategy may require creation of a standing inter-university body charged with shepherding the performance and utilization of EA licenses. Modeled perhaps on PIPRA,

316. *See* Glidden, *supra* note 288, at 178-80 (describing the importance of non-patented trade secrets in the manufacture of biologics); Gil Y. Roth, *Biomanufacturing Report*, CONTRACT PHARMA, June 2003 (describing challenges of producing a protein identical to a branded drug without materials or know-how), http://www.contractpharma.com/June032.htm.

317. A university could deem a licensee who failed to provide enabling know-how or materials in breach of the original primary license, and use its own powers of persuasion to facilitate the enablement. However, a university may be reluctant to travel this road. *Cf. supra* text accompanying note 281.

318. *See* Press Release, IMS Health, IMS Health Reports 2004 Global Pharmaceutical Sales Grew 7 Percent to $550 Billion (Mar. 9, 2005) ("Biotech products accounted for 27 percent of the active research and development pipeline, and 10 percent of global sales in 2004. IMS expects that over the next five years, innovative products derived from biotechnology will continue to grow in the double digits and represent an increasing share of the overall market."), http://www.imshealth.com/ims/portal/front/articleC/0,2777,6599_3665_71496463,00.html.

such a body would have to include a staff and collective funding mechanism and would be named specifically as an assignee of the university's rights under each EA license entered by a member university.

B. Addressing the R&D Gap: Implementing Neglected Disease Clauses and Innovative Partnerships

The EA licensing approach is designed to harness technologies developed through university technology transfers to industry. This approach will do little to address the lack of direct investment into research for neglected disease. Additional changes in the way that universities manage their IP portfolios can reduce barriers to R&D in these areas.

1. Neglected Disease Exemptions

The first strategy universities can adopt is one we term Neglected Disease ("ND") licensing. If a university chooses to enter into an exclusive license for a research tool (a practice that we do not mean to advocate by making this proposal), it can insert a specially tailored research exemption into the license. Utilizing the same notification structure as the EA provision, the ND clauses would grant scientists worldwide the freedom to engage in research to address neglected diseases using the licensed technology. Just as importantly, ND exemptions guarantee those who conduct this research the right to market resulting products in LMI countries. Such an exemption could be applied to all technologies useful in biomedical research, from research tools to compounds intended for end products. The ND exemption we propose utilizes an open licensing approach, like the EA license, and is similarly commons-based.

Our ND proposal draws on the model proposed by the PIPRA initiative[319] and adapts it to provide researchers and producers in LMI countries freedom to operate with biomedical research tools. Unlike the EA provisions, the ND clauses do not necessarily entail obtaining a cross-license from the licensee. Instead, the ND clauses simply must carve-out of any exclusivity granted to the university's licensee a set of provisions for freedom to operate pertaining to neglected diseases. Such a core clause specifies that, notwithstanding anything else in the agreement, the university retains the right to license use of its technology for research on neglected diseases anywhere in the world and for commercial purposes in LMI countries. In this case, notification provided to the university alone will result in open licenses to conduct such activities. As with EA clauses, implementing an open licensing structure would minimize transaction costs

319. *See* Pub. Intellectual Prop. Res. for Agric., Draft Definition of Humanitarian Use, *supra* note 208.

and allow any party to engage in research for a neglected disease after simple notification. A more robust model, which would more closely mirror the equitable access approach, would also capture all licensee improvements on the university's technology in the open licensing pool.[320] Critically, the ND clauses are not limited in geographical scope. Any entity is eligible to conduct research using the university's patented innovation—and if the more robust version is used, any licensee improvements to it—without paying a royalty, provided that the research targets a neglected disease.

Two approaches to defining the scope of the ND research exemption are possible. First, uses could be limited to academic institutions and other nonprofit entities (such as the Drugs for Neglected Diseases Initiative) that have as their primary aim producing products predominantly for patient populations in developing countries. The second approach would allow any entity to make use of the exemption, but to carefully establish the universe of applicable diseases. An ND license could, for example, allow an open license to any institution, public or private, for research targeting any disease on a list included in the license.[321] A more comprehensive and flexible approach would be to provide a standard for identifying rare diseases and to grant an open license to any scientist working on any disease meeting that standard. Current U.S. law defines a rare condition as one with an incidence of less than 200,000 persons in the United States or for which there is no reasonable expectation of recouping the necessary R&D investment in the U.S. market.[322] The FDA makes available a cumulative

320. This approach has, in fact, been taken in both the WARF standard licenses and the BIOS initiative. *See supra* notes 187, 205 and accompanying text. Like the EA approach, ND provisions should not conflict with the patent misuse doctrine. *See* Feldman, *supra* note 305.

321. One such list might be derived from the influential Trouiller study. *See* Trouiller et al., *supra* note 81, at 2189. An alternative might include any disease with some set proportion of its burden in developing countries. *Cf.* Kremer, *supra* note 58, at 71 tbl.3.

322. 21 U.S.C. §§ 360bb(a)(2), 360ee(b)(2) (2000); 21 C.F.R pt. 316.1 (2004). The FDA has developed specific criteria for classifying orphan drugs. *See* FDA, Cumulative List of Orphan Drug Products Designated and or Approved Through 2005 (Apr. 6, 2005), *at* http://www.fda.gov/orphan/designat/alldes.rtf. To incentivize development of drugs for rare conditions, companies are rewarded with regulatory-based exclusivity. 21 U.S.C. § 360cc. As a result, orphan drugs for serious, chronic diseases with small but steady U.S. patient populations may yield substantial revenues. For example, Genzyme makes several hundred million dollars a year on Ceredase, an expensive orphan drug for Gaucher Disease. *See, e.g.*, James Love, *The Other Drug War*, AM. PROSPECT, June 1993, at 121, *available at* http://www.prospect.org/web/page.ww?section=root&name=ViewPrint&articleId=5121.

list of all drugs for such diseases,[323] including a number for diseases of particular significance in LMI countries.[324] By taking the United States' approach, universities would be adopting a widely accepted definition of indications for which markets fail to provide.

As noted above, in practice, the most significant IP barriers to research may result from potential exclusion from commercializing a resulting invention. Therefore, the most important part of an ND exemption may be the assurance of freedom to exploit any eventual product in LMI countries. This can be accomplished by guaranteeing freedom to operate vis-à-vis the licensed technology in LMI countries.

A researcher acting under the ND exemption we propose would not have the right to commercialize an end product in a high-income market, unless she negotiated the necessary cross-license(s). An ND clause might mandate that the licensee receive an option for a cross-license for all high-income markets. Such a provision would likely appeal to licensees, and would ensure that the end product would not be barred from high-income markets on account of failed cross-licensing negotiations. However, avoiding such a mandate might provide greater incentives for private firms to engage in the relevant research.

Critics may express concern that the contractual creation of a worldwide neglected research exemption—both for the underlying university patent and any licensee or sublicense improvements—will actually lead to scientists using these technologies in research on non-neglected diseases. However, such research is not authorized by an ND clause, and would constitute actionable infringement. The pertinent question is whether ND uses can readily be distinguished from other uses in an infringement context. We argue that ND uses can be distinguished in the ways that matter most, and that where they cannot, little harm is done. When a drug is registered with a regulatory agency, any misuse of the ND research exemption would likely become apparent. Of course, even with researchers acting in good faith, early-stage research may produce results applicable to a variety of indications, including non-neglected diseases. The ND exemption does not prevent a researcher from negotiating cross-licenses in order to exploit such an innovation. Where, on the other hand, the attempt to li-

323. *See*, FDA, Cumulative List of Orphan Drug Products Designated and or Approved Through 2005, *supra* note 322.

324. *Id.* (including dengue fever, Chagas disease, leishmaniasis, malaria, and TB). The FDA has approved orphan drug status for products intended to treat subsets of disease populations; a drug indicated to treat a particular stage or strain of a disease, or a particular category of patients (for example, AIDS patients with symptomatic HIV infection and CD4 count below 200/mm3), may qualify for orphan drug designation. *Id.*

cense or patent an innovation does not reveal the infringement, the infringement is likely of the class that is difficult to detect, and thus commonplace even in an environment without ND licensing.

2. Promoting Partnerships

The second component of universities' neglected disease agenda would be a more proactive approach to out-licensing of research tools. Universities need not wait until they exclusively license a technology to ensure that the technology is available to researchers working on neglected diseases. Universities can affirmatively grant scientists royalty-free licenses to use their tools for commercial and noncommercial research. This might be facilitated by the creation of simple, ready-to-sign agreements that could be posted on a TTO's website.

Universities can also seek out opportunities to license to public-private partnerships, and try to bring foundations into the agreement to provide support for the development of the technology.[325] Again, such initiatives should not be limited to attempts to produce medicines, but should also include diagnostics.[326] Finally, universities should explore the option of licensing early-stage inventions directly to entities in LMI countries that have the ability and desire to commercialize products for both neglected and non-neglected diseases.[327] Such agreements might sometimes offer limited forms of geographical exclusivity or co-exclusivity,[328] or leverage public or foundation financing to support the development of the technology. Such partnerships have many potential benefits. For example, they could help meet goals of development and technology transfer, and make use of the relatively low cost of research in LMI countries.[329]

---

325. One model here could be the agreement between the University of California at Berkeley, OneWorld Health, Amyris Biotechnologies, and the Gates Foundation to secure both research freedom and funding to develop a steady, affordable supply of the antimalarial drug artemisinin. Press Release, University of California at Berkeley, $43 Million Grant from Gates Foundation Brings Together Unique Collaboration for Antimalarial Drug (Dec. 13, 2004), http://www.berkeley.edu/news/media/releases/2004/12/13_gates.shtml.

326. *See* Rayasam, *supra* note 86 (describing an effort at the University of Texas to develop an inexpensive, rapid technology to count CD4 cells, an important component of care for individuals with HIV/AIDS).

327. The NIH Office of Technology Transfer has begun to adopt this approach, and has licensed or is in the process of licensing technologies to institutions in Mexico, Brazil, India, Chile, Argentina, China, Korea, Egypt, Indonesia, South Africa, and other sub-Saharan African countries. Salicrup et al., *supra* note 189, at 6.

328. The NIH has offered exclusive, partially exclusive, and nonexclusive licenses covering both patents and biological materials. *See id.* at 10 tbl.2.

329. *Id.* at 9.

C. Intersections Between EA Clauses, ND Clauses, and Partnerships

These approaches can, of course, be combined. EA and ND clauses can be implemented together to ensure freedom for suppliers of an eventual end product in LMI countries as well as freedom for researchers in high-income countries who seek to develop the compound for use against a neglected disease.[330] Similarly, EA clauses can be inserted into ND licenses to ensure that any resulting products must be licensed under terms that guarantee generic companies freedom to operate in LMI countries. Finally, when licensing to a nonprofit entity such as OneWorld Health, universities could adopt either EA or ND clauses.[331] Exactly how and when to supplement one approach with another will likely depend on the particular technologies and partners in question.

VI. CONCLUSION

We have highlighted a series of institutional innovations that could constitute the backbone of a new agenda for access to biomedical innovations and research on treatments for neglected diseases. One strong advantage of our approach is that it can be undertaken in the absence of any changes to national or international IP regimes. By collectively adopting such an agenda, as well as clear and binding policies governing the use of these approaches, universities can maximize their joint potential to close the R&D and access gaps and improve the lives of people living in LMI countries. No one—not pharmaceutical companies, not patients in developed nations, and not universities—benefits from letting people in poor countries die from conditions that could be prevented or treated.

---

330. For a model license that aims to integrate the EA and ND provisions in this way, see Model Provisions for an Equitable Access License, *at* http://www.essentialmedicine.org/EAL.pdf. We are grateful to all those who participated in the interdisciplinary working group, based at Yale University and organized by Universities Allied for Essential Medicines, to develop this document.

331. However, such clauses may be unnecessary or inappropriate when dealing with nonprofits of this sort, whose ethos and mission are already closely aligned with those of universities. Because these nonprofits are generally expert in their particular areas of research, they may be better positioned than universities to determine the best strategy to ensure access for researchers and patients. In the artemisinin deal between Berkeley and OneWorld Health, for example, Berkeley granted OneWorld Health an exclusive right to the University's relevant patent rights. *See* E-mail from Carol Mimura, Director, Office of Technology Licensing, to Yochai Benkler, Professor of Law, Yale Law School (May 6, 2005) (on file with authors).

We must find ways around the many myopic and technical stumbling blocks that contribute to millions of preventable deaths each year. In the best case scenario, this voluntary solution will pave the way for IPR disarmament among a wide range of actors both in the United States and elsewhere—including universities, scientists, federal legislators, federal agencies, nonprofit drug development companies, and the pharmaceutical industry itself. On the other hand, there may be no spillover effects beyond providing access to university-generated medicines and research tools. Perhaps only a small percentage of research on neglected diseases is redirected, abandoned, or delayed because of problems accessing research tools. Perhaps patent-based costs account for only a few percent of preventable deaths from diseases in low- and middle-income countries. Perhaps open access to university-based technologies would only avert a fraction of these deaths and free up a fraction of the research tools relevant to neglected diseases. But preventing even a fraction of one percent of deaths in low- and middle-income countries would translate into saving tens of thousands of lives every year. The opportunity to prevent these deaths is a worthy goal for the community of scientists and universities to pursue, and to pursue together.

THE DEATH OF THE PUBLIC FORUM IN CYBERSPACE

By Dawn C. Nunziato[†]

TABLE OF CONTENTS

I. INTRODUCTION

Throughout the past decade, the Internet has been conceptualized as a forum for free expression with near limitless potential for individuals to express themselves and to access the expression of others. Indeed, some Internet law scholars have claimed that the Internet will enable us to realize, for the first time in our nation's history, the free speech values embod-

---

© 2005 Dawn C. Nunziato

† Associate Professor, The George Washington University Law School. I am grateful to Jerome Barron, Robert Brauneis, Laura Heymann, Ira Lupu, Todd Peterson, David Post, as well as to the participants of the Penn-Temple-Wharton Colloquium, for their helpful comments on a prior draft. I am also grateful to David Ludwig, who provided excellent research assistance, and Leonard Klein, who provided excellent library assistance, for this Article.

ied in the First Amendment. Other Internet law scholars have claimed that the Internet should be conceptualized as one grand "public forum"—First Amendment parlance for a place in which the right to free speech receives its strongest protection. These scholars claim that the Internet's architecture and structure facilitate freedom of expression and that, to advance the cause of freedom of expression, the United States government should simply get out of the way and hand over the regulation of Internet speech to private regulators.[1]

During the Clinton Administration, the United States government essentially acceded to these calls for privatization, undertook measures to turn over many aspects of the Internet to private entities, and convinced Congress to do the same. With the government's withdrawal from management of the Internet, private entities assumed control. The end result is that the vast majority of speech on the Internet today occurs within private places and spaces that are owned and regulated by private entities such as Internet service providers (ISPs) like America Online (AOL) and Yahoo!, Internet access providers like employers and universities, content providers like washingtonpost.com and nytimes.com, and pipeline providers like Comcast and Verizon. In contrast to real space (which enjoys a mixture of privately- and publicly-owned places in which speech occurs) and in contrast to media channels such as broadcast and cable television (which enjoy publicly-subsidized and public forums), speech in cyberspace occurs almost exclusively within privately-owned places. The public/private balance that characterizes real space and renders the First Amendment meaningful within it is all but absent in cyberspace.

Private regulation of speech on the Internet has grown pervasive, and is substantially unchecked by the Constitution's protections for free speech, which generally apply only to state actors' regulations of speech. At an earlier stage of the Supreme Court's First Amendment jurisprudence, such private speech regulation might have been subject to the dictates of the First Amendment under the state action doctrine.[2] The Supreme Court, however, has substantially limited the application of the state action doctrine in past decades, and courts have been unwilling to extend this doctrine to treat private regulators of Internet speech as state actors for purposes of subjecting such regulation to First Amendment scrutiny.

Given the judicial contraction of the state action doctrine, ISPs and other private regulators of Internet speech will likely continue to be exempt from First Amendment scrutiny. The result of such privatization is

1. *See infra* Part III.A.
2. *See infra* Part II.A.

that the vast majority of Internet speech regulation is conducted by private parties who are not subject to the First Amendment's protections for free speech. Those private entities have sole discretion over whether and how speech is regulated in such privately-owned places. Some scholars applaud this result and contend that free speech values are best advanced by facilitating a proliferation of private speech decisions without intervention or control by the government.[3] According to this conception of the First Amendment, the sole function of this constitutional guarantee is the negative one of checking the government's restrictions of speech.

This negative conception of the First Amendment, however, fails to account for the important affirmative role that the government has played, and should continue to play, in facilitating freedom of speech and correcting imperfections in the market for free expression. In particular, this negative conception fails to account for the important role the government plays in providing public forums for expression and protecting speech from censorship within such forums. Under the Supreme Court's public forum doctrine, the government has the affirmative obligation to dedicate public property, of the kind traditionally well-suited to the expression and exchange of ideas, to the public's use for free speech purposes. Indeed, it is only within such public forums that free speech rights are accorded their most robust protection. Government regulations of speech within such public forums are subject to the most exacting First Amendment scrutiny, and, accordingly, individuals enjoy their strongest free speech rights within such forums.

The required existence of public forums ameliorates the inequalities that disparities in private property ownership would otherwise impose on individuals' free speech rights. To exercise their free speech rights meaningfully, individuals need forums from which to express themselves. Yet many individuals do not own property, much less property from which they can effectively express themselves on matters of importance within our democratic system. The requirement that the government maintain public forums compels the government to provide such individuals with forums from which to exercise their First Amendment rights. Public forums, at least in real space, therefore subsidize the speech of those who otherwise would not be able to express themselves effectively.

On the Internet, however, essentially no places exist to serve as "public forums" because the places within which expression occurs are overwhelmingly privately owned. As a result, the government's affirmative role of advancing free speech by providing speakers with meaningful fo-

3. *See infra* Part III.A.

rums from which to express themselves free of censorship is non-existent on the Internet.

The absence of public forums in cyberspace augurs the absence of meaningful protection for free speech under the First Amendment. In real space, the existence of government-owned property as a forum for speech available to all comers provides an important guarantee for such speech. In contrast, as increasingly more speech takes place in cyberspace, the affirmative constitutional protections for free speech that exist in real space are in danger of being sacrificed. In particular, the government's abdication of control over Internet speech regulation may well result in the loss of protection for speech that is insufficiently protected within an unregulated market for speech (viz., unpopular and poorly-subsidized speech).

Congress and the courts have declined to take the steps necessary to update First Amendment jurisprudence to account for modern communications media and the radical shift in the public/private balance within such media. The Supreme Court recently declined to accord public forum status (and therefore declined to extend meaningful First Amendment protection) to even the comparatively minor portion of public "property" on the Internet. In *United States v. American Library Ass'n*,[4] the Court held that public libraries' provision of Internet access via publicly-owned computers did not constitute a public forum and therefore that restrictions on speech in that context were not subject to meaningful First Amendment scrutiny. As a result of this and similar developments, the important functions served by the First Amendment in general and by the public forum doctrine in particular are in danger of being seriously eroded in cyberspace. Despite the common perception among members of the public and Internet law scholars that the Internet is a forum for free expression of unprecedented scope and importance, in fact there are essentially no places on the Internet where speech is constitutionally protected against censorship.

In Part II of this Article, I describe the scope and extent of private ownership of the Internet and private regulation of Internet forums for speech. Having been encouraged by Congress to assume the mantle of Internet speech regulation, private entities have imposed substantial controls on Internet speech. Courts, following the Supreme Court's contraction of the state action doctrine, have likewise declined to subject private Internet speech regulation to First Amendment scrutiny.

In Part III, I analyze the important role served by public forums within our system of democratic self-government. I set forth two competing conceptions of the First Amendment, the first of which is consistent with the

4. 539 U.S. 194 (2003) (plurality) [hereinafter *Am. Library Ass'n II*].

privatization of the Internet and the second of which informs the public forum doctrine. Under the first conception, the First Amendment is conceptualized purely as a restraint on government restrictions on speech and has no role in scrutinizing private regulation of speech. Under the second conception, the First Amendment imposes obligations on the government and other regulators of speech to provide meaningful opportunities for expression free of censorship. This conception finds its clearest expression in the public forum doctrine, under which courts impose the affirmative obligation on the government to dedicate certain publicly-held property for the use and benefit of individuals seeking to exercise their free speech rights. I explore the speech facilitating roles served by public forums in real space, and examine what the absence of public forums means for cyberspace.

In Part IV, I analyze *United States v. American Library Ass'n*. After analyzing this and other cases declining to meaningfully apply First Amendment scrutiny to the government's restrictions of speech within Internet forums, I set forth in Part V several ways in which courts and legislatures should act to reintroduce the values of the public forum into cyberspace.

II. PRIVATE REGULATION OF SPEECH ON THE INTERNET

A. The Early Years: The Internet's Promise as an Unprecedented Forum for Free Expression

Over the past decade, the Internet has been conceptualized as a forum for free expression of unprecedented scope and importance. Once Congress lifted limitations on the permissible uses of the Internet, the Internet opened up as a forum for expression of all kinds,[5] and speakers and publishers from all walks of life and from every corner of the world flocked to the Internet.[6] As one court explained, "It is no exaggeration to conclude that the Internet has achieved, and continues to achieve, the most participatory marketplace of mass speech that this country—and indeed the world—has yet seen."[7]

---

5. In 1992, Congress granted authority for commercial uses of NSFNET, which was to become an Internet backbone. Management of Internet Names and Addresses, 63 Fed. Reg. 31,741, 31,742 (June 10, 1998), *available at* http://www.ntia.doc.gov/ntiahome/domainname/6_5_98dns.htm.

6. *See, e.g.*, ACLU v. Reno, 929 F. Supp. 824, 831 (E.D. Pa. 1996), *aff'd*, 521 U.S. 844 (1997).

7. *Id.* at 881.

Several features constitutive of today's Internet[8] render it a uniquely powerful vehicle for speakers and publishers to express themselves to worldwide audiences at very low cost.[9] For the (very low) cost of establishing a website, an individual can express herself in the context of a whole host of media—text, images, audio, video—to a virtually unlimited array of listeners. The barriers to entry that exist in other mediums of expressions, such as traditional print publication and broadcast media, are drastically reduced in the context of the Internet. In contrast to traditional broadcast media, where the ability to express oneself widely is constrained by substantial licensing requirements and associated fees, the Internet is not shackled by spectrum scarcity, nor by the onerous licensing requirements or fees necessitated by a limited broadcast spectrum. As a result, the Internet, to a much greater extent than traditional methods of expression, has the potential to facilitate a true marketplace of ideas, one that is not dominated by the few wealthy speakers who are able to express themselves effectively via traditional media.[10]

Because of the Internet's unprecedented speech-facilitating characteristics, early commentators contended that the Internet should be conceptualized as one grand public forum.[11] Consistent with the speech utopian rhetoric of early court decisions like the district court opinion in *ACLU v. Reno*, commentators viewed the Internet as constituting an important new forum for public discourse. Perhaps because it was not precisely clear to what extent the government would involve itself in the ownership and control of the Internet, early commentators believed that the "National Information Infrastructure" would preserve an important place for genuine public forums in cyberspace. As government ownership and control of the

---

8. This is not to say that the *inherent* nature of the Internet presumes such features. *See generally* LAWRENCE LESSIG, CODE AND OTHER LAWS OF CYBERSPACE (1999) (arguing against an essentialist conception of the Internet's "nature").

9. *See Reno*, 929 F. Supp. at 877.

10. Indeed, because of the Internet's combination of such speech-friendly features,

> [i]ndividual citizens of limited means can speak to a worldwide audience on issues of concern to them. Federalists and Anti-Federalists may debate the structure of their government nightly, but these debates occur in newsgroups or chat rooms rather than in pamphlets. Modern-day Luthers still post their theses, but to electronic bulletin boards rather than the door of the Wittenberg Schlosskirche. More mundane (but from a constitutional perspective, equally important) dialogue occurs between aspiring artists, or French cooks, or dog lovers, or fly fishermen.

Id. at 880.

11. *See, e.g.*, David Goldstone, *The Public Forum Doctrine in the Age of the Information Superhighway*, 46 HASTINGS L.J. 335 (1995).

Internet has substantially receded, however, the role of public forums in cyberspace has substantially differed from these commentators' predictions.

B. Private Regulation of Internet Speech Forums

Despite the Internet's potential as a forum for expression of unprecedented scope and importance, today private actors wield the vast majority of power over Internet speech—power unchecked by the First Amendment. While it is often presumed that speech on the Internet will be "uninhibited, robust, and wide-open,"[12] the private entities that own and control the forums for Internet speech enjoy and often exercise the unfettered power to impose substantial restrictions on such speech.

The extent of such private speech restrictions is staggering. Each of the major ISPs establishes and enforces Terms of Service by which it prohibits the expression of certain types of speech that fall within the protection of the First Amendment. AOL, for example, specifies in its Terms of Service that AOL and its agents "have the right at their sole discretion to remove any content that, in America Online's judgment . . . [is] harmful, objectionable, or inaccurate."[13] AOL explains in its Community Guidelines that "like any city, we take pride in—and are protective of—our community."[14] Unlike any other city, however, AOL enjoys the unfettered discretion to censor constitutionally-protected speech in its discussion forums and other online spaces, including "vulgar language" (which, it warns, is "no more appropriate online than [it] would be at Thanksgiving dinner"), "crude conversations about sex," and "discussions about . . . illegal drug abuse that imply it is acceptable."[15] AOL members may hope to escape from these speech-restrictive Terms of Service by leaving AOL-sponsored forums and expressing themselves elsewhere on the Internet. AOL, however, informs its members that "as an AOL member, you are required to follow our [Terms of Service] no matter where you are on the Internet," and warns that it may terminate the accounts of users who violate its Terms of Service by engaging in AOL-prohibited speech anywhere on the Internet.[16]

---

12. *See* N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964).

13. America Online, Agreement to Rules of User Conduct, *at* http://www.aol.com/copyright/rules.html (last visited Apr. 25, 2005).

14. America Online, Community Guidelines, *at* http://legal.web.aol.com/aol/aolpol/comguide.html (last visited Apr. 25, 2004).

15. *Id.*

16. *Id.*

Internet users seeking stronger protection for their expression might turn to an ISP other than AOL. They will find, however, similar restrictions on speech imposed by many other major ISPs. Yahoo!'s Terms of Service, for example, prohibit users from making available content that is, inter alia, "objectionable," and specify that Yahoo! may pre-screen and remove any such "objectionable" content.[17] Similarly, Comcast prohibits users from disseminating material that "a reasonable person could deem to be objectionable, embarrassing, . . . or otherwise inappropriate, regardless of whether this material or its dissemination is unlawful."[18] And Comcast, by its Terms of Service, "reserves the right . . . to refuse to transmit or post and to remove or block any information or materials . . . that it, in its sole discretion, deems to be . . . inappropriate, regardless of whether this material or its dissemination is unlawful."[19]

Colleges and universities, both private and public, serve as Internet access providers for millions of students across the United States. Many of these have established and enforced "acceptable use policies" that substantially restrict First Amendment protected speech. To list just a representative sample: Colby College restricts speech that may cause "a loss of self-esteem" within the individual to whom it is addressed;[20] Brown University prohibits speech that produces "feelings of impotence, anger, or disenfranchisement," whether "intentional or unintentional";[21] while Bowdoin College restricts jokes and stories that are "experienced by others as harassing."[22]

The amount of communication via e-mail has far surpassed the amount of communication via "snail" mail. While the U.S. Postal Service is subject to the dictates of the First Amendment when performing its duties, the private entities that are predominantly responsible for relaying billions of e-mails per day are not, thus these entities are free to monitor and censor

---

17. Yahoo!, Terms of Service, *at* http://docs.yahoo.com/info/terms (last visited Apr. 25, 2005).

18. Comcast Cable Communications, LLC, Comcast High-Speed Internet Acceptable Use Policy, *at* http://www.comcast.net/terms/use.jsp (last visited Apr. 25, 2005).

19. *Id.*

20. Found. for Individual Rights in Educ., *Colby College*, Speechcodes.org, *at* http://www.speechcodes.org/schools.php?id=659 (last visited Apr. 25, 2005). *See generally* Found. for Individual Rights, *at* http://www.speechcodes.org (last visited Apr. 25, 2005) (detailing speech-restrictive acceptable use policies for private and public universities throughout the United States).

21. Found. for Individual Rights in Educ., *Brown University*, Speechcodes.org, *at* http://www.speechcodes.org/schools.php?id=2483 (last visited Apr. 25, 2005).

22. Found. for Individual Rights in Educ., *Bowdoin College*, Speechcodes.org, *at* http://www.speechcodes.org/schools.php?id=2481 (last visited Apr. 25, 2005).

the content of the e-mails that they are responsible for delivering. Private employers, which serve as Internet access providers for millions of employees across the United States, routinely monitor and restrict e-mail (and Internet use generally), with approximately 50-60% of employers monitoring e-mail.[23] In short, the vast majority of Internet access and service providers, which are privately owned, assert and exercise substantial control over the expression that flows through their Internet places.

In the remainder of this Section, I set forth a few representative scenarios that elucidate the nature and extent of such private regulations of Internet speech. These examples demonstrate the ways in which private entities, including Internet providers like Google, AOL, and Yahoo!, have broadly exercised the power to regulate and censor speech on the Internet wholly exempt from First Amendment scrutiny.

Google is the largest and most popular Internet search engine, providing access to billions of websites and serving as the first place most people turn to find information on the Internet. In addition to providing links to webpages as results for search terms, Google also provides "sponsored links," which are advertisements keyed to search terms entered by web surfers. Because of Google's dominant position in the search engine market, the ability to secure a sponsored link is a valuable medium of expression on the Internet. Indeed, for many less well-known advocacy groups and causes whose sites do not otherwise enjoy substantial web traffic, sponsored links serve as an important means of drawing attention to the sites' content. Google, however, has been wielding its power as a private speech regulator to censor a considerable amount of valuable expression, including political speech that would fall within the core of the First Amendment's protection. Adhering to its policy of refusing to accept sponsored links that "advocate against any individual, group, or organization,"[24] Google has refused to host a range of politically-charged, religious, and critical social commentary in the form of advertisements themselves, as well as the websites to which these advertisements link. Google has also required prospective advertisers to alter the content within their sponsored links—as well as within their websites—as a condition for Google's hosting such content. Furthermore, Google's speech regulations

23. *See* AMA RESEARCH, 2004 WORKPLACE E-MAIL AND INSTANT MESSAGING SURVEY SUMMARY, *at* http://www.amanet.org/research/pdfs/im_2004_summary.pdf.

24. Google, Google Ad Sense Program Policies, *at* http://www.google.com/adsense/policies (last updated Mar. 8, 2005).

and restrictions apply not only on Google's site but also on other websites that run the sponsored links, including AOL, Ask Jeeves, and EarthLink.[25]

A few instances of such censorship illuminate the extent of Google's power as a private regulator of speech. In August 2004, W. Frederick Zimmerman, who maintains a political website called the Zimmerblog, sought to advertise his book *Basic Documents About the Detainees at Guantanamo and Abu Ghraib*, which contained, inter alia, the full text of the *Hamdi v. Bush*, *Rumsfeld v. Padilla*, and *Rasul v. Bush* opinions, as well as various applicable Geneva Convention documents. Google suspended Zimmerman's account once it became aware of the material Zimmerman was advertising via this sponsored link, informing Zimmerman that "Google policy does not permit the advertisement of websites that contain 'sensitive issues.'"[26] In a similar incident in June 2004, John Perr, author of the PERRspectives website, which contains "left-of-center" political commentary, sought to advertise his website via a Google Sponsored Link, with advertisements entitled "The Liberal Resource: Analysis, Commentary, and Satire" and "Complete Liberal Resource Center." The linked-to website contained an article written by Perr that was critical of President George W. Bush and which characterized the president, inter alia, as "secretive, paranoid, and vengeance-filled."[27] Once Google became aware of this language within Perr's article, a Google representative informed Perr that Google was removing his sponsored link because it linked to a website that contained text critical of Bush and therefore "advocates against an individual, group, or organization," in violation of Google's policy.[28]

---

25. *See* Verne Kopytoff, *Google's Ad Rules Complex, Controversial*, S.F. CHRON., Aug. 9, 2004, at F1.

26. W. Frederick Zimmerman, *Guantanamo/Abu Ghraib Ads Banned by Google*, Nimble Books, *at* http://www.wfzimmerman.com/index.php?page=3 (last visited Mar. 16, 2005).

27. Jon Perr, *Google's Gag Order: An Internet Giant Threatens Free Speech*, PERRspectives, June 20, 2004, *at* http://www.perrspectives.com/articles/art_gagorder01.htm.

28. *Id.* In another example, when *Unknown News* sought to advertise anti-Iraq-war bumper stickers on Google's Sponsored Links with an ad headlined "Who Would Jesus Bomb?," Google censored the ad, claiming that the ad was in violation of its policy against "sites that promote hate, violence, racial intolerance, or advocate against any individual, group, or organization." Unknown News, *Google Refuses Our Ad*, *at* http://www.unknownnews.net/google.html (last visited Apr. 25, 2005). When Unknown News responded to Google's censorship decision by explaining that it merely "advocate[s] against killing thousands of Iraqis," Google explained that it would reinstate the ad only if the website was edited "to show both sides of the argument" over attacking Iraq. *Id.* The ad was reinstated only after multiple appeals to Google. *Id.* Similar instances of

It is not only political or socially-charged speech that runs the risk of censorship by Google; religious speech also appears to be a target. For example, while Google generally permits sponsored links for abortion services, it prohibits all sponsored links for abortion services that make any reference to religion.[29] Google also singles out for special treatment links sponsored by the Church of Scientology and requires that particular language appear in any link sponsored by this church.[30] No other religion is subject to the requirement that particular language be contained within its advertisements. In short, Google enjoys and exercises substantial censorial control over the content of sponsored links (as well as the content of websites linked to by such sponsored links) that appear on Google's search page.

AOL is by far the largest ISP in the world, with over twenty-four million subscribers in the United States.[31] Indeed, AOL is responsible for hosting the exchange of more messages on a daily basis than the United States Post Office.[32] It is also the largest single forum for individuals to express themselves online, offering thousands of discussion forums on a

---

Google wielding its policy to censor speech abound. Amy Harmon, *Is a Do-Gooder Company a Good Thing?*, N.Y. TIMES, May 2, 2004, at 12 (rejection by Google of a sponsored link by an owner of a T-shirt shop in Los Angeles unless the owner removed from his site all T-shirts with slogans critical of President Bush); Kopytoff, *supra* note 25, at F1 (rejection by Google of a sponsored link by an individual who sought to advertise playing cards that featured fifty-four reasons to defeat President Bush because the ad was "advocating against" the president); Michael Liedtke, *Google Bans Environmental Groups' Ads*, Boston.com, Feb. 12, 2004 (allowance by Google for Oceana, a nonprofit environmental group, to run an ad criticizing Royal Caribbean Cruise Line's environmental policies under the headline "Help us protect the world's oceans," quickly followed by rejection from Google because of its policy prohibiting advertisements criticizing groups or companies), *at* http://www.boston.com/business/technology/articles/2004/02/12/google_bans_environmental_groups_ads; Katherine C. Reilly, *Google's 'Haphazard' Ad Policy*, NATION, Aug. 12, 2004 (refusal by Google to host an advertisement by *The Nation*, which was headlined "Bush Lies," because of Google's policy against ads that advocate against any individual), *available at* http://www.thenation.com/doc.mhtml%3Fi=20040830&s=reilly; *Google Censorship*, MediaRights, *at* http://www.mediarights.org/news/announcement.php?ann_id=04489 (last visited Apr. 25, 2005) (censorship by Google of an advertisement by The Cat's Dream, an independent filmmaker seeking to advertise its controversial documentary *XXI Century*, which contained the text "American Voices Against Bush," because of Google's policy against ads that advocate against individuals).

29. Kopytoff, *supra* note 25, at F1.

30. *Id.*

31. *See* Alex Goldman, Top 22 U.S. ISPs by Subscriber: Q4 2004, ISP-Planet, Mar. 25, 2005, *at* http://www.isp-planet.com/research/rankings/usa.html.

32. *See* Andy Kessler, Wired: Stop the U.S. Mail, Jan. 25, 2005, *at* http://andykessler.com/wired_stop_the_us_mail.html.

wide variety of topics. Within these discussion groups, AOL's twenty million plus subscribers can express their views on topics of concern to them and engage in online debates with other interested individuals.

One of the discussion forums hosted by AOL is the "Irish Heritage" discussion group for individuals interested in Irish history and politics.[33] In 1998, AOL's discussion forum "monitors" grew concerned about the heated nature of the discussions within this forum and apparently felt that some of the contributions were getting out of hand.[34] AOL determined that several exchanges on the Irish Heritage discussion group were in violation of AOL's terms of service, which do not allow members to "harass, threaten, embarrass, or do anything else to another member that is unwanted." Accordingly, AOL shut down the discussion group for a "cooling off" period, and in the process removed all traces of the earlier heated postings, contrary to AOL's regular practice of preserving past discussions in its archives.[35] After a seventeen-day cooling off period, AOL reopened the discussion group, while encouraging members to make this forum "a more amiable place" and announcing that members who committed three or more violations would face suspension or termination of their accounts.[36]

AOL's speech restrictions do not stop with the Irish Heritage discussion group. As discussed above, AOL has long imposed rigorous regulations on the vast amount of speech that it hosts in AOL-owned places, prohibiting speech that advocates the illegal use of drugs, speech that uses "crude" or "inappropriate" terminology in connection with sex, and the like.[37] AOL's censorial actions within its discussion groups are now well-known. Accordingly, the "unprecedented forum for free expression" that is the Internet in actuality is largely dominated by AOL's forums for expression, wherein the regulation of speech is subject to AOL's Terms of Service.

AOL is, of course, a private company, and the servers on which it hosts its discussion forums are its private property. Because it is a private

---

33. *See* Amy Harmon, *Worries About Big Brother at America Online*, N.Y. TIMES, Jan. 31, 1999, at 1.

34. *Id.*

35. *Id.*

36. AOL has also wielded more targeted tools to prevent certain individuals from engaging in heated discussion with one another. For example, two AOL subscribers were instructed by AOL never to speak to one another in AOL space again, or else face suspension or termination of their accounts. *Id.*

37. *See* Dawn Nunziato, *Exit, Voice, and Values on the Net*, 15 BERKELEY TECH. L.J. 753, 756-57 (2000).

entity, the determinations that it makes regarding the types of expression allowed within its property are not subject to scrutiny under the First Amendment. AOL enjoys complete discretion to allow or disallow whatever speech it likes on the vast number of discussion forums and websites that it hosts, and to enforce its speech restrictions by removing unwanted speech and prohibiting unwanted speakers from partaking in its discussion forums. Accordingly, for the great majority of Internet speakers, it is not the First Amendment, but AOL's (or other ISPs') terms of service, that determine the contours of protection accorded to their Internet expression.

The popular ISP Yahoo! recently fought and won a highly publicized international battle on behalf of free speech and First Amendment values, only to turn around and exercise its prerogative as a private speech regulator to censor the same constitutionally protected speech that it fought to protect. In *La Ligue Contre le Racisme et l'Antisémitisme v. Yahoo! Inc.*,[38] several French organizations committed to combating anti-Semitism brought suit in French court against Yahoo!.[39] Plaintiffs alleged that Yahoo!'s auction site was hosting auctions of materials such as *The Protocols of the Elders of Zion*, *Mein Kampf*, and other Nazi materials and memorabilia, the display of which within France violated French law. The French lawsuit, which involved novel issues of international jurisdiction and choice of law, ultimately resulted in a French court order compelling Yahoo! to cease making available the specified anti-Semitic content to French citizens (which essentially required Yahoo! to cease making this content available on the Internet at all).[40]

Yahoo! was concerned that the French court would be able to enforce this order against it within the United States. To forestall the enforcement of the French court's order within the United States, Yahoo! filed suit in U.S. district court against the French organizations. The company claimed that enforcement of the French court judgment in the United States would violate the First Amendment.[41] The U.S. district court agreed, holding that principles of international comity that would generally favor enforcing international courts' judgment against United States entities were out-

38. T.G.I. Paris, May 22, 2000, Gaz. Pal. 2000, somm. jurispr. 1307. An English translation is available Yahoo! Case Tribunal De Grande Instance De Paris May 22, 2000, JURISCOM.NET, *at* http://juriscom.net/txt/jurisfr/cti/yauctions20000522.htm (last visited May 12, 2005).

39. *Id.*

40. *Id.*

41. Yahoo! Inc. v. La Ligue Contre le Racisme et l'Antisémitisme, 169 F. Supp. 2d 1181, 1186, 1194 (N.D. Cal. 2001).

weighed by the First Amendment values at play in this case.[42] Because the First Amendment protected Yahoo!'s dissemination of anti-Semitic speech, the enforcement of the French court order enjoining such dissemination would violate the First Amendment.[43]

Yet in a surprising turn of events, shortly after securing this unprecedented victory for the dissemination of First Amendment protected speech over the Internet, Yahoo! chose to exercise its prerogative as a private regulator of Internet speech to prohibit the dissemination of the Nazi-related content at issue in the case.[44] Other Internet search engines and service providers also refuse to host Nazi-related and other controversial content, even though such speech is protected by the First Amendment against government censorship.[45]

C. Congress's Encouragement of Private Internet Speech Regulation

One might suppose that the censorship of Internet speech described above would cause concern among those charged with protecting our First Amendment freedoms. Surprisingly, however, neither Congress nor the courts have looked critically upon speech restrictions imposed by private Internet actors. Rather, courts have rejected challenges to private Internet actors' speech restrictions on the grounds that such actors are not state actors, nor the functional equivalent of state actors, under applicable First Amendment doctrine.[46] Further, Congress, far from looking critically upon such "private" speech restrictions, has affirmatively encouraged Internet actors to exercise discretion to restrict First Amendment protected expression.

In passing the Communications Decency Act of 1996 (CDA),[47] Congress sought to remedy perceived ills caused by certain types of offensive Internet expression (primarily sexually-themed expression). Congress set out to remedy these speech harms by using two different approaches. First,

---

42. *Id.* at 1193.

43. *Id.* However, a Ninth Circuit panel overturned the district court's decision because the district court improperly asserted personal jurisdiction over the French parties. 379 F.3d 1120 (9th Cir. 2004). The Ninth Circuit has granted en banc review of the case. 399 F.3d 1010 (9th Cir. 2005).

44. *See* Lori Enos, *Yahoo! To Ban Nazi-Related Auctions*, E-COMMERCE TIMES, Jan. 3, 2001, *available at* http://www.ecommencetimes.com/story/6432.html.

45. Alexander Tsesis, *Hate in Cyberspace: Regulating Hate Speech on the Internet*, 38 SAN DIEGO L. REV. 817, 866 (2001).

46. *See infra* Part II.D.2.

47. Pub. L. No. 104-104, 110 Stat. 56 (codified as amended at 47 U.S.C. § 223 (2000)).

the CDA prohibited the transmission of certain types of sexually-themed expression anywhere on the Internet.[48] These provisions were insufficiently attentive to the First Amendment rights of individuals, and were readily struck down by the Supreme Court.[49] Second, and far more successfully, Congress encouraged private Internet actors to do what it could not do itself—restrict harmful, offensive, and otherwise undesirable speech, the expression of which would nonetheless be protected by the First Amendment (if restricted by a state actor). In Section 230 of the CDA,[50] Congress sought to encourage the proliferation of a free market in Internet speech; a market in forums for expression that would be largely unfettered by government intervention and defined predominantly by private actors' speech choices. Accordingly, Section 230 provides that:

> The Congress finds [that] the Internet and other interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity, [and] the Internet and other interactive computer services have flourished, to the benefit of all Americans, with a minimum of government regulation . . . .
>
> . . . .
>
> It is the policy of the United States . . . to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation.[51]

Recognizing the importance of the Internet as a forum for expression, Congress sought in this section of the CDA to advance the goal of free expression by minimizing the government's role in regulating Internet expression, while at the same time handing over the reins of regulating Internet expression to private actors. In lieu of creating an affirmative role for Congress in protecting free speech on the Internet, Congress chose to defer to private actors to regulate speech as they saw fit, to let a "free market" in expression and in the regulation of expression reign on the Internet. Congress sought to encourage ISPs and other owners of Internet speech forums to restrict expression, and access to expression, that the providers found undesirable. Accordingly, Section 230(c)(2) of the CDA provides:

> No provider . . . of an interactive computer service shall be held liable on account of any action voluntarily taken in good faith to

48. *Id.*
49. *See* Reno v. ACLU, 521 U.S. 844 (1997).
50. 47 U.S.C. § 230 (2000).
51. *Id.*

> restrict access to or availability of material that the provider . . . considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, *or otherwise objectionable, whether or not such material is constitutionally protected*.[52]

Through Section 230, the Government excised itself from the role of affirmatively protecting free expression on the Internet, and passed the mantle of speech regulation over to private entities.[53]

Accordingly, one significant result of the government's privatization of the Internet is the concomitant privatization of regulation of Internet expression. What follows from such privatization is that today there are essentially no places on the Internet where free speech is constitutionally protected. Rather, such Internet speech is only protected, if at all, by the grace of the private entities who control the private spaces in which such speech is hosted. Far from the speech utopian theorists' predictions that the Internet would constitute a public forum in which constitutional protections for free expression were at their apex, today's Internet is constituted by an amalgam of private forums within which constitutional protection for free expression is non-existent. Furthermore, Congress apparently wanted it that way.

D. Private Speech Regulation and the First Amendment

Like Congress, courts have consistently protected the right of private entities to regulate Internet speech unchecked by the First Amendment. In accordance with the Supreme Court's recent First Amendment jurisprudence, courts have refused to apply any First Amendment scrutiny to private Internet actors' restrictions of constitutionally-protected speech.

For the first century and a half of constitutional interpretation, courts consistently applied the First Amendment only to speech restrictions imposed by government actors. In *Marsh v. Alabama*,[54] however, the Supreme Court began to discard formalistic distinctions between public and private regulations of speech and to carefully scrutinize the speech restrictions imposed by entities that wielded government-like power over individuals' speech. In *Marsh*, the Court inaugurated the "company town" doctrine, in which it treated a private corporation that performed certain "traditional government functions" as the equivalent of a state actor for

---

52. *Id.* § 230(c)(2) (emphasis added). Section 230(c) also provides immunity for ISPs from liability for defamatory content posted by their subscribers. *See, e.g.*, Zeran v. Am. Online, Inc., 129 F.3d 327 (4th Cir. 1997).

53. The government also passed on to private entities the task of managing and regulating the Internet's infrastructure. *See* text accompanying notes 105-06.

54. 326 U.S. 501 (1946).

First Amendment purposes. This doctrine was applied and extended through the 1960s, but was curtailed in several decisions in the 1970s.[55] Consistent with this trend of declining to treat private speech regulators as state actors for First Amendment purposes, courts have declined to subject private Internet actors' speech restrictions to any scrutiny whatsoever under the First Amendment.

1. Private Regulation of Speech in Real Space

Marsh v. Alabama involved speech regulations imposed by a "company town," a phenomenon of the Deep South in the early Twentieth Century, in which economically ailing regions encouraged capital investments by allowing corporations to build and operate towns.[56] Though privately maintained, such towns "had all the characteristics of any other American town," including streets, sidewalks, sewers, public lighting, police and fire protection, business and residential areas, churches, postal facilities, and schools.[57] The town of Chickasaw, Alabama, was one such company town, with "nothing to distinguish [it] from any other town and shopping center, except the fact that the title to the property belong[ed] to a private corporation."[58] The streets and sidewalks of the town, which under the public forum doctrine would be considered public forums where individuals would enjoy their strongest free speech rights, were privately owned and regulated.

Grace Marsh, a Jehovah's Witness, sought to distribute religious literature and express her religious views from a sidewalk in Chickasaw. A town official warned her that she could not distribute literature on the town's sidewalks or streets—or anywhere else in the town—without a permit, and that no permit would be issued to her. Marsh was asked to leave the sidewalk and the town, but she refused to do so. She was subsequently arrested and charged with violating state trespass law.[59]

Marsh claimed that the application of the state trespass law under these circumstances would unconstitutionally abridge her First Amendment rights.[60] The Supreme Court agreed, holding that the town's private status did not insulate its regulations of speech from First Amendment scrutiny. In extending the First Amendment's protections to private regulations of speech, the Court emphasized the fact that the streets, sidewalks, and other

55. *See, e.g.*, Lloyd Corp. v. Tanner, 407 U.S. 551 (1972).
56. 326 U.S. at 502.
57. *Id.*
58. *Id.* at 503.
59. *Id.* at 503-04.
60. *Id.* at 504.

places within the town that would be categorized as public forums if owned by the state, were "accessible to and freely used by the public in general," and generally served the same functions as such places serve when publicly owned.[61] The Court looked beyond the formalistic public/private distinction and held that, despite the fact that such places were privately owned, they were "built and operated primarily to benefit the public" and "their operation is essentially a public function."[62] Because such property was open to the public and because the public had an interest in keeping the channels of communication open and uncensored to enable the public to make well-informed decisions as members of a self-governing democracy, even private restrictions on speech within such forums were subject to scrutiny.[63] *Marsh* therefore places primacy on the government's affirmative obligations under the First Amendment to establish and protect the pre-conditions of democratic self-government.[64] As the Court explained, to participate in democratic self-government, individuals must have access to uncensored information and open channels of communication. For the purposes of advancing this goal, it does not matter whether the restrictions on speech are imposed by private property owners or by the government as a property owner. Rather, the dispositive inquiry is whether the speech regulation interferes with the channels of communication essential for individuals to participate meaningfully in democratic self-government.

The *Marsh* Court's rejection of the public/private distinction was carried forward in *Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc.*[65] *Logan Valley* involved the picketing of Weis Market, a non-union supermarket located within a privately-owned shopping center on privately-owned property adjacent to the shopping center. The shopping center had no publicly-owned sidewalks or streets adjacent to the targeted supermarket, and so the picketers' only effective option was to stage their picket on private property adjacent to the supermarket.[66] Members of the Amalgamated Food Employees Union picketed Weis by carrying (truthful) signs stating that the supermarket was non-union and that its employees did not receive union wages or benefits.[67] The picket was

61. *Id.* at 501.
62. *Id.* at 506.
63. *Id.*
64. *See* Thomas Emerson, *Toward a General Theory of the First Amendment*, 72 YALE L.J. 877 (1963).
65. 391 U.S. 308 (1968).
66. *Id.* at 310-16.
67. *Id.* at 311.

staged in the privately-owned areas where the supermarket's customers would pick up their groceries and the adjacent portion of the parking lot.[68] The picketers selected this location so as to effectively convey their message to Weis managers, employees, and customers. In response to the picketing, the owners of the supermarket and of the shopping center instituted an action to enjoin the picketing. The lower courts granted an injunction prohibiting such expression, finding that the picketing constituted a trespass on private property not privileged by the First Amendment.[69]

The Supreme Court reversed. The Court first compared the features and characteristics of the places where the picketing occurred to the place where the expressive activity involved in *Marsh* occurred, and found them to be functionally similar for purposes of the First Amendment inquiry.[70] The Court explained, "[w]e see no reason why access to a business district in a company town for the purpose of exercising First Amendment rights should be constitutionally required, while access for the same purpose to property functioning as a business-district should be limited . . . ."[71] Because the Logan Valley shopping center enjoyed the same features as the sidewalks in *Marsh* and public forums like sidewalks and streets, regulations of speech on such private property was subject to First Amendment scrutiny.[72]

Because the picketers sought to exercise their free speech rights "in a manner and for a purpose generally consonant with the use to which the [private property at issue was] actually put," the Court held that the First Amendment precluded the private property owners from enjoining such expression.[73] As in the *Marsh* decision, the Court in *Logan Valley* looked to the functional characteristics of the property at issue, instead of to the formalistic distinction between public and private ownership of such property, in determining whether and how to protect free speech values within such property. In so doing, the Court properly looked to the characteristics of the property on which the speech regulation occurred, the functional similarities between such private forums and public forums, the openness of the property to the public, and the suitability of such property for expressive purposes, instead of simply to whether the property was held privately or publicly.

---

68. *Id.*

69. Logan Valley Plaza, Inc. v. Amalgamated Food Employees Union Local 590, 227 A.2d 874, 878 (Pa. 1967).

70. *Logan Valley*, 391 U.S. at 319.

71. *Id.*

72. *Id.*

73. *Id.* at 319-20.

Marsh and *Logan Valley* represent the high water mark of the Court's treatment of private speech regulators as state actors for First Amendment purposes. Shortly after the *Logan Valley* decision, the Supreme Court scaled back its protection of free speech against private speech regulation in *Lloyd Corp. v. Tanner*.[74] *Lloyd* involved individuals' efforts to distribute leaflets and other materials to protest the Vietnam War within the privately-owned property of Lloyd Center, a large shopping mall complex.[75] Lloyd Center encompassed approximately fifty acres and contained sixty establishments, including offices of doctors, dentists, lawyers, bankers, travel agents, and persons offering a variety of other services.[76] The private entities who acquired Lloyd Center purchased the land from the city of Portland, which vacated acres of public streets and other public land to make room for the shopping mall complex. Given the extent of goods and services available at the shopping center complex, as well as its central location, "for many Portland citizens, Lloyd Center [would] so completely satisfy their wants that they would have no reason to go elsewhere for goods or services."[77] Indeed, as a testament to the Center's potential for reaching broad general audiences, presidential candidates from both major parties selected the Lloyd Center as the forum from which to reach the broadest audience of Portland residents.[78] Recognizing the Center's potential for reaching a broad cross-section of Portland citizens, several anti-war protestors sought to distribute anti-war materials from within the mall's walkways.[79] The owners of the complex instructed their security guards to warn the protestors that they would be arrested unless they ceased their expressive activities on the mall's private property.[80] The protestors brought suit against the mall owners, claiming that the First Amendment privileged their expressive activities.

The Supreme Court held that the protestors did not enjoy the First Amendment right to express themselves on the mall's private property. The Court distinguished *Logan Valley* by explaining that the picketers in *Logan Valley* were expressing themselves on a subject matter that was directly related to the shopping center's operations—the non-union nature of the Weis supermarket—in circumstances in which there were "no other reasonable opportunities for the picketers to convey their message to their

74. 407 U.S. 551 (1972).
75. *Id.* at 553.
76. *Id.*
77. *Id.* at 580 (Marshall, J., dissenting); *see id.* at 575-81 (Marshall, J., dissenting).
78. *Id.* at 555.
79. *Id.* at 556.
80. *Id.*

intended audience were available."[81] In contrast, the Court explained, the protestors in *Lloyd* were not protesting on a matter related to the shopping center's operations,[82] and the protestors had reasonable alternative opportunities to convey their message to their intended audiences, such as by distributing their literature on the public streets and sidewalks adjacent to Lloyd Center.[83]

Notwithstanding the Court's attempts to distinguish these earlier cases meaningfully, it is difficult to explain the Court's movement from *Logan Valley* to *Lloyd* as anything other than a deliberate doctrinal shift to insulate private speech regulation from First Amendment scrutiny. The Court's subsequent First Amendment state action cases, including *Hudgens v. National Labor Relations Board*,[84] more definitively narrowed the exceptions left open by *Lloyd* to all but preclude First Amendment protection for speech subject to regulation by private actors.

2. Private Regulation of Speech in Cyberspace

Individuals whose speech has been restricted by private Internet actors have sought to extend the state action doctrine as articulated in *Marsh* and *Logan Valley* (and subsequently honed in *Lloyd*) to private Internet actors, and have attempted to subject such online speech regulations to First Amendment scrutiny. These lawsuits involve challenges to Internet actors' refusal to deliver e-mails, removal of content posted on websites, and termination of individuals' accounts as a penalty for speech infractions.

Several of these cases involve Cyber Promotions, Inc., a company specializing in the dissemination of unsolicited e-mail. In *Cyber Promotions, Inc. v. America Online, Inc.*,[85] Cyber Promotions sought an injunction to prevent AOL from blocking messages it was attempting to send to AOL e-mail addresses. Cyber Promotions argued that AOL had opened its network to the public and devoted a portion of its property to public use by providing Internet e-mail services and acting as the sole conduit to its members' Internet e-mail boxes.[86] Cyber Promotions concluded that AOL had thereby effectively established a public forum in which its speech regulations were subject to First Amendment scrutiny.[87]

---

81. *Id.* at 563.
82. *Id.* at 563-64.
83. *Id.* at 566-67.
84. 424 U.S. 507, 518 (1976) (explaining that "the rationale of *Logan Valley* did not survive the Court's decision in the *Lloyd* case").
85. 948 F. Supp. 436 (E.D. Pa. 1996).
86. *Id.* at 450.
87. *Id.*

In analyzing whether AOL was a state actor for purposes of the First Amendment, the district court adopted a three-factor state action analysis gleaned from earlier Supreme Court cases, which required it to consider (1) whether AOL assumed a "traditional public function" in undertaking to perform the conduct at issue; (2) whether an elaborate financial or regulatory nexus existed between AOL's challenged conduct and the state; and/or (3) whether a symbiotic relationship existed between AOL and the state.[88] As to whether AOL "has exercised powers that are traditionally the exclusive prerogative of the state," the court answered in the negative, resting its analysis in part on the (simplistic) observation that the provision of access to the Internet through an e-mail system did not constitute the exercise of a public service traditionally exercised by the state.[89]

Regarding the "exclusive public function" prong of the state action test, the court once again readily concluded that AOL's provisions of e-mail service to its subscribers did not constitute a traditional exclusive public function.[90] In contrast with the private property owner in *Marsh*, in which "the owner of the company town was performing the full spectrum of municipal powers and stood in the shoes of the State,"[91] AOL exercised no powers that were the exclusive prerogative of the State. Cyber Promotions contested this point, citing *Logan Valley* and *Lloyd* and arguing that AOL's provision of Internet e-mail service constituted an exclusive public function because there were no alternative avenues of communication available to Cyber Promotions to disseminate its message to AOL members via e-mail.[92] The court rejected this argument, finding that Cyber Promotions had other means available to reach AOL members, including U.S. mail, telemarketing, television, cable, newspapers, magazines, and leafleting.[93] The *Cyber Promotions* court thus essentially concluded that the only alternative avenues of expression for Cyber Promotions to reach its desired audience of AOL members would be non-Internet channels.

The Supreme Court has recently clarified that (at least when the government is regulating) the relevant constitutional inquiry is whether the speech regulation at issue leaves open alternative avenues of expression within the speaker's chosen medium of expression. This issue was confronted by the Court in the *Reno v. ACLU* decision,[94] in which the Court

88. *Id.* at 447 n.2.
89. *Id.* at 451.
90. *Id.* at 452.
91. *See* Lloyd Corp. v. Tanner, 407 U.S. 551, 569 (1972).
92. *Cyber Promotions*, 948 F. Supp. at 451-52.
93. *Id.*
94. 521 U.S. 844 (1997).

reviewed the constitutionality of a statute prohibiting "indecent" and "patently offensive" communications on the Internet.[95] Part of the government's argument for the statute's constitutionality was that, even though the statute proscribed certain types of speech on the Internet, there were ample real space avenues of communication available for speakers. The Supreme Court rejected this argument, explaining that it would foreclose an entire medium of communication from constitutional protection.[96] To assess properly whether alternative avenues of expression exist for purposes of determining whether to subject private speech regulation to First Amendment scrutiny, courts should look to whether there are adequate alternative avenues of communication within the Internet medium itself for the speaker to communicate her message. If the speech regulation at issue fails to leave open such avenues on the Internet, the regulation—whether imposed by public or private hands—should be held to fail First Amendment scrutiny.[97]

In short, the Supreme Court's First Amendment jurisprudence makes clear that the relevant inquiry into "adequate alternative means of expression" turns on whether such alternative means exist within the speaker's chosen medium of expression. To conduct the inquiry otherwise would be tantamount to foreclosing an entire medium of expression to the speaker. Because the *Cyber Promotions* court looked to non-Internet media to con-

---

95. *See, e.g.*, Dawn C. Nunziato, *Toward a Constitutional Regulation of Minors' Access to Harmful Internet Speech*, 79 CHI.-KENT L. REV. 121 (2004).

96. *Reno*, 521 U.S. at 879-80.

97. Several Supreme Court Justices reiterated this interpretation of the alternative avenues of communication analysis in their ruling on the constitutionality of the CDA's successor statute, the Child Online Protection Act. Pub. L. No. 105-277, § 1401, 112 Stat. 2681-736 (1998). In *Ashcroft v. ACLU*, 535 U.S. 564 (2002), the Supreme Court considered the constitutionality of a statute criminalizing the dissemination of expression over the Internet that is obscene for minors, where obscenity was measured by reference to "local community standard." Because community standards on such matters may vary from one locality to another, in order to avoid liability under the statute, Internet speakers would be required either to tailor their expression to the most puritanical community's standards, or else abandon the Internet (with its heightened speech restrictions) as a vehicle for expression. *Ashcroft*, 535 U.S. at 593 (Kennedy, J., concurring). In defending this statute, the government once again argued that, despite these restrictions, the statute left individuals with ample alternative avenues of expression under *Logan Valley* and *Lloyd* because speakers of such potentially restricted expression could turn to non-Internet mediums to express themselves. Justice Kennedy, joined in his concurrence by Justices Souter and Ginsburg, once again rejected this argument, stating that "it is no answer to say that the speaker should take the simple step of utilizing a different medium," and explaining that "our prior decisions have voiced particular concern with laws that foreclose an entire medium of expression." *Ashcroft*, 535 U.S. at 596 (Kennedy, J., concurring).

clude that alternative avenues of expression existed for Cyber Promotions to reach its intended audience, its analysis was flawed.

The application of the state action doctrine to subject private Internet speech regulation to First Amendment scrutiny has been rejected in other contexts as well. In *Intel Corp. v. Hamidi*, Kenneth Hamidi, a disgruntled former employee of Intel, sent a series of e-mails critical of Intel's employment practices to several thousand Intel employees using their Intel e-mail addresses.[98] Intel, not wanting its employees to receive critical information about it, asserted its private property rights in its e-mail servers, and claimed that Hamidi, by sending such e-mails, was trespassing upon its personal property.[99] Among his other defenses, Hamidi asserted a First Amendment defense, claiming that he had a free speech right to send such an e-mail and that Intel's maintenance of an e-mail system connected to the Internet subjected Intel's speech regulations to First Amendment scrutiny.

The lower court rejected Hamidi's First Amendment defense to Intel's trespass to chattels claim, holding that Intel's e-mail servers were its private property and that "Intel is as much entitled to control its e-mail system as it is to guard its factories and hallways."[100] Although Hamidi ultimately prevailed on property law grounds,[101] the First Amendment issues in this case merit closer inspection.[102] In particular, the analysis of Hamidi's First Amendment claims by the Justices of the California Supreme Court reveal some common misperceptions regarding channels of communication that are available for individuals to express themselves on the Internet. Judge Mosk in his dissenting opinion, for example, took pains to criticize Hamidi's First Amendment defense on the grounds that Hamidi's expression occurred on a "private, proprietary intranet," and not within "the public commons of the Internet."[103] Mosk further explained, "Hamidi is not communicating in the equivalent of a town square or a . . . mailing through the United States Postal Service [but is rather] crossing from the public Internet into a private intranet."[104]

Justice Mosk's analysis embodies a common misperception about the existence of public spaces and forums on the Internet—that there exists

98. 71 P.3d 296, 301 (Cal. 2003).
99. *Id.* at 301-02.
100. Intel Corp. v. Hamidi, 114 Cal. Rptr. 2d 244, 257 (Ct. App. 2001).
101. *Hamidi*, 71 P.3d at 296.
102. *See, e.g.*, Dan Hunter, *Cyberspace as Place and the Tragedy of the Digital Anticommons*, 91 CALIF. L. REV. 439 (2003).
103. *Hamidi*, 71 P.3d at 326 (Mosk, J., dissenting).
104. *Id.* (Mosk, J., dissenting).

some sort of "public commons on the Internet" and indeed that the functional "equivalent of a town square" exists somewhere on the Internet. If Hamidi had gone there to express his message, Mosk suggests, his speech would have been protected by the First Amendment. The problem, however, is that there is no there there. Given the government's retreat from ownership and control of the Internet's infrastructure and the component spaces of the Internet, and given the overwhelming private ownership of expressive forums on the Internet, no such "public commons" or "town square" equivalents exist. Virtually all speech on the Internet is subject to the same type of private regulation as Intel's "private, proprietary intranet." In the same way that Intel's speech regulations on its servers are immune from First Amendment scrutiny, so too are other regulations of speech at the hands of private entities.

In addition to ceding control over the forums for speech to private entities, the U.S. government has ceded control over certain gateways for expression to private entities. As I explore at greater length elsewhere,[105] the government in 1998 ceded control over the Internet domain name system[106] to a nominally private entity, the Internet Corporation for Assigned Names and Numbers (ICANN). Domain names are the names that uniquely correspond to Internet protocol addresses assigned to computers to enable them to be connected to the Internet. Management of the domain name system translates into management of the gateways of communication via the Internet.

Prior to transferring control of the domain name system to ICANN, the United States vested a private corporation, Network Solutions, Inc. (NSI), with exclusive control over the registration of domain names. NSI, in turn, exercised this control by refusing to register certain domain names that it deemed inappropriate.[107] Its refusal to register certain domain names was challenged in several instances by the entities seeking to use such domain names. In one such case, an brought suit challenging NSI's refusal to register certain domain names containing sexually-themed words protected by

---

105. Dawn C. Nunziato, *Freedom of Expression, Democratic Norms, and Internet Governance*, 52 EMORY L.J. 187 (2003).

106. *See id.* at 192.

107. According to its policy in effect at the time, Network Solutions claimed that it

> has a right founded in the First Amendment to the Constitution to refuse to register, and thereby publish, on the Internet registry of domain names words that it deems to be inappropriate [because under the First Amendment] no corporation can be compelled to engage in publication which the corporation finds to be inappropriate.

See Island Online, Inc. v. Network Solutions, Inc., 119 F.Supp. 2d 289, 291 (E.D.N.Y. 2000).

the First Amendment. The disappointed domain name registrant, Island Online, claimed that NSI was a state actor and that NSI's refusal to register constitutionally-protected terms as domain names violated the First Amendment.[108] The court rejected Island Online's First Amendment claims, holding, inter alia, that the function of registering domain names did not constitute an "exclusive public function" and therefore the performance of this function did not render NSI a state actor. Relying once again on the fact that Internet-related functions are not in any sense "traditional" public functions, the court explained:

> Although the [U.S. government] was an instrumental agent in the Internet's origins, the Internet is by no stretch of the imagination a traditional and exclusive public function. For most of its history, its growth and development have been nurtured by and realized through private action. Moreover, registration of Internet domain names, the focal point of this case, has never been a public function at all.[109]

The court accordingly rejected Island Online's argument that NSI's conduct satisfied the exclusive public function test—or any of the other tests—for establishing that NSI's speech-restrictive actions constituted state action. It therefore concluded that NSI's content-based domain name registration policy was "purely private conduct" that was immune from scrutiny under the First Amendment.

The court in *National A-1 Advertising Inc. v. Network Solutions, Inc.* further analyzed whether NSI's speech-restrictive decisions were subject to First Amendment scrutiny.[110] In that case, NSI once again refused to register as domain names certain sexually-oriented (but First Amendment-protected) terms. The court first considered and rejected the argument that NSI was a state actor because it was performing a traditional state function by registering Internet domain names. Undertaking a more sophisticated analysis than the *Island Online* court, this court observed that although the "tradition" of serving as a registrar of domain names was not a long one, it was indeed one that had been performed and overseen by the government since the Internet's inception.[111] The court stated, however, that mere performance of a public function by itself was insufficient to qualify an entity as a state actor; rather, it must be further shown that the function is one

108. *See id.* at 296.
109. *Id.* at 306.
110. 121 F. Supp. 2d 156 (D.N.H. 2000).
111. *Id.* at 167.

that is exclusively reserved by the state.[112] This test, the court explained, is designed to "flush out a state's attempt to evade its responsibilities by delegating them to private entities."[113] The court held that NSI failed this portion of the test for establishing state action. The court went on to hold that NSI did not satisfy the second prong of the state action test because the government did not exercise coercive power or provide encouragement such that the actions at issue (that is, the refusal to register the desired domain names) could be deemed to be the conduct of the government.[114] Finally, the court concluded that NSI failed the third prong of the state action test because the relationship between the government and NSI could not properly be viewed as "symbiotic."[115] Notwithstanding the existence of a cooperative agreement between the government and NSI regarding NSI's role as a domain name registrar, the court found that the government was not fairly considered a "joint participant" in the challenged conduct.

NSI defended its conduct in the above cases not only on the grounds that it was a private actor and that its speech-restrictive decisions were therefore immune from First Amendment scrutiny, but also on the grounds that as a private actor it enjoyed the right not to sponsor (or be compelled to express) speech with which it disagreed.[116] Indeed, private Internet entities like NSI have increasingly wielded the First Amendment not only as a shield to insulate themselves from First Amendment liability for their speech regulations, but also as a sword to claim First Amendment protection as speakers and publishers for their exercise of "editorial" discretion. Invoking a line of cases beginning with *Miami Herald v. Tornillo*,[117] in which the Supreme Court upheld a newspaper's right not to be compelled to publish right-of-reply speech that was not of its choosing, private Internet actors like NSI have asserted that their content-based decisions are in furtherance of their First Amendment rights as speakers and publishers.[118]

---

112. *Id.* at 167-68.

113. *Id.* (quoting Barrios-Velazquez v. Asociacion de Empleados del Estado Libre Asociado de Puerto Rico, 84 F.3d 487, 494 (1996)).

114. *Id.* at 168.

115. *Id.* at 169.

116. *Id.* at 166.

117. 418 U.S. 241 (1974).

118. *See, e.g.*, Comcast Cablevision of Broward County, Inc. v. Broward County, 124 F. Supp. 2d 685 (S.D. Fla. 2000) (holding unconstitutional a county's requirement that cable television systems offering high-speed Internet service must allow competitors equal access to their broadband systems because it deprives cable operators of editorial discretion over their programming and forces cable operators to alter their content to conform to an agenda they do not set).

In *Island Online*, NSI claimed that it "has a right founded in the First Amendment of the Constitution to refuse to register, and thereby publish, . . . words that it deems to be inappropriate [because under the First Amendment] no corporation can be compelled to engage in publication which the corporation finds to be inappropriate."[119]

This First Amendment argument has been wielded in contexts far afield from its original domain of protecting speakers and publishers from being compelled to adopt speech with which they disagree. For example, Internet pipeline providers have successfully asserted this First Amendment argument to fend off governmental attempts to provide competitors with equal access to their pipelines. In *Comcast Cablevision of Broward County, Inc. v. Broward County*, Comcast challenged the county's "equal access" regulation, which required cable television systems offering high-speed Internet service to allow competitors equal access to their broadband systems.[120] Comcast successfully argued that, like publishers and speakers, it enjoyed editorial discretion over its programming and that the government requirement of equal (or "forced") access would violate its First Amendment rights to host only the content of its choosing.[121] Accordingly, the First Amendment has not only insulated private speech regulators from First Amendment scrutiny; it has also affirmatively protected their "editorial" decisions regarding which content to publish within their Internet places.

In sum, courts have resoundingly concluded that private entities' regulation of speech on the Internet does not constitute state action and that such private speech regulation is wholly immune from First Amendment scrutiny. Consistent with Congress's intent (as embodied in Section 230 of the Communications Decency Act) to turn the reins of Internet speech regulation over to private entities, private Internet actors have been allowed to wield substantial control over Internet expression, wholly unchecked by the First Amendment.

---

119. *See* Island Online, Inc. v. Network Solutions, Inc., 119 F. Supp. 2d 289, 296 (E.D.N.Y. 2000).

120. 124 F. Supp. 2d at 685.

121. *Id.* at 693-94.

III. PUBLIC OWNERSHIP OF SPEECH FORUMS AND THE PUBLIC FORUM DOCTRINE

A. Negative Versus Affirmative Conceptions of the First Amendment

The Internet indeed provides an unprecedented forum—or more accurately, an amalgam of forums—for expression. The overwhelming majority of these forums, however, are private, and accordingly, decisions regarding speech regulation are wholly immune from First Amendment scrutiny. Moreover, as will be discussed in Part V, even the comparatively insignificant publicly-owned forums for Internet speech have not been held to the stringent First Amendment standards applicable to speech regulations within genuine public forums.

According to one theory of Internet speech—the Net libertarian school that informed the privatization of the Internet in the first place—it is this very privatization of speech forums that best advances free speech values on the Internet. Net libertarians, such as Professor Richard Epstein,[122] claim that the primary purpose of the First Amendment is to insulate private individuals' speech decisions from government interference. Accordingly, such theorists claim that a free market for speech and for regulations of speech fully and completely serves the goals of freedom of expression, absent any government involvement to protect speech. Under this Net libertarian view, if there are low barriers to entry in the speech market, then the speech-protective goals of the First Amendment will be perfectly advanced by the aggregation of private forums and private speech decisions within these forums.[123] Affirmative government involvement in the market for speech in general—and in creating and maintaining public forums for speech in particular—will be unnecessary.

A competing school of thought conceptualizes the First Amendment as encompassing an affirmative, speech-protective role for the government, beyond the negative role of strictly scrutinizing governmental regulation of speech. According to this affirmative conception, the free speech values embodied in the First Amendment cannot be advanced solely by allowing private property owners free rein to determine what speech to permit and what speech to restrict within their property. As Cass Sunstein contends, a well-functioning system of free expression "is not intended to aggregate existing private preferences," but rather must incorporate certain collective values, values that will not necessarily be realized in an unregulated mar-

122. *See, e.g.*, Richard A. Epstein, *Cybertrespass*, 70 U. CHI. L. REV. 73 (2003).
123. *See, e.g.*, *id.*

ket for speech.[124] Sunstein explains that a free market for free speech could generate a range of serious problems. If the allocation of speech rights was decided through an ordinary pricing system, such a system would fail to incorporate certain speech-regarding values, and would ensure that dissident and other disfavored or unpopular speech would be foreclosed.[125] The affirmative conception of the First Amendment requires the government's involvement in the market for free speech to establish conditions allowing each citizen to exercise meaningfully his or her right to freedom of expression, a right that is integral to our system of democratic self-government. In the words of Professor Laurence Tribe, another advocate of the affirmative conception of the First Amendment, "it is not enough for the government to refrain from invading certain areas of liberty. The State may, even at some cost to the public fisc, be required to provide at least a minimally adequate opportunity for the exercise of certain freedoms."[126]

B. The Public Forum Doctrine

1. Development of the Doctrine

This affirmative conception of the First Amendment finds its clearest judicial expression in the development of the public forum doctrine,[127] under which courts impose on the government the affirmative obligation to dedicate certain publicly-held property for the use and benefit of individuals seeking to exercise their free speech rights. Prior to *Hague v. CIO*,[128] in which the Supreme Court first adopted the public forum doctrine, the government as a property owner enjoyed the same unfettered discretion as private property owners to regulate speech on its property. The Supreme Court articulated this position in *Davis v. Massachusetts*.[129] *Davis* involved a First Amendment challenge to government-imposed speech regulations on the (publicly-owned) Boston Common. The Court rejected this constitutional challenge and held that "there was no right in the plaintiff

124. CASS SUNSTEIN, DEMOCRACY AND THE PROBLEM OF FREE SPEECH 18 (1993).

125. *Id.* at 57-58.

126. LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW 964 (2d ed. 1988).

127. *See, e.g.*, Richard A. Posner, *Free Speech in an Economic Perspective*, 20 SUFFOLK U. L. REV. 1, 52 (1986) ("Until fairly recently it was assumed that the purpose of the first amendment was the negative one of preventing undue interference with private markets in ideas rather than the positive one of promoting the effective functioning of such markets. The 'public forum' doctrine, however, requires the government in some cases to make public facilities available for persons wanting to express themselves.").

128. 307 U.S. 496 (1939) (plurality).

129. 167 U.S. 43 (1897).

. . . to use the common except in such mode and subject to such regulations as the legislature in its wisdom may have deemed proper to prescribe."[130] Accordingly, prior to the Supreme Court's adoption of the public forum doctrine, the government as private property owner enjoyed unfettered discretion to restrict speech on its property, free from any First Amendment scrutiny.

In *Hague v. CIO*, the Supreme Court rejected the approach articulated in *Davis* and held that the government did not enjoy the same unfettered discretion as a private property owner to regulate speech on its property.[131] Rather, the Court imposed on the government the requirement that it accord the widest possible latitude to speech within property that constituted a "public forum" such as public parks, sidewalks, and streets.[132] These public forums are the places in which individuals are guaranteed not just the right but the meaningful opportunity to express themselves. Accordingly, under the public forum doctrine, the government serves as the guarantor of citizens' free speech rights.

Not all government-owned property enjoys public forum status. Property such as offices within government-owned office buildings, state prisons, and the like are not held open by the government for members of the public for expressive purposes.[133] But, within government-owned property that is deemed a "public forum," all speakers are permitted to express themselves on whatever viewpoints and whatever subjects they choose (as long as those subjects fall within the general parameters of speech for which the forum was designated).[134] It is within these public forums that speakers enjoy the fullest and broadest First Amendment protection. Thus, within public forums, the government must permit all manner of speech within the scope of the First Amendment's protection, regardless of the speakers' viewpoint or the content of such speech.[135]

Government restrictions on speech within such forums are also subject to the strictest judicial scrutiny. Accordingly, the ability to express oneself within a public forum—where regulation of the viewpoint or content of

---

130. *Id.* at 47.

131. 307 U.S. at 514-16.

132. *Id.*

133. *See, e.g.*, Perry Educ. Ass'n v. Perry Local Educators' Ass'n, 460 U.S. 37, 46 (1983).

134. As I explain later, within a designated public forum devoted to particular subjects, however, the government may impose restrictions limiting expression to the particular subject matter(s) for which the forum is designated. *See* text accompanying notes 167-70.

135. *See Perry*, 460 U.S. at 45.

speech is substantially prohibited[136]—is among the most important components of the First Amendment's protection for free speech. As Stephen Gey explains:

> The public forum doctrine . . . derives from the most basic mythological image of free speech: an agitated but eloquent speaker standing on a soap box at Speakers' Corner, railing against injustices committed by the government, whose agents are powerless to keep the audience from hearing the speaker's damning word. . . . .
>
> . . . .
>
> . . . [T]he essential reality grasped by the public forum doctrine remains as valid today as it was when thousands of Socialists packed into Union Square in the early days of [the twentieth] century to hang on every word of great progressive orators such as Eugene Debs. The larger reality behind the myth of the debate on the public street-corner is that every culture must have venues in which citizens can confront each other's ideas and ways of thinking about the world. Without such a place, a pluralistic culture inevitably becomes Balkanized into factions that not only cannot come to agreement about the Common Good, but also will not even know enough about other subcultures within the society to engage effectively in the deal-making and horse-trading that is the key to every modern manifestation of democratic government.[137]

Thus, while private property owners enjoy unfettered discretion to regulate or censor speech on any grounds whatsoever, under the public forum doc-

---

136. *Id.*

137. Steven G. Gey, *Reopening the Public Forum—From Sidewalks to Cyberspace*, 58 OHIO ST. L.J. 1535, 1538-39 (1998). The availability of centrally-located forums in which individuals can express themselves and where members of the general public are—willingly or unwillingly—exposed to such expression has been central to freedom of expression, and to democratic self-government, from time immemorial. The Supreme Court emphasized in *Hague* that

> [S]treets and parks . . . have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions. Such use of the streets and public places has, from ancient times, been a part of the privileges, immunities, rights, and liberties of citizens.

307 U.S. 496, 515 (1939) (Roberts, J., plurality). Because of the important function that places such as streets and parks have served in facilitating the exchange of ideas and expression in democracies, the public forum doctrine imposes upon the government the obligation to preserve such places for free expression and communication.

trine, government actors are substantially restrained in their ability to regulate speech within their "property" when that property constitutes a public forum.

In sum, in its foundational public forum decisions, the Supreme Court made clear that streets, sidewalks, parks, and similar places that are open to the public and conducive to open and free expression are to be dedicated to the public as forums for individuals to exercise their free speech rights. In real space, the requirement imposed upon the government to preserve such places for free and open expression serves as a critical safeguard of First Amendment rights. Absent the public forum doctrine, individuals would be restricted to expressing themselves on their own private property or on property owned by others only if they could convince these other property owners to permit such speech on their property.

The existence of public forums ameliorates the effect of economic disparities in property ownership upon individuals' right to free speech. Along these lines, the Supreme Court has demonstrated special solicitude in its public forum jurisprudence for the free speech rights of poorly-financed speakers and causes.[138] The public forum doctrine imposes upon the government the obligation to facilitate and subsidize the speech of poorly-financed speakers, by granting them an effective forum from which to express themselves.[139]

Given that property in real space generally consists of a mix of public and private forums, and given that most places in the United States contain centrally-located public forums like public streets and parks, in real space all speakers—however poorly financed or unpopular their cause—are guaranteed an effective, centrally-located forum from which to express their views and to reach a broad general audience. Accordingly, the mandate that the government dedicate such centrally-located public places to the public for free speech purposes provides a crucial safety valve for free expression. By granting poorly-financed and unpropertied speakers access to centrally located public property, public forums enable individuals to express themselves effectively to broad audiences.

The required existence of public forums advances free speech interests in another important manner. By granting speakers access to interstitial public forums—such as streets and sidewalks located adjacent to private

---

138. *See* Molly S. Van Houweling, *Distributive Values in Copyright*, 83 TEX. L. REV. (forthcoming 2005).

139. Professor Jack Balkin has elucidated the subsidization function of the public forum doctrine. *See* J.M. Balkin, *Some Realism About Pluralism: Legal Realist Approaches to the First Amendment*, 1990 DUKE L.J. 375, 400.

property—public forums enable individuals effectively to target specific private property owners. Interstitial public forums serve a different purpose than centrally-located public forums like parks. Interstitial public forums enable individuals to target specific private property owners by providing a forum from which individuals can address the precise targets of their speech.[140] For example, individuals who wish to criticize a company's employment practices are ensured the right under the public forum doctrine to protest on the public streets and sidewalks adjacent to that company's headquarters. Absent such interstitial public forums, individuals could not effectively criticize private property owners because the targeted entities would refuse to allow speech critical of them on or near their property. Given the characteristics of our real space landscape, individuals accessing private property must generally pass through interstitial public forums and along the way may be subject to hearing speech that they might otherwise choose to avoid. The existence of interstitial public forums requires listeners to be confronted with expression that they might otherwise choose to avoid and prevents individuals from exercising perfect control over which expression will reach them.[141] If an individual chooses to shop at a non-union grocery store, and desires to insulate herself from criticisms of the store's employment practices, for example, the existence of interstitial public forums requires that the individual be confronted with such speech nonetheless.[142]

In short, public forums (and public ownership of property, the prerequisite for public forums) serve important roles in facilitating freedom of expression in real space. The existence of publicly-owned property and the scrutiny imposed on regulations of speech within such property are critical to the protection of First Amendment rights.

---

140. *See generally* Noah D. Zatz, Note, *Sidewalks in Cyberspace: Making Space for Public Forums in the Electronic Environment*, 12 HARV. J.L. & TECH. 149, 151-52 (1998) (discussing the difference between "general access" and "specific access" public forums and advocating the creation of "sidewalks in cyberspace" so as to "enable ordinary citizens to engage one another as they move between the places where they conduct their affairs").

141. Professor Cass Sunstein, for one, has criticized the use of the Internet to tailor and control all the information an individual receives. *See* CASS SUNSTEIN, REPUBLIC.COM 3 (2001) (explaining that, with the advent of advanced Internet search and filtering capabilities, one will "need not come across topics and views that one has not sought out . . . and will be able to see exactly what one wants to see, no more and no less").

142. *See* Amalgamated Food Employees Union Local 590 v. Logan Valley Plaza, Inc., 391 U.S. 308 (1968).

2. Categorization of Forums

Although the justifications for public forums and the role public forums play in our democratic system are compelling and straightforward, the development of the public forum doctrine has become quite complex in recent years. Since the inception of this doctrine, the Supreme Court has rendered the doctrine intricate, complex, and rather convoluted. While this Article primarily analyzes the roles that so-called traditional public forums like streets, sidewalks, and parks serve within our system of democratic self-government, a brief foray into the Court's complex, trifurcated analysis of government-owned forums may prove helpful. This case law breaks out forums into the following three categories: (1) traditional public forums; (2) designated public forums; and (3) non-public forums.

"Traditional" public forums consist of streets, sidewalks, parks, and other places that "have immemorially been held in trust for the use of the public and, time out of mind, have been used for purposes of assembly, communicating thoughts between citizens, and discussing public questions."[143] "Designated public forums" consist of public property that has not "immemorially" been used as a forum for expression, but which the government has explicitly opened and designated as a place for expressive activity by the public.[144] The government may choose, for example, to open up property within a public school,[145] university meeting facilities[146] or municipal theaters,[147] as forums for expression generally or for expression on certain designated subjects. Within a limited-purpose designated public forum, once the government has defined the subject matter limitations of the forum, regulation of such property is subject to the same limitations as those governing a traditional public forum.[148] Thus, both within traditional public forums and designated public forums, individuals enjoy their most robust rights of free expression. Government restrictions on speech within such public forums are subject to the most stringent scrutiny under the First Amendment such that no speech restrictions will be upheld unless they serve compelling government interests and are the least restrictive means of restricting such speech.

---

143. Hague v. CIO, 307 U.S. 496, 515 (1939) (Roberts, J., plurality).

144. *See* Perry Educ. Ass'n v. Perry Local Educators' Ass'n, 460 U.S. 37, 45 (1983).

145. City of Madison Joint Sch. Dist. v. Wisc. Employment Relations Comm'n, 429 U.S. 167 (1976).

146. Widmar v. Vincent, 454 U.S. 263 (1981).

147. S.E. Promotions, Ltd. v. Conrad, 420 U.S. 546 (1975).

148. *See* Int'l Soc'y for Krishna Consciousness, Inc. v. Lee, 505 U.S. 672, 678 (1992).

The third category of publicly-owned forums is non-public forums: places like military bases, jail grounds, and federal workplaces, that the government owns but which it has not opened up for expressive activity on the part of the public.[149] Furthermore, the Supreme Court has made clear that, to constitute a public forum, the place in which speech occurs need not be an actual physical place. Rather, public forums may also include virtual forums, like funding and solicitation schemes,[150] the airwaves,[151] and cable television.[152]

The classification of a forum into one type of forum or another is all but dispositive of the First Amendment challenge. If a forum is deemed to fall within the traditional or designated public forum category, courts will apply strict scrutiny to content-based regulations of speech within the forum and will almost certainly strike down such regulations. Regulations of speech within a non-public forum, on the other hand, are subject to reduced scrutiny and will most likely withstand constitutional challenge. How courts classify speech forums on the Internet thus becomes a critical factor in the extension of First Amendment protections to speech in cyberspace.

IV. COURTS' REFUSAL TO SUBJECT PUBLIC INTERNET ACTORS TO STRINGENT FIRST AMENDMENT STANDARDS

Public forums embody the government's guarantee that citizens will enjoy meaningful free speech rights. Yet, as we have also seen, the vast majority of speech forums in cyberspace are privately owned and privately regulated, with the consequence that virtually no public forums exist, and speech regulations within private forums are immune from First Amendment scrutiny. A small fraction of Internet forums for speech, however, are government-owned. One might suppose that such publicly-owned Internet spaces would be deemed public forums: places where individuals could enjoy their free speech rights most fully and the constitutional guarantee of free expression could be rendered meaningful.

In several recent challenges to speech regulations imposed by government within public Internet spaces, however, courts—including the Su-

149. *Perry* , 460 U.S. at 46.

150. *See, e.g.*, Rosenberger v. Rector & Visitors of the Univ. of Va., 515 U.S. 819 (1995).

151. *See, e.g.*, Ark. Educ. Television Comm'n v. Forbes, 523 U.S. 666 (1998).

152. *See, e.g.*, Denver Area Educ. Telecomm. Consortium, Inc. v. FCC, 518 U.S. 727 (1996).

preme Court—have concluded that such spaces are not public forums and that therefore governmental regulation of speech within these forums are immune from meaningful First Amendment scrutiny. Most notably, in *United States v. American Library Ass'n*,[153] the Supreme Court held that Internet access provided by public libraries did not constitute a public forum, and that speech restrictions imposed within such forums were therefore immune from meaningful First Amendment scrutiny.

In *American Library Ass'n*, plaintiffs challenged the constitutionality of the Children's Internet Protection Act (CIPA), which required that all public libraries that provide Internet access to their patrons must impose software filters upon such access, or else forgo substantial federal funding.[154] CIPA makes the use of software filters by public libraries and schools a condition on their receipt of two kinds of federal subsidies: grants under the Library Services and Technology Act (LSTA)[155] and "E-rate" discounts for Internet access and support under the Telecommunications Act.[156] To receive LSTA funds or E-rate discounts, CIPA essentially requires public libraries and schools to certify that they are using "technology protection measures" that prevent patrons from accessing visual depictions that are "obscene," "child pornography," or in the case of minors, "harmful to minors."[157] While CIPA's scheme allows library officials under certain circumstances to disable software filters for certain patrons engaged in bona fide research or other lawful purposes, the disabling of such filters on computers used by minors is prohibited if the library or school receives E-rate discounts.[158] In challenging CIPA's constitutional-

153. 539 U.S. 194 (2003) (plurality).

154. *Id.*

155. *See* Am. Library Ass'n v. United States, 201 F. Supp. 2d. 401 (E.D. Pa.) (three-judge court) [hereinafter *Am. Library Ass'n I*], *rev'd*, 539 U.S. 194 (2003) (plurality).

156. *See id.* at 408.

157. *Id.* at 407.

158. Thus, in order to receive E-rate discounts, libraries and schools must certify that, during any use of Internet-accessible computers by minors (those 16 and under), *id.* at 406, filtering technology is being used to block access to material that is obscene, child pornography, or deemed "harmful to minors." While the terms "obscene" and "child pornography" are given their (constitutionally acceptable) standard meaning, CIPA defines material that is "harmful to minors" as

> any picture, image, graphic image file, or other visual depiction that—(i) taken as a whole and with respect to minors, appeals to a prurient interest in nudity, sex, or excretion; (ii) depicts, describes, or represents, in a patently offensive way with respect to what is suitable for minors, an actual or simulated sexual act or sexual contact, actual or simulated normal or perverted sexual acts, or a lewd exhibition of the genitals;

ity, the American Library Association claimed that the required software filters imposed unconstitutional restrictions on their patrons' access to protected speech.

To understand the speech restrictions at issue in CIPA, it is important to understand the mechanics of software filtering.[159] Software filtering programs generally operate by comparing website addresses that a user wishes to access against a "blacklist." A typical filtering software program operates by examining various parts of an Internet address, or URL, against this internal blacklist to see if the URL is forbidden.[160] Prohibited sites may also be compared against separate exception lists or "whitelists." Some types of filtering software can be set so that everything not prohibited is permitted (blacklist only) or only that which is explicitly allowed is permitted (whitelist only). The software can also be designed to operate via some combination of blacklists and whitelists, with one list overriding another.[161]

The default blacklists and whitelists used by filtering software programs are created by those who design such software and constitute a substantial portion of the programs' value to consumers. As such, these lists are typically protected as trade secrets. Although the library may choose to configure the filtering software to filter out certain pre-defined categories of websites (such as "Adult/Sexually Explicit"), the library has no way of knowing the criteria used by the software developers to select which websites fall into this category, nor which websites will actually be found to fall within this category. Thus, a library implementing a filtering software program has no way of knowing which websites will actually be rendered inaccessible by the filtering software program.

> and (iii) taken as a whole, lacks serious literary, artistic, political, or scientific value as to minors.

Pub. L. No. 106-554 § 1721(c), 114 Stat. 2763A-335 (2000) (codified at 47 U.S.C. § 254(h)(7)(G) (2000)). With respect to adults' use of Internet-accessible computers, CIPA provides that a library official is permitted to "disable the technology protection measure concerned, during use by an adult, to enable access for bona fide research or other lawful purpose." 47 U.S.C. § 254(h)(5)(D). However, CIPA's amendments to the E-rate program do not permit libraries or schools to disable filters to enable bona fide research or other lawful use for minors. 47 U.S.C. § 254(h)(5)(A). CIPA's amendments to the Library Services and Technology Act (LSTA) condition funding under the LSTA upon parallel restrictions. 47 U.S.C. § 254(h)(5)(F).

159. My discussion of software filtering follows closely that provided by filtering experts Seth Finkelstein and Lee Tien in their extremely lucid article. Seth Finkelstein & Lee Tien, *Blacklisting Bytes*, *in* FILTERS & FREEDOM 2.0 (Electronic Privacy Info. Ctr. eds., 2001).

160. *Id.* at 67.

161. *Id.* at 67-68.

The constitutional challenge to CIPA was first heard by a special three-judge panel, which struck down the statute after a thorough analysis of the application of the public forum doctrine to the circumstances presented by this case. The court explained that under the public forum doctrine, "the extent to which the First Amendment permits the government to restrict speech on its own property depends on the character of the forum the government has created."[162] The threshold determination was whether libraries' provision of Internet access constituted a traditional public forum, a designated public forum of some type, or a non-public forum.[163] Because the category of traditional public forums appears to be limited to streets, sidewalks, public parks, and other such public places that have "immemorially been held in trust for the use of the public" for expressive purposes,[164] the court concluded that libraries' provision of Internet access did not fall within this category.

The court was then required to determine whether libraries' provision of Internet access constituted a "designated public forum," in which case the speech restrictions would be subject to strict First Amendment scrutiny, or a non-public forum, in which case strict scrutiny would not apply. The court distinguished libraries' provision of Internet access from other types of non-public forums (including military bases, jail grounds, and the federal workplace)[165] and found that the purpose of a public library's provision of Internet access is "for use by the public . . . for expressive activity, namely, the dissemination and receipt by the public of a wide range of information."[166] Accordingly, the court concluded that the government's provision of Internet access in a public library constituted a designated public forum.[167]

The court next considered the level of First Amendment scrutiny that was applicable to the speech regulations CIPA imposed within this designated public forum. It explained that if a very narrow range of speech was facilitated in the first place within the designated limited-purpose public forum at issue, then the government's restrictions of speech within such a forum would be accorded substantial deference.[168] As the Supreme Court

---

162. *Am. Library Ass'n I*, 201 F. Supp. 2d at 454.

163. *See supra* Part III.B.2.

164. *Am. Library Ass'n I*, 201 F. Supp. at 454-55.

165. *Id.* at 457.

166. *Id.*

167. *See also* Mainstream Loudoun v. Bd. of Trs. of Loudoun County Library, 24 F. Supp. 2d 552 (E.D. Va. 1998) (holding that government's provision of Internet access in a public library constitutes a designated public forum).

168. *Am. Library Ass'n I*, 201 F. Supp. 2d at 458.

explained by way of example on the related subject of government-subsidized speech, "[w]hen Congress established the National Endowment for Democracy to encourage other countries to adopt democratic principles, it was not constitutionally required to encourage competing lines of political philosophy such as communism and fascism."[169] Rather, only speech that was within the scope for which the forum was designated was permitted within that forum, and speech that fell outside of this designated range could be constitutionally excluded by the government. Conversely, the broader the range of speech the government facilitates within a designated public forum, the less deference the First Amendment accords to the government's content-based restrictions on the speech within that forum. Thus, "where the government creates a designated public forum to facilitate private speech representing a diverse range of viewpoints, the government's decision selectively to single out particular viewpoints for exclusion is subject to strict scrutiny."[170] The court concluded that libraries' provision of Internet access fell within the latter category of designated public forums—that is, those in which a broad range of expression was permitted and, concomitantly, those in which the government's speech regulations are subject to strict First Amendment scrutiny.

Adverting to the Supreme Court's decision in *Rosenberger v. Rector & Visitors of the University of Virginia*,[171] the district court in *American Library Ass'n* explained:

> [T]he more widely the state opens a forum for members of the public to speak on a variety of subjects and viewpoints, the more vulnerable is the state's decision selectively to exclude certain speech on the basis of its disfavored content, as such exclusions distort the marketplace of ideas that the state has created in establishing the forum.
>
> . . . [W]here the state designates a forum for expressive activity and opens the forum for speech by the public at large on a wide range of topics, strict scrutiny applies to restrictions that single out for exclusion from the forum particular speech whose content is disfavored. . . .
>
> . . . [T]o the extent that the government creates a public forum expressly designed to facilitate the dissemination of private speech, opens the forum to any member of the public to speak on virtually any topic, and then selectively targets certain speech for exclusion based on its content, the government is singling out

169. *Id.* (quoting Rust v. Sullivan, 500 U.S. 173, 194 (1991)).
170. *Id.* at 460.
171. 515 U.S. 819 (1995).

> speech in a manner that . . . [is subject to] heightened First Amendment scrutiny. . . .[172]

Applying the *Rosenberger* Court's analysis, the court explained that libraries' provision of Internet access to their patrons, unlike their provision of print materials, enables their patrons to receive speech on a "virtually unlimited number of topics, from a virtually unlimited number of speakers, without attempting to restrict patrons' access to speech that the library, in the exercise of its professional judgment, determines to be particularly valuable."[173] Because the libraries' provision of Internet access enables patrons to receive speech on a broad and diverse range of topics, the restrictions on sexually-themed expression imposed by mandatory software filters were subject to strict First Amendment scrutiny. Accordingly, the court found that the use of filtering software mandated by CIPA erroneously blocked a huge amount of speech that is protected by the First Amendment,[174] estimating the number of web pages erroneously blocked to be "at least tens of thousands."[175] The court observed that the government's expert himself found that popular filtering software packages overblock at rates between 6% and 15%,[176] that such programs inevitably overblock harmless Internet content, which adults and minors have a First Amendment right to access, and underblock obscene and child pornographic content, which neither adults nor minors have a First Amendment right to access.[177] The court also found that the provisions of CIPA permitting libraries to unblock wrongfully blocked sites upon the request of an adult[178] (or in some cases a minor)[179] who is engaged in "bona fide research or other lawful purpose[s]" were insufficient to render the statute constitutional.[180] The court concluded that, "[g]iven the substantial amount

---

172. *Am. Library Ass'n I*, 201 F. Supp. 2d at 461.

173. *Id.* at 462.

174. *See id.* at 406.

175. *Id.* at 449; *see id.* at 475 (finding that filtering software programs "block many thousands of Web pages that are clearly not harmful to minors, and many thousands more pages that, while possibly harmful to minors, are neither obscene nor child pornography").

176. *Id.* at 475-76. That is to say, the expert concluded that between 6% and 15% of blocked web pages contained no content that met even the software's own definitions of sexually-themed content, let alone the constitutional definitions of obscenity or child pornography. *Id.*

177. *Id.* at 475.

178. *Id.* at 484.

179. *Id.* at 485.

180. *Id.*; *see* 47 U.S.C. § 254(h)(6)(D) (2000). In addition to the constitutional infirmities inherent in refusing to permit libraries to unblock wrongly blocked sites for minors, the court found that many adult patrons were "reluctant or unwilling to ask librari-

of constitutionally protected speech blocked by the filters studied," CIPA failed strict scrutiny because it was not narrowly tailored to advance its compelling government interests.[181]

The Supreme Court reversed, holding that the restrictions CIPA required were not unconstitutional, primarily based on the Court's conclusion that these speech restrictions were not imposed within a public forum.[182] Chief Justice Rehnquist, who authored a plurality opinion in which Justices O'Connor, Scalia, and Thomas joined, held that the provision of Internet access in public libraries did not constitute a public forum and that strict scrutiny was therefore not the proper level of scrutiny for analyzing CIPA's constitutionality.[183] Rehnquist first explained that Internet access in public libraries did not constitute a "traditional public forum" within the constitutional meaning of that term because "this resource—which did not exist until quite recently—has not immemorially been held in trust for the use of the public [or], time out of mind, . . . been used for purposes of assembly, communication of thoughts between citizens, and discussing public questions."[184]

Rehnquist then explained that Internet access in public libraries did not constitute a "designated public forum," a forum with respect to which "the government [has made] an affirmative choice to open up its property for

---

ans to unblock Web pages or sites that contain only materials that might be deemed personal or embarrassing, even if they are not sexually explicit or pornographic." *Am. Library Ass'n I*, 201 F. Supp. 2d at 427. Because libraries were not required under CIPA's scheme to permit Internet users to make anonymous unblocking requests, the vast majority of patrons confronted with wrongfully blocked sites apparently decline to request the unblocking of such sites. *See Am. Library Ass'n I*, 201 F. Supp. 2d at 427. Furthermore, the court found that even where unblocking requests were submitted and acted upon, the unblocking process took too long—between twenty-four hours and one week. *Am. Library Ass'n I*, 201 F. Supp. 2d at 487. The court concluded that:

> The content-based burden that the library's use of software filters places on patrons' access to speech suffers from the same constitutional deficiencies as a complete ban on patrons' access to speech that was erroneously blocked by filters, since patrons will often be deterred from asking the library to unblock a site and patron requests cannot be immediately reviewed.

Am. Library Ass'n I, 201 F. Supp. 2d at 489.

181. *Am. Library Ass'n I*, 201 F. Supp. 2d at 476.

182. *Am. Library Ass'n II*, 539 U.S. 194, 205 (2003) (plurality).

183. *See id.* at 216-18 (Rehnquist, C.J., plurality). On this point, Rehnquist explained that "[w]e require the Government to employ the least restrictive means only when the forum is a public one and strict scrutiny applies." *Id.* at 207 n.3 (Rehnquist, C.J., plurality).

184. *Id.* at 205 (Rehnquist, C.J., plurality) (internal quotation marks omitted) (quoting Int'l Soc'y for Krishna Consciousness, Inc. v. Lee, 505 U.S. 672, 679 (1992)).

use as a public forum."[185] The Chief Justice found, with little elaboration, that "[a] public library does not acquire Internet terminals in order to create a public forum for Web publishers to express themselves, [but] . . . to facilitate research, learning, and recreational pursuits by furnishing materials of requisite and appropriate quality."[186] He observed further that "even if appellees had proffered more persuasive evidence that public libraries intended to create a forum for speech by connecting to the Internet, we would hesitate to import the public forum doctrine . . . wholesale into the context of the Internet."[187] Having concluded that libraries' provision of Internet access did not constitute a public forum, Rehnquist analyzed CIPA's constitutionality under a framework of reduced scrutiny, and merely inquired into whether libraries' use of filtering software was "reasonable," which he readily found that it was.[188]

Despite the fact that the libraries themselves contended that they had provided Internet access to their patrons to facilitate communication and exchange on a "virtually unlimited number of topics," Rehnquist declined to extend public forum status (either traditional or designated) to the Internet forum at issue and accordingly declined to extend meaningful scrutiny to the government's content-based exclusions from that forum effected by the statutorily mandated filters. The Court's refusal to accord public forum status to libraries' provision of Internet access establishes a dangerous, speech-restrictive precedent for the Internet. In this rare instance of public ownership and control over Internet speech forums, in which the public entity acknowledges that it created the forum to facilitate the exchange of ideas and communication among members of the public on a virtually unlimited number of topics, the Court nonetheless held that no public forum was involved and that speech restrictions within the forum were therefore immune from meaningful First Amendment scrutiny. If no pub-

185. *Id.* at 206 (Rehnquist, C.J., plurality).

186. *Id.* (Rehnquist, C.J., plurality).

187. *Id.* at 207 n.3 (Rehnquist, C.J., plurality) (internal quotation marks omitted).

188. *Id.* at 208 (Rehnquist, C.J., plurality) ("[I]t is entirely reasonable for public libraries to . . . exclude certain categories of content, without making individualized judgments that everything they do make available has requisite and appropriate quality."). This holding, in turn, was the predicate for the Supreme Court's ultimate holding that CIPA was a valid exercise of Congress's spending power and imposed no unconstitutional conditions upon libraries: "Because public libraries' use of Internet filtering software does not violate their patrons' First Amendment rights, CIPA does not induce libraries to violate the Constitution, and is a valid exercise of Congress' spending power. Nor does CIPA impose an unconstitutional condition on public libraries." *Id.* at 214 (Rehnquist, C.J., plurality).

lic forum for expression is found in these circumstances, it is unlikely that a public forum will ever be recognized in the context of the Internet.

The *American Library Ass'n* decision is not the first (and likely will not be the last) to refuse to extend meaningful First Amendment scrutiny to speech restrictions imposed by the government on government-owned computers. Several other cases involving government ownership of Internet-accessible computers have declined to subject speech restrictions within such forums to strict judicial scrutiny. The case of *Urofsky v. Gilmore* also involved government censorship of First Amendment protected speech on government-owned computers.[189] The challenged speech restrictions originated in 1998 when the Commonwealth of Virginia grew concerned about the use of public computers by its employees to access sexually-themed expression on the Internet. In an attempt to remedy this perceived problem, the legislature enacted the "Restrictions on State Employee Access to Information Infrastructure Act."[190] Several Virginia public college and university professors challenged the constitutionality of the Act,[191] which restricted the ability of hundreds of thousands of Virginia public employees to access sexually explicit (but constitutionally protected) content on computers that were owned or leased by the State. Under the statute, public employees were prohibited from accessing (without securing advanced written agency approval) sexually explicit content on the Internet, where "sexually explicit" was defined quite broadly to include, inter alia, "any . . . visual representation . . . depicting . . . a lewd exhibition of nudity."[192]

In ruling on the professors' First Amendment challenge, the district court held that the statute unconstitutionally restricted "the ability of more than 101,000 public employees at all levels of state government to read, research, and discuss sexually explicit topics within their areas of expertise, [including] inquiry and debate by academics in the fields of art, literature, medicine, psychology, anthropology, and law."[193] The statute further restricted the rights of members of the public to receive and benefit from the speech of state employees on matters within their expertise.

Because the Act restricted public employees' free speech rights, the court was required by First Amendment precedent to conduct its analysis under the special test crafted by the Supreme Court for evaluating the First

189. 216 F.3d 401 (4th Cir. 2000) (en banc).

190. Urofsky v. Allen, 995 F. Supp. 634, 635 (E.D. Va. 1998), *rev'd sub nom.* Urofsky v. Gilmore, 216 F.3d 401 (4th Cir. 2000) (en banc).

191. *Id.*

192. *Id.* at 635.

193. *Id.* at 638.

Amendment rights of public employees.[194] Because the government enjoys greater latitude to restrict the speech of its employees versus speech of members of the public generally, this test applies deferential scrutiny to restrictions of public employees' work-related expression. Under this test, set forth in the *Connick*[195] and *Pickering*[196] cases, courts must first consider whether the speech at issue is that of the employee as a private citizen speaking on a matter of public concern. If so, then the court must consider whether the employee's interest in her First Amendment expression outweighs her employer's interest in regulating such speech for the appropriate operation of the workplace. If, however, the court determines that the speech at issue is not that of an employee qua private citizen speaking on a matter of public concern, then the state, as employer, may regulate the speech without infringing any First Amendment protection.[197]

Applying this test, the district court held that the speech of Virginia state employees on sexually explicit topics includes speech on matters of public concern that is entitled to the fullest First Amendment protection under the required *Connick*/*Pickering* analysis. The court then held that the Act's restrictions were not properly tailored to address the harm that the government allegedly aimed to protect, and therefore that the Act was fatally over- and under-inclusive.[198]

The Fourth Circuit disagreed. Sitting en banc, the Fourth Circuit concluded, with little discussion, that the restricted speech at issue did not touch upon a matter of public concern and that the state, as employer, could therefore regulate it without infringing any First Amendment protection.[199] The Fourth Circuit held that the statute "does not affect speech by [the professors] in their capacity as private citizens speaking on matters of public concern [and that therefore] it does not infringe the First Amendment rights of state employees."[200] In short, even though the state-imposed restrictions on the Internet expression at issue in *Urofsky* applied to state-owned property, and even though we would expect that the First Amendment would have a meaningful role to play in holding in check such government restrictions on speech, the Fourth Circuit applied reduced scrutiny to the challenged speech restrictions.

---

194. *Id.* at 636.
195. Connick v. Myers, 461 U.S. 138 (1983).
196. Pickering v. Bd. of Educ., 391 U.S. 563 (1968).
197. *Urofsky*, 995 F. Supp. at 638.
198. *Id.* at 638-41.
199. Urofsky v. Gilmore, 216 F.3d 401 (4th Cir. 2000) (en banc).
200. *Id.* at 409.

Because private Internet actors—who have the power to regulate and suppress a vast amount of Internet speech—are not subject to any First Amendment scrutiny (much less the stringent scrutiny imposed under the public forum doctrine), and because after *American Library Ass'n*, even public owners of Internet speech forums will likely not be held to meaningful First Amendment scrutiny under the public forum doctrine, restrictions of Internet speech are essentially no longer constitutionally prohibited. Thus, the important functions served by the First Amendment generally—and by the public forum doctrine in particular—are in danger of being seriously eroded in cyberspace.

V. RESTORING THE VALUES OF THE PUBLIC FORUM WITHIN CYBERSPACE

The evisceration of meaningful First Amendment protection for Internet speech and the absence of public forums in cyberspace have important consequences. First, individuals who seek to express unpopular views will find it increasingly difficult to do so, absent general public forums within which viewpoint discrimination is constitutionally prohibited. Second, individuals who seek to criticize others will find it more difficult to do so, absent the functional equivalent of interstitial public forums on the Internet.

A. Restoring the Values of the General Public Forum Within Cyberspace

In real space, general public forums provide effective forums for individuals who wish to reach broad general audiences but who otherwise would be unable to compete within the marketplace for speech. Such speakers' inability (absent the government's intervention via the public forum doctrine) may arise from their lack of financial means or from the unpopularity of their message. Because the government, under the public forum doctrine, has an obligation to open up certain of its property for the use and benefit of all speakers, without regard to the content or viewpoint of their message, such speakers are guaranteed an opportunity to express their message effectively in real space.

Some might contend that no such subsidization of unpopular or poorly-financed speech is called for in cyberspace. Certainly, as discussed above, it is less expensive to express oneself through an Internet discussion forum, website, blog, or e-mail message than it is to engage in such expression in real space. And in the past, the Internet has generally been a hospitable forum for a broad range of expression. Yet, as private Internet actors become less hospitable to unpopular speech in their Internet places

and modify their terms of use to prohibit communication on unpopular (but First Amendment protected) subjects, it will become more difficult for individuals to secure the same type of speech protection in cyberspace that they enjoy in real space. Similarly, as public Internet actors—like public libraries throughout the United States—become less hospitable to unpopular speech within the Internet forums they control, the obstacles confronted by speakers of unpopular messages on the Internet will become formidable. To remedy this problem, we need to introduce spaces in which individuals' free speech is constitutionally protected instead of leaving the protection of free speech at the mercy of private speech regulators. Doing so will require either the courts or the legislature to act.

Several aspects of First Amendment doctrine must be reconceptualized in order for courts to introduce public forum values into cyberspace. First, courts need to reconceptualize the "traditional government function" component of the state action doctrine. Second, courts need to reconceptualize their analysis of "traditional public forums" within public forum jurisprudence. Because current tests require "traditional" actions and places in order for First Amendment scrutiny to apply to speech regulations, Internet-related actions and places *ipso facto* will be found to fall outside the protection of these doctrines. No action undertaken by a private entity regulating Internet speech will ever be deemed the performance of a "traditional state action" sufficient to subject such regulation to First Amendment scrutiny under the state action doctrine as it is currently conceived. No expressive forum on the Internet will ever be deemed a "traditional" public forum—one that has "immemorially" and "time out of mind" been held in trust for the use of the public for expressive purposes—under the public forum doctrine as it is currently conceived. As a result, speech in such forums will not enjoy the full measure of protection under the First Amendment. Courts should therefore rework the "traditionality" component of these First Amendment doctrines to incorporate a functional analysis of the places in which speech, and speech regulations, occur.

1. Reconceptualizing the "Traditionality" Component

Courts, notably the Supreme Court in *American Library Ass'n*, have interpreted the public forum analysis too parsimoniously and have placed undue emphasis on whether the forum at issue is a traditional one that has "immemorially" or "time out of mind" been used for purposes of assembly, communication of thoughts between citizens, and discussion of public questions.[201] This emphasis on traditionality and history led Chief Justice

201. *See Am. Library Ass'n II*, 539 U.S. 194, 204-05 (2003) (plurality).

Rehnquist to conclude in *American Library Ass'n* that libraries' provision of Internet access did not constitute a public forum because "this resource . . . did not exist until quite recently."[202]

Courts should reject such a simplistic analysis of public forums, which forecloses by its very terms the recognition of an Internet forum as a public forum for First Amendment purposes. Instead, courts should undertake a functional analysis to determine whether such places are currently widely used for purposes of "communication of thoughts between citizens, and discussing public questions"[203] and serve the same speech-facilitating purposes served in real space by public sidewalks and parks. The lower court's three-judge panel in *American Library Ass'n* is instructive in setting forth a reinterpretation of this aspect of the public forum analysis:

> Regulation of speech in streets, sidewalks, and parks is subject to the highest scrutiny not simply by virtue of history and tradition, but also because the speech-facilitating character of sidewalks and parks makes them distinctly deserving of First Amendment protection. Many of these same speech-promoting features of the traditional public forum appear in public libraries' provision of Internet access.
>
> . . . Just as important as the openness of a forum to listeners is its openness to speakers. Parks and sidewalks are paradigmatic loci of First Amendment values in large part because they permit speakers to communicate with a wide audience at low cost. . . . Similarly, given the existence of message boards and free Web hosting services, a speaker can, via the Internet, address the public, including patrons of public libraries, for little more than the cost of Internet access. . . .
>
> . . . .
>
> . . . A faithful translation of First Amendment values from the context of traditional public fora such as sidewalks and parks to the distinctly non-traditional public forum of Internet access in public libraries requires that content-based restrictions on Internet access in public libraries be subject to the same exacting standards of First Amendment scrutiny as content-based restrictions on speech in traditional public fora such as sidewalks, town squares, and parks.[204]

---

202. *Id.* at 205 (Rehnquist, C.J., plurality).

203. *Id.* (Rehnquist, C.J., plurality).

204. *Am. Library Ass'n I*, 201 F. Supp. 2d 401, 466-70 (E.D. Pa. 2002) (three-judge panel), *rev'd*, 539 U.S. 194 (2003) (plurality).

In translating the values underlying the public forum doctrine from real space to cyberspace, courts should follow the careful analysis of the three-judge panel in *American Library Ass'n* and look to the speech-facilitating functions served by Internet forums.

Furthermore, the original justification for treating only historical and traditional public forums as public forums is no longer persuasive. In 1939, the Supreme Court initially justified its creation of the public forum doctrine in *Hague* by adverting to the property law doctrine of prescriptive easements (akin to the doctrine of adverse possession), through which trespassers can acquire the right to use another's property if they have so used the property continuously for an extended period of time. Because the *Hague* Court was seeking a justification for removing from the government its plenary rights as a property owner, the Court relied on well-established property law doctrine as a justification for so doing.[205] The Court explained that since citizens have used streets, sidewalks, and public parks "time out of mind" for expressive purposes, they have in effect secured an easement by prescription to continue to do so. While the prescriptive easement justification, and its reliance on long-term historic use of public property by private citizens, may have been important in ushering in the public forum doctrine, subsequent courts and theorists have abandoned the prescriptive easements justification underlying this prong of the public forum analysis.[206] And, as the three-judge panel in *American Library Ass'n* explained, it is more conceptually coherent to look to the present purpose and function of the subject forum within our system of democratic self-government, rather than the historical uses of such a forum, in determining the level of scrutiny to apply to restrictions of speech within it.

It might be countered that the "traditional public forum" prong of the public forum analysis need not be translated to account for new mediums of expression because, after all, there is a second prong—the "designated public forum" prong—of the public forum analysis. Even if a forum is not deemed a "traditional public forum," it can still be deemed a "designated public forum" and thereby secure full First Amendment protection as a

---

205. Interestingly, however, individuals generally cannot acquire rights via prescription with respect to government-owned property. *See, e.g.*, JOSEPH SINGER, PROPERTY LAW: RULES, POLICIES, AND PRACTICES 221 (3d ed. 2002).

206. *Cf.* Harry Kalven, *The Concept of a Public Forum:* Cox v. Louisiana, 1965 SUP. CT. REV. 1; Geoffrey Stone, *Fora Americana: Speech in Public Places*, 1974 SUP. CT. REV. 233, 238 (criticizing the Court's focus on historical uses of places like parks and streets, "instead of forthrightly recognizing that access to public property for speech purposes is essential to effective exercise of First Amendment rights").

public forum. The designated public forum prong of the analysis, however, is similarly fraught with conceptual difficulties. To be deemed a designated public forum, the government must have made "an affirmative choice to open up its property for use as a public forum."[207] However, the relevant government decision maker knows that once it makes such an affirmative choice, any regulations that it imposes upon speech within that forum will be subject to strict—and likely fatal—scrutiny. Certainly, in every case in which plaintiffs challenge such regulations and claim that such regulations are subject to strict scrutiny because they are imposed within a designated public forum, the government will defend by claiming that it has not made the requisite affirmative choice to designate the forum as a public forum. Accordingly, the "traditional public forum" prong of the public forum analysis remains an important one that should be updated and translated to enable the First Amendment to protect speech within new communications mediums.

2. Following the Lead of State Courts

If the Supreme Court persists in its parsimonious interpretation of the state action doctrine and declines to subject private actors' speech regulations to scrutiny under the First Amendment, state courts should interpret their state constitutions' free speech clauses to extend to private speech regulations. Precedential support for speech-protective interpretations of state constitutions exists in contexts similar to those presented by widespread private regulation of Internet speech. In *Robins v. Pruneyard Shopping Center*,[208] the California Supreme Court interpreted the free speech protections in the California Constitution to apply to regulations imposed by private entities. In *Pruneyard*, several California high school students sought to protest a United Nations' resolution opposing "Zionism" by distributing leaflets in a large privately-owned shopping mall located in California.[209] The case arose subsequent to *Lloyd Corp. v. Tanner*, in which the United States Supreme Court held that students protesting the Vietnam War had no First Amendment right to do so within the confines of a privately-owned shopping center.[210] While recognizing that the First Amendment, per the Supreme Court's decision in *Lloyd*, did not grant the high school activists the right to so protest, the California Supreme Court held that the California Constitution's free speech clause granted the pro-

207. *See, e.g.*, Perry Educ. Ass'n v. Perry Local Educators' Ass'n, 460 U.S. 37 (1983).

208. 592 P.2d 341 (1979), *aff'd*, 447 U.S. 74 (1980).

209. *Id.* at 342.

210. 407 U.S. 551 (1972).

testors this right—notwithstanding the fact that their protest took place on private property.[211] In weighing the shopping center's right to exclude individuals from its property against the free speech rights of the protestors, the California Court interpreted its state constitution to hold that the protestors' free speech rights outweighed the mall's private property rights.[212]

The shopping mall challenged the California Supreme Court's decision in the United States Supreme Court, claiming that the ruling deprived it of its property without just compensation in violation of the Fifth and Fourteenth Amendments.[213] The Supreme Court rejected this challenge, holding that California enjoyed the power to interpret its state constitution's free speech provisions more broadly than the First Amendment (and that so doing did not constitute a taking of property without just compensation). The Court explained, "[o]ur reasoning in *Lloyd* does not . . . limit the authority of the State to exercise its police power or its sovereign right to adopt in its own Constitution individual liberties more expansive than those conferred by the Federal Constitution."[214] The United States Supreme Court also rejected the shopping mall's arguments that the state court's holding violated the mall's First Amendment rights by forcing it to use its property as a forum for the speech of others with which it disagrees. The Court explained, first, that the protestors' speech was unlikely to be identified with the shopping mall, and that, in any case, the shopping mall could simply post a notice disclaiming any sponsorship of the protestors' speech.[215]

The California Supreme Court further championed the primacy of free speech rights over property rights in its recent decision in *Intel Corp. v. Hamidi*.[216] While not explicitly relying on Hamidi's free speech rights, the California Supreme Court declined to construe Intel's property rights so broadly as to include the right to prohibit Hamidi from sending e-mails via Intel's e-mail server. The court recognized that extending property rights so broadly would hamper open Internet communication and impose costs on society, and implicitly privileged the right to free expression over private property rights in this context.

In short, states are free to define their citizens' free speech rights under their state constitutions to incorporate individuals' right to express themselves on private property, even if their First Amendment rights do not

211. *Id.* at 346.
212. *Id.*
213. Pruneyard Shopping Ctr. v. Robins, 447 U.S. 74 (1980).
214. *Id.* at 81.
215. *Id.* at 86-88.
216. 79 P.3d 296 (Cal. 2003).

extend so far. And such an extension of free speech rights does not violate the property or free speech rights of the owner of the property on which such speech occurs. Following California's lead, states should interpret their own constitutions' free speech clauses to grant individuals the right to express themselves in privately-owned forums for expression that are the functional equivalent of traditional public forums.[217]

States, through their courts or legislatures, should also explicitly define public forums to include Internet forums that are generally open to the public for free speech purposes, even where such forums are privately owned. Once again, California is illustrative. Concerned with what it found to be a "disturbing increase in lawsuits brought primarily to chill the valid exercise of the constitutional right of freedom of speech,"[218] the California legislature enacted a statute aimed at deterring SLAPP suits—"strategic lawsuits against public participation."[219] This statute grants individuals the right to speak and petition freely within "public forums"—whether such forums are publicly or privately owned—free from harassing and meritless lawsuits aimed at chilling such speech. In particular, California's anti-SLAPP statute grants individuals the right to "dismiss at an early stage non-meritorious litigation meant to chill the valid exercise of constitutional rights of freedom of speech and petition in connection with a public issue."[220] The statute defines an "act in furtherance of a person's right of petition or free speech" to include "any written or oral statement or writing made in a place open to the public or a public forum in connection with an issue of public interest."[221] In interpreting this language, a California court of appeal recently found that privately-owned Internet

217. Several other states have followed California's lead in extending free speech rights to private property in real space. *See, e.g.*, Batchelder v. Allied Stores Int'l, Inc., 445 N.E.2d 590 (Mass. 1983) (finding that the state constitution's free speech guarantee extends to political speech within private shopping centers); N.J. Coalition Against War in the Middle E. v. J.M.B. Realty Corp., 650 A.2d 757 (N.J. 1994) (holding that the state constitution's guarantee extends to political and "societal" speech within private shopping centers); Alderwood Assocs. v. Wash. Envtl. Council, 635 P.2d 108 (Wash. 1981) (finding that the state constitution's free speech guarantee extends to political speech within private shopping centers).

218. Bidbay.com, Inc. v. Spry, No. B160126, 2003 Cal. App. Unpub. LEXIS 2057 (Ct. App. Mar. 4, 2003).

219. CAL. CODE CIV. PROC. § 425.16 (Supp. 2004). In passing this legislation, California joined a number of other states that have also passed anti-SLAPP laws. *See* California Anti-SLAPP Project, *available at* http://www.casp.net/menstate.html (last visited Aug. 2, 2004) (listing states that have passed and are considering passage of anti-SLAPP legislation).

220. CAL. CODE CIV. PROC. § 425.16.

221. *Id.* § 425.16(e)(3).

chat rooms and message boards constitute public forums where they are "open to the public or to a large segment of the interested community."[222] California's anti-SLAPP statute thereby grants individuals the right to speak freely on matters of public importance in Internet forums, whether publicly or privately owned, free from reprisals in the form of meritless lawsuits.

States should further grant individuals the right to speak freely on matters of public importance within Internet forums, free from reprisal in the form of lawsuits designed to chill the exercise of their free speech rights or in the form of self-help censorship efforts by Internet actors. With the technological tools available to Internet actors to censor speech with the click of a mouse, technological measures restricting speech present an even greater harm to speakers than lawsuits designed to chill their speech. States should define public forums to include privately-owned Internet forums for expression that are open to the public and should protect individuals' right to express themselves on matters of public interest, broadly construed, within such forums.

In short, the Supreme Court should meaningfully translate First Amendment values for twenty-first century communications media by reconceptualizing the "traditionality" requirements in First Amendment doctrine. Courts should analyze the current function of the forum at issue within our system of democratic self-government, rather than the historic uses of such forums. If the Supreme Court persists in its unwillingness to translate First Amendment values to render the right to free expression meaningful in the new technological age, then states should interpret their own constitutions' free speech clauses—or, in the alternative, enact legislation—to provide individuals with meaningful rights to express themselves on the Internet.

B. Restoring the Values of the Interstitial Public Forum Within Cyberspace

The existence of interstitial public forums in real space provides speakers with forums from which to target effectively their speech toward adjacently-located, privately-owned establishments. Such forums enhance the ability of speakers to target their speech effectively toward their desired audience. Because real space is generally characterized by the intermingling of publicly and privately-owned property, public property adjacent to private property can be used as a launching point from which to

222. *Bidbay.com*, 2003 Cal. App. Unpub. LEXIS 2057, at *14-*15.

target private entities, through expression such as picketing, boycotting, and general protesting.

In translating this function of public forums into the Internet realm, we must first consider the appropriate cyberspace analogue to the real space characteristic of adjacency.[223] In real space, interstitial public forums are valuable because, as a result of their physical proximity to privately-owned property, they are particularly well-suited forums from which to target privately owned property. Because such physical proximity or adjacency has no direct correlate in the Internet realm, we must look to other features that enable individuals effectively to target their speech at private entities.

One important way in which protestors target the objects of their criticism in cyberspace is by the use of search terms within Internet search engines. Internet speakers who desire to criticize an entity can utilize search terms to capture the attention of Internet users generally seeking information about such entities. Just as a real space protestor might protest on the sidewalk adjacent to General Motors (GM) to launch a targeted attack on GM and to reach individuals seeking out GM in real space, so too a cyberspace protestor might choose to launch a targeted attack on GM by utilzing a user's search for GM to reach individuals seeking out GM in cyberspace. The cyberprotestor may do this in several ways: by using "General Motors" (or other General Motors trademarks) as one of the metatags for her critical website; by using a GM-related mark as a keyword in her advertisement criticizing GM; or by using GM as part of the domain name associated with her critical website.

In recent years, powerful business owners have wielded trademark law in their efforts to silence critics who have made use of the business owners' trademarks to reach audiences seeking information on such business owners. Although the case law in this area is not uniform, it is clear that trademark owners have been far more successful in shutting down such cyberprotestors than they have been in silencing those who launch protests against them from adjacent interstitial public forms in real space. First, trademark owners have successfully wielded trademark law to prohibit critics from using their trademarks as keywords or metatags to drive interested searchers toward information critical of the trademark owners.[224] Under the recently-enshrined doctrine of "initial interest confusion,"

---

223. *See* Zatz, *supra* note 140, at 185-86.

224. *See, e.g.*, J.K. Harris & Co. v. Kassel, 62 U.S.P.Q.2d (BNA) 1926 (N.D. Cal 2002) (holding that direct competitor's use of plaintiff business's trade name in its metatags leading to website with unfavorable information about plaintiff constituted actionable initial interest confusion).

courts have held that even where individuals seeking out information about a business are not confused by such critical sites, the sites nonetheless infringe the marks of the business owner.[225] Second, trademark owners have successfully wielded trademark law[226] (as well as the Uniform Dispute Resolution Policy promulgated by ICANN[227]) to prohibit critics from using their trademarks as part of domain names for their critical websites.

To translate the values of interstitial public forums into the Internet context, courts should limit trademark owners' relief to circumstances in which defendants' websites result in actual confusion. Courts should also recognize and protect protestors' First Amendment right to use others' marks in critical, non-confusing contexts on the Internet. Indeed, surmounting the public-private distinction to protect critics' First Amendment rights is not as formidable a hurdle as in the context of general public forums. Although it might be supposed that critics have no First Amendment right to use others' marks (which are the private intellectual property of these entities) to advance their criticisms, courts have in fact long recognized such a right. The case of *L.L. Bean, Inc. v. Drake Publishers*[228] is illustrative. In that case, L.L. Bean sought to wield infringement and dilution causes of action to silence expressive speech incorporating its trademark. High Society, a commercial, adult-oriented magazine, published an article parodying the popular L.L. Bean sportswear catalog, under the title "L.L. Beam's Back-to-School-Sex-Catalog." The article included variations on L.L. Bean's trademarks and featured pictures of nude models in sexually explicit positions using products similar to those offered in L.L. Bean catalogs and described the products in a "crudely humorous fashion."[229] L.L. Bean, not amused, sued the publisher of High Society for trademark infringement and dilution. The district court, while rejecting

---

225. *Id.* at 1929.

226. *See, e.g.*, PETA v. Doughney, 263 F.3d 359 (4th Cir. 2001) (holding that defendant's website, peta.org, which parodied People for the Ethical Treatment of Animals (PETA), infringed PETA's trademark).

227. *See* Burlington Coat Factory Warehouse Corp. v. Smartsoft, L.L.C., WIPO Case No. D2001-1792 (WIPO Arbitration & Mediation Ctr. Mar. 1, 2000) (concluding that the use of domain names burlingtonmurderfactory.com, burlingtonkillfactory.com, burlingtondeathfactory.com, burlingtonblood-factory.com, and burlingtonholocaust.com for website critical of Burlington Coat Factory was in violation of ICANN's Uniform Domain Name Dispute Resolution Policy and thus could be enjoined), *available at* http://arbiter.wipo.int/domains/decisions/html/2000/d2000-1792.html. *See generally* Nunziato, *supra* note 105 (criticizing policy's effect on critical expression).

228. 811 F.2d 26 (1st Cir. 1987).

229. *Id.* at 27.

L.L. Bean's infringement cause of action due to lack of confusion, sustained L.L. Bean's cause of action for trademark dilution, finding that High Society's parodic use of the L.L. Bean mark within this context diluted the mark's distinctive qualities.[230] The court enjoined publication of the article.[231]

High Society appealed, claiming that the district court's order enjoining publication of the article violated its right to freedom of expression and asserting that it enjoyed a First Amendment right to use the Bean trademark for expressive purposes. The First Circuit agreed, holding that although this case involved a suit between two private entities, the lower court's interpretation of the state anti-dilution law to enjoin defendant's speech constituted state action restricting expression in violation of the First Amendment.[232]

In the Internet context of metatags, key words, and domain names, courts should follow the rationale of the L.L. Bean court and protect critics' First Amendment right to use the trademarks of another in a non-confusing manner to direct targeted criticism at the owner of such marks. Critical websites or advertisements that incorporate their target's trademarks as metatags or keywords or as part of their domain names should be immune from trademark liability.

In translating the speech-protective functions of interstitial public forums from real space to cyberspace, courts (and arbitrators) should grant broad protection to critical speech on the Internet, even where such critical speech incorporates the property of the entity subject to criticism and even where such critical speech occurs in expressive forums owned and regulated by private actors. Because actual, publicly-owned interstitial public forums do not exist in cyberspace, the functions served by such forums—namely, the facilitation of effective criticism and protest of "adjacent" private property owners—must be protected by courts by according wide berth to the use of others' intellectual property in cyberspace for purposes of criticism.

VI. CONCLUSION

Contrary to the widely-held perception of the Internet as one great public forum for individuals to express themselves, the Internet has become transformed by privatization into a collection of largely privately-owned and privately-regulated places. Because the relevant "property"

230. *Id.*
231. *Id.*
232. *Id.* at 33.

that makes up the Internet is overwhelmingly privately owned, the restrictions on speech imposed by the property owners within such places have been held to be outside the purview of the First Amendment. With the Supreme Court's contraction of the state action doctrine in recent years, private regulations of speech within expressive Internet forums have become essentially immune from scrutiny under the First Amendment. Furthermore, even government restrictions on speech within expressive Internet forums have become immune from meaningful First Amendment scrutiny.

In this Article, I have argued that the death of public places in cyberspace brings with it the erosion of important First Amendment values. Most importantly, the death of public places in cyberspace heralds the death of public forums in cyberspace—the most important vehicle for the protection of free speech in real space. With the death of public forums in cyberspace, long-standing constitutional protections for speech are in danger of being seriously eroded in cyberspace. Courts and legislatures must act to remedy this problem and faithfully translate First Amendment values to render the values meaningful in the new technological age.

CHOOSING AMONG ANTITRUST LIABILITY STANDARDS UNDER INCOMPLETE INFORMATION: ASSESSMENTS OF AND AVERSIONS TO THE RISK OF BEING WRONG

By Barbara Ann White[†]

David McGowan observes in his symposium article, *Between Logic and Experience: Error Costs and United States v. Microsoft Corp.*,[1] that conceivably in the realm of theory, the different schools of thought on how to evaluate and treat suspicious anticompetitive activity could coincide with each other.[2] What prevents this from happening, he suggests, is that in the realm of reality, courts must make decisions under conditions of imperfect information, not only as to the anticompetitiveness of the scrutinized conduct but also as to the impact of their decisions, since incomplete information implies a risk that whatever the court decides, the decision is wrong.[3] This yields the possibility of "error costs,"[4] both of "false positives" (finding anticompetitive conduct when in fact it is welfare-enhancing) and "false negatives" (finding the conduct efficient when in fact it is anticompetitive). McGowan argues that since the gaps in actual knowledge in any particular case must be overcome by certain leaps of faith, the guidelines an antitrust scholar advocates (or a particular court adopts) to evaluate specific conduct must to a certain extent turn on ideology; pure logic cannot resolve completely the uncertainty embedded in the incomplete information.[5] As a corollary to his thesis, McGowan noted during the presentation that it would be interesting to explore why different antirust scholars adopt different ideological solutions to the logically unresolvable problems.

© Barbara Ann White

† Associate Professor of Law, University of Baltimore; Chair, Section on Antitrust and Trade Regulation of the Association of American Law Schools. This Commentary discusses several papers presented at an antitrust panel, Evolving Antitrust Treatment of Dominant Firms, at the Association of American Law Schools (AALS) 2005 Annual Meeting. I wish to thank Steve Salop, David McGowan, Doug Melamed, and Alan Meese for useful comments.

1. David McGowan, *Between Logic and Experience: Error Costs and* United States v. Microsoft Corp., 20 BERKELEY TECH. L.J. 1185 (2005).
2. *Id.* at 1187, 1244-45.
3. *Id.* at 1187.
4. *Id.* at 1186 n.2.
5. *Id.* at 1242.

The other two papers for this symposium, Douglas Melamed's article, *Exclusionary Conduct Under the Antitrust Laws: Balancing, Sacrifice, and Refusals to Deal*,[6] and Steven Salop's paper, "Section 2, Consumer Welfare Effects, and the Flawed Profit-Sacrifice Standard,"[7] each take their own approach regarding conduct evaluation. The two also discuss extensively the problem of error costs, albeit reaching different conclusions as to their impact and treatment.[8] Melamed proposes a specifically defined measure of evaluating a firm's exclusionary conduct based on whether or not the conduct's profitability depends solely on rivals exiting, or the "profit sacrifice test," in order to expand on capturing false negatives without much sacrifice in the way of false positives.[9] Salop proposes a more comprehensive evaluation by suggesting that courts consider the scrutinized conduct's overall impact on consumer welfare, or the "consumer welfare effects standard."[10] Salop's proposal is to expand even further the capture of false negatives.

I agree with McGowan's overall analysis of the source of conflict between the current predominant views regarding the treatment of potentially anticompetitive conduct, particularly as it pertains to the conduct of dominant firms. I agree that the differences arise from the incomplete information that necessarily informs and renders uncertain the impact of decisions regarding firms' conduct. Though with additional analysis regarding whether antitrust decisions should be based on our experience rather than logic leads McGowan to conclude that his particular antitrust perspective—relying on the market rather than the courts for correction of anticompetitive conduct, or the "market correction approach"—is the superior one,[11] I am more persuaded, within the same framework, by Melamed's more interventionist standard of the profit sacrifice test and am most persuaded by Salop's even more encompassing consumer welfare effects standard.

I also think McGowan's insight that the differences among scholars (and courts) are ideologic more than differences in logic can be usefully refined further. McGowan ascribes the source of ideological differences to the differences in faith in the marketplace versus faith in the government

---

6. A. Douglas Melamed, *Exclusionary Conduct Under the Antitrust Laws: Balancing, Sacrifice, and Refusals to Deal*, 20 BERKELEY TECH. L.J. 1247 (2005).

7. Steven Salop, Section 2, Consumer Welfare Effects, and the Flawed Profit-Sacrifice Standard (May 11, 2005) (unpublished manuscript, on file with author).

8. Melamed, *supra* note 6, at 1257-61; Salop, *supra* note 7, at 53-54.

9. Melamed, *supra* note 6, at 1255-57.

10. Salop, *supra* note 7, at 3, 24-28, 48.

11. McGowan, *supra* note 1, at 1242-45.

to regulate effectively, whether the errors are with regard to anticompetitive conduct or with regard to the errors in the decisions themselves.[12] This suggests, for him, that the various points of view fall into one of two categories: either a market correction approach or an integrationist approach, in which consideration of the potential error costs of the decision are integrated into the decision process itself.[13] McGowan advocates the former while Melamed and Salop presumably fall into the latter.

I believe, however, that the differences in viewpoint are affected not only by which correction tool a scholar or court find more effective (the market or the government), but also by the differences in their evaluation of the consequences of the false positives versus the false negatives in the specific instance under consideration. Inherent in any antitrust decision is a trade-off between the risk of a false positive and a false negative. As such, the differences in the valuation of those consequences may lead scholars and courts to differ in their reluctance to undertake the risks of particular outcomes when they choose to guard against one error as opposed to the other in a particular case. Therefore, I suggest that one's antitrust position is not only informed by one's belief in the ability of the marketplace relative to the government to correct anticompetitive problems, but also to the degree, in any particular circumstance, one is more risk-averse to the consequences of one error cost over the other. Incorporating a risk-averse analysis into the framework set forth by McGowan allows for a continuum of approaches rather than a stark categorization into two camps. This then allows for a distinction, in the context of that continuum, between Melamed's and Salop's papers. Thus, I would place McGowan's perspective—primary reliance on the marketplace—toward one end; Salop's perspective, which recommends a relatively broad and more fine-tuned government evaluation of conduct, toward the other; and Melamed's perspective of specific finite criteria to determine intervention, somewhere in the middle.

McGowan expands on his analysis of differences arising from the fusion of the ideologic with the logic by assessing how informative actual experience is with regard to practical application of the different views, as compared with what decisions the logic of these different views would dictate. He suggests that even though theoretical developments might guide us to more nuanced decision making in antitrust matters, the same element of imperfect information that requires us to supplement logical conclusions with ideological choices also compels us at times to choose

12. *Id.* at 1197-99.
13. *Id.* at 1186-88.

what experience tells us is a more effective practical course over what our theory would choose.[14] McGowan uses the *Microsoft* case to prove his point: that at times experience (of the practicalities) should trump logic (no matter how insightful or far-reaching theory takes us in the theoretical plane).[15] In particular, McGowan argues that though the *Microsoft* court couched its opinion in a theoretical framework which suggested a fairly strong integrationist approach, ultimately the actual liabilities and remedies accomplished very little in the way of change.[16] This, he notes, is close to the outcome that would have occurred if a market correction approach had been adopted instead (and the case never brought).[17] His analysis of the *Microsoft* case, therefore, argues for more reliance on the market correction approach for practical reasons even if, in theory, one supports a more integrationist approach. For McGowan, the *Microsoft* case even suggests that the integrationist approach is, in practical terms, a failure, as clearly the track record of government intervention has been rather abysmal, at least in the general purpose computing market.[18]

Unstated in McGowan's assertion—that the various approaches would coincide in theory if not for the practical impingement of imperfect information—is that, currently, most antitrust scholars adhere to the view that the goal is to maximize consumer welfare and therefore that standard should underlie antitrust judgments with regard to specific conduct. This, of course, contrasts with antitrust's past judicial history, when there was a strong sentiment in favor of preserving the number of competitors for its own sake as an ideal for maintaining competitive markets. This older standard gave way to the current view of judging conduct in light of consumer welfare in the wake of the Chicago School revolution. Certainly, not only is the consumer welfare standard implied by McGowan's discussion,[19] it is also more explicitly expressed in both Melamed's and Salop's papers.[20] Consumer welfare is a concept from economics, and to the extent that there is agreement among modern day antitrust scholars that not only is consumer welfare the benchmark by which business conduct should be

14. *Id.* at 1197-99.
15. *Id.* at 1211-12.
16. *Id.* at 1223-25.
17. *Id.* at 1244-45.
18. *Id.* at 1189.
19. *Id.* at 1188.
20. Melamed, *supra* note 6, at 1252 ("The principal function of an antitrust rule in a law enforcement regime is to create appropriate incentives for the avoidance of welfare-reducing conduct."); Salop, *supra* note 7, at 2 ("[A]ntitrust law is said to be a 'consumer welfare prescription.'"); *see* Salop, *supra* note 7, at 23-40 (describing further the consumer welfare standard).

judged but also as to how it is defined,[21] McGowan's assertion that the various views regarding antitrust treatment would coincide with each other in a world of perfect information is all the more convincing.[22]

In order to explain why I am most persuaded, within the context of McGowan's framework, by the approach at the other end of the continuum—Salop's consumer welfare effects standard—it might be useful to examine the ways the three papers fall along a continuum. The philosophies of the three papers might be summarized as follows.

McGowan's market correction view holds that the government, whether by court or administrative action, should only interfere with business conduct when the conduct is unambiguously anticompetitive in nature and, moreover, egregiously so. Any lesser standard increases the risk of a false positive, as more ambiguous conduct is more likely to be judged erroneously in violation of antitrust law.[23] The subsequent interference by the government to "correct" this erroneously judged violation will cause a disruption in the workings of the marketplace, creating inefficiencies that are consumer welfare-reducing.[24] This perspective recognizes that by leaving alone more ambiguous behavior, the false negatives that otherwise might have been caught, although welfare-reducing in the short term, will be remedied by forces the market will create to correct and, thus, remove the resultant inefficiencies. The assessment of the "market correctionists" is that harm done by the government from false positives (far) exceeds the harm done by false negatives because the market is more capable and efficient at correcting the latter than either the government or the market is able to correct the former.[25] Of course, this reflects the valuation assessment by those who support this view and demonstrates a high degree of risk averseness to the consequences of false positives.

---

21. There is debate as to what are the welfare measures that antitrust seeks or ought to seek to maximize—that is whether to maximize "consumer welfare" or "aggregate (or total) welfare." *See* Salop, *supra* note 7, at 24 n.58. There is also the issue of whether courts adopt the consumer welfare standard in section 2 cases. Alan Meese observes that some courts have rejected such an approach since 1955. E-mail from Alan Meese, Professor of Law, William & Mary School of Law, to Barbara Ann White, Professor of Law, University of Baltimore School of Law (May 13, 2005) (on file with author).

22. One sometimes hears of positions such as Melamed's and Salop's as being of the old school view, in that they do focus on what happens to rivals in the course of the defendant firm's scrutinized conduct. Thus, it is important to recognize, as clearly McGowan does by implication, that observations of effects on rivals of particular conduct do not mean abandonment of the modern day measure of consumer welfare.

23. McGowan, *supra* note 1, at 1190-91.

24. *Id.*

25. *Id.* at 1189.

Melamed advocates the profit sacrifice test. To a certain extent, the motivations underlying the test are concerns similar to the market correction approach, in that the goal is to create a standard that has a low probability of a false positive.[26] The profit sacrifice test also has, however, the advantage of casting a wider net to capture true positives than does the market correction view. As Melamed defines the profit sacrifice test, the question is whether the incremental cost to the firm of the conduct under scrutiny exceeds the incremental revenues or cost savings the firm achieves (not including increased revenues or reduced costs that are due solely to rivals' exit as a result of the scrutinized conduct).[27] If the incremental revenues exceed the incremental costs, then the activity makes sense for the firm to undertake (that is, it is profitable) regardless of whether rivals exit or not. The conclusion is then that the exit of the rivals is ancillary to the conduct and not an antitrust violation. However, if the conduct is profitable only if rivals exit—that is, without their exit, the firm would actually be sacrificing profits—then the profit sacrifice test concludes that the primary intent of the conduct is to exclude rivals from the market for profit purposes.[28] Furthermore, the conduct is necessarily consumer welfare-reducing and thus properly condemned as anticompetitive.[29] The advocates of the profit sacrifice test argue that this measure captures more of what would otherwise be false negatives under the market correction approach while not, in any significant way, increasing false positives.

The consumer welfare effects standard is a more wide-ranging balancing test than either of the other two. It is designed to cast a much broader net for true positives and, through government intervention, come closer to the goal of enhancing consumer welfare, the benchmark goal underlying all three views.[30] The consumer welfare effects standard acknowledges an inherent problem in judging antitrust conduct, and that problem is, for the most part, that most business conduct has both procompetitive and anti-

---

26. Melamed, *supra* note 6, at 1257.

27. *Id.* at 1255-56.

28. *Id.*

29. Melamed notes that it is not necessary for the firm to "sacrifice" profits—that is, lose profits but for the exit of the rivals—for the conduct to fail profit sacrifice test. Presumably, even if the conduct was a break-even proposition for the firm without its rivals exiting but profitable upon their exit, then the inference would be that the intent behind the conduct was merely anticompetitive and would still be condemned. *Id.*

30. Salop emphasizes that the consumer welfare effects standard focuses solely on the welfare of consumers and not on aggregate welfare. *See* Salop, *supra* note 7, at 24 n.58; discussion *supra* note 21.

competitive effects.[31] In other words, with most business activity, one can find welfare-enhancing as well as welfare-reducing aspects. The court's task is to weigh the factual evidence of welfare enhancement against the evidence of welfare reduction and if the enhancement exceeds the reduction the firm is not held liable; vice versa, the firm is found in violation of the antitrust laws.[32]

The primary criticism of the market correction approach is fairly obvious: it leaves too many false negatives—that is, it leaves too many firms engaging in welfare-reducing conduct unconstrained. Those who critique the market correction approach, as McGowan says, do not have much faith in the market to correct false negatives.[33] The advocates of both the profit sacrifice test and the consumer welfare effects standard—or, as McGowan would have it, the integrationists—assess the persistence of the welfare reduction from the exclusionary behavior of dominant firms as a much graver consequence than do the market correctionists. The integrationists not only perceive the government as capable of properly assessing anticompetitive conduct, but they also evaluate the consequences of the false positives as less severe than do the market correctionists.[34] Typically, the integrationists also see the consequential effects of false positives as much less than the consequential effects of the false negatives if those conducts are left unchecked.[35] Thus, the integrationists are highly motivated to find and refine the tools the government can employ to gauge more correctly the net welfare effects of specific business conduct.

The criticism of the profit sacrifice test, from the perspective of the market correctionists, is the perceived increase in the risk of false positives,[36] which the market correctionists evaluate at a very high cost to society. The advocates of the consumer welfare effects standard—the consumerists—view the profit sacrifice test as good but not enough. The profit sacrifice test alone is not sufficient to capture as many of the true positives as the consumerists believe is possible, and thus leaves too many false negatives for the consumerists' comfort. More than the profit sacrifice test is needed and is possible, the consumerists argue. Salop, through an example of suspect predatory pricing, asserts that the profit sacrifice test may create a false positive and thus it may fail in its effort to increase

31. Salop, *supra* note 7, at 47.
32. *Id.* at 52-54.
33. McGowan, *supra* note 1, 1192-93.
34. *See, e.g.*, *id.* at 1192-99.
35. *See, e.g.*, *id.*
36. Melamed, *supra* note 6, at 1259-60.

the number of true positives (that is, reduce the number of false negatives) without simultaneously increasing the number of false positives.[37]

The supporters of the profit sacrifice test argue in their defense that the test yields a fairly simple and straightforward rule that not only is relatively easy for jurists to apply but also creates a measure of certainty for firms, assuring that the firms can continue to engage in innovative procompetitive conduct without being caught in a tangled net of antitrust accusations that may have no merit.[38] Clearly, the advocates of the profit sacrifice test weigh the consequences of false negatives far more heavily than the market correctionists, but like the market correctionists, they also measure the consequences of false positives fairly strongly. The profit sacrifice test seems to its supporters to take skillful advantage of the economic insights into the workings of the firm to fashion a rule that captures anticompetitive conduct with more precision than previous standards with small risk of increasing false positives. This middle ground, to its supporters, seems to be the best of all choices. The test also appeals to our instinct that a profit-maximizing firm would not forgo profits unless it anticipates recouping them at some later date.

The primary criticism of the consumer welfare effect standard is that its biggest advantage is also its biggest disadvantage. This is also when McGowan's "experience ought to trump logic in the world of practicalities" argument has its most persuasive import. The consumer welfare effect standard is a broader-ranged nuanced analysis requiring examination of the impact of scrutinized conduct on consumers' welfare overall, with particular attention paid to price, quality, and innovation.[39] In addition, the standard considers a much broader range of conduct. Though in theory this has great appeal because it addresses directly the presumptive goal of antitrust—maximizing consumer welfare—and it should reduce dramatically the likelihood of false negatives, from a practical perspective, its critics hold that its application is untenable. Judges are not skilled enough to evaluate the complex economic arguments and data analyses asserting consumer welfare enhancement or reduction, nor are they able to properly weigh the measures of the two to conclude what the net effects are. Juries are even less capable. Moreover, even examining data to put forth plaintiff or defendant arguments is extremely difficult and expensive, taking great

37. Salop, *supra* note 7, at 19-20.
38. Melamed, *supra* note 6, 1256-27.
39. Salop, *supra* note 7, at 3-4, 25.

talent and skill. To some, it seems perhaps impossible to achieve properly.[40]

That is one of the points McGowan makes in his article with regard to the *Microsoft* case. There the very best analysts in the country argued on both sides. In theory the court sided with the consumerists in its analysis but concluded with a "tapioca pudding" of effect.[41] McGowan argues that the conduct that was in fact most market-enhancing for Microsoft went unpunished and the conduct with trivial implications was what was sanctioned with trivial results.[42]

Finally, a serious problem with a welfare approach for evaluating conduct is that each case is very situation specific. The welfare effects standard creates an environment with limited precedential value. Businesses will not know how to proceed when desiring to undertake innovation; what seems like an arbitrary antitrust web may fall on them at any time, and is more likely to do so the more successful they are, as great success usually breeds the exit of others. Some argue that the welfare effects approach could hamper creativity and economic progress.[43] Even if it could be properly implemented, the welfare effect standard seems doomed not to succeed.

Thus, the consumer welfare effect standard seems to be a paradigm of what McGowan was referring to in his argument that sometimes experience should trump logic. The consumer welfare effect standard is essentially perfect in theory, comporting precisely and logically with all modern scholars' views that the goal of antitrust law should be consumer welfare enhancement. But in practice, given its complexity, the difficulties in implementing it effectively, and its confusing implications for predictable precedent,the consumer welfare effects standard seems doomed to failure. Experience seems to tell us that this logic should be trumped.

With such dire evaluations, why then might the consumer welfare effects standard seem the most persuasive? There are several reasons.

First, its appeal does not rely just on a lack of faith in the market correction approach. The appeal is more than just ideologic. Not only is the lack of faith in the market's ability to correct anticompetitive behavior a factor, but also the recognition of the unlikelihood that, even when the market does manage to move the actors to more efficient conduct, the in-

40. *See* Melamed, *supra* note 6, at 1253-54. Melamed also concludes that such balancing tests would improperly evaluate product innovation and price-cutting behavior, as well. *See id.* at 1267.

41. McGowan, *supra* note 1, at 1189.

42. *Id.* at 1226.

43. *See, e.g.*, Melamed, *supra* note 6, at 1253-54.

dustry will now progress on an economic path that is the not one consumers would have chosen but for the initial anticompetitive conduct. It is one thing for the marketplace to move all parties to some efficient conduct, it is another to adopt the same efficient path that was emerging before the exclusionary conduct.

Despite the fact that economic efficiency is often expressed as synonymous with maximizing welfare—it is not. Efficiency is, of course, a necessary condition for a maximally welfare-enhancing allocation of resources. However, more than firms operating efficienctly is needed. What is well-known by those steeped in economic analysis of antitrust law but seems to have been forgotten—or at least not mentioned during discussions of antitrust issues—is that there is not a sole, unique economically efficient path on which an economy can progress. As basic economic analysis teaches us, for every given allocation of initial resources, there are a multitude of efficient paths; and for every set of initial resources, there are multiple efficient allocations. It is the interaction of constraints from different sources outside the marketplace as well as efficiency forces within the marketplace that affect which of these economic paths will be followed. Much of the outside force is derived from consumer preferences, but other factors can affect the direction as well. Anticompetitive behavior at critical points in an industry's development can alter the direction of economic progress and is likely to do so counter to consumer preferences.

It is in that context that full faith in market correction of anticompetitive conduct seems, at best, tenuous. Though most agree that inefficiency by firms will be eroded by market forces—and anticompetitive behavior is typically viewed as inefficient—it is not convincing that the efficient path the industry will now be on, post-anticompetitive behavior, is the same as the efficient one that consumers would have preferred. After all, the path is one geared to a dominant firm's conduct designed to enhance its own economic welfare through the means of subverting consumer choice.

Anticompetitive conduct is by definition undermining consumer choice. It is necessary for a dominant firm to use these tactics to exclude rivals only if enough consumers prefer the rivals' products and the dominant firm cannot woo them otherwise. So although, in the long run, the market may move to make the dominant firm efficient once again, it will be efficient around the direction generated by the firm, against the consumers' wishes and without, perhaps, preferred rivals.

This view clearly affects one's risk analysis of consequences of error costs and will also lead to an evaluation of the costs of false negatives as being very high. It cautions decision makers to be risk averse to decisions that leave open increases in false negatives.

The profit sacrifice test therefore has great appeal compared to the market correction approach, for all the reasons those who advocate the profit sacrifice test suggest: it probes the implication of a firm's conduct more thoroughly and thus is more able to capture true negatives than the market correction approach (its improvement over the older Areeda-Turner test has been much discussed in the literature); it has clarity and simplicity, making it easier for courts to apply; and it can provide sufficient precedential value to guide firms as to what will be considered problematic aggressive conduct when battling it out in the marketplace.

However, there was a brief discussion of a point during another presentation of these viewpoints that struck me and caused me to shift in favor of the consumer welfare effects test.[44] The shift in position is also consistent with McGowan's view that sometimes experience ought to trump logic. The discussion was with regard to what the effect would be on firms if the courts indeed adopted a consumer welfare effects standard. It was suggested that firms would probably, right from the start, begin accruing evidence that their conduct enhanced consumer welfare over any reduction. This implies that dominant firms would engage in their own consumer welfare effects analysis when evaluating a new course of conduct, and, though motivated by antitrust concerns, this self-evaluation would ultimately benefit consumers because of its impact on the firms' decision-making process.

If this is correct, other advantages would occur as well. As it becomes standard for firms to incorporate consumer welfare assessments along with the other measures of feasibility and profitability of any new venture, a likely outcome is that the means and methods of measuring such effects will improve over time. It also suggests that the plaintiffs and defendants will become more efficient at presenting cogent evidence and courts more efficient at evaluating net consumer welfare benefits. Experience teaches us that implementing legal standards to create incentives for the actors to better achieve certain social goals is effective. Certainly we have found that the implementation of the exclusionary rule under the Fourth Amendment dramatically improved police investigative activity, occupational health and safety laws have made the workplace dramatically safer, and product liability for design defects has caused firms to produce far safer goods than in the past century.

Though it is has been debated for sometime now whether the implementation of these standards in other areas of law has gone too far, there is

---

44. This discussion took place at the American Bar Association Section of Antitrust Law Annual Spring Meeting (Mar. 30, 2005).

no question regarding the vast improvement in these matters compared with over a century ago. Thus, if it becomes routine in the corporate world to incorporate consumer welfare analysis in the decision-making process, this atmosphere will permeate throughout the market system. The effect is likely to motivate firms to be more consistent with the goals of antitrust law, and the firms are likely to give consumer welfare effect its best and most efficient expression. Experience teaches us so.

BETWEEN LOGIC AND EXPERIENCE: ERROR COSTS AND *UNITED STATES V. MICROSOFT CORP.*

By David McGowan[†]

TABLE OF CONTENTS

© 2005 David McGowan

† Professor of Law, University of Minnesota Law School. My thanks to Mark Lemley, Alan Meese, Steve Salop, and Howard Shelanski for comments. Remaining mistakes are the sole fault of the author. A prior draft of this Article was presented at an antitrust panel, Evolving Antitrust Treatment of Dominant Firms, at the Association of American Law Schools (AALS) 2005 Annual Meeting.

This Article uses *United States v. Microsoft Corp.*[1] to assess the ongoing debate over how antitrust doctrine should deal with the risk of mistakes and the consequent problem of error costs.[2] I discuss the case in the context of the Justice Department's longstanding involvement in general-purpose computing, with an eye to the ways free and open-source software development practices ("F/OSS") may affect the structure of those markets in the future. In particular, I focus on the D.C. Circuit's two-tiered approach to causation with respect to the monopolization cause of action. The Circuit's approach is best understood as a doctrinal tool designed to minimize error costs.

There are two main schools of thought regarding relative error costs. One school, for which Judge Frank Easterbrook has been a consistent and eloquent advocate, holds that society suffers more when courts wrongly find a defendant liable than when they wrongly find no liability.[3] On this view, monopolization cases should be restricted to targeting the few business practices that we understand thoroughly enough to say with confidence that such practices undermine competition. Outside this set of practices, judges should stay the heavy hand of antitrust. A key claim of this school is that markets correct mistakes better than courts do, so I refer to it as the market correction approach. I agree with this view because experience teaches that this approach causes less harm than extensive judicial management of large firms. As I discuss in Part I, however, I concede that,

---

1. My nomenclature for the opinions in the case is as follows: United States v. Microsoft Corp., 147 F.3d 935 (D.C. Cir. 1998) [hereinafter *Microsoft I*] (contempt proceeding); United States v. Microsoft Corp., 84 F. Supp. 2d 9 (D.D.C. 1999) [hereinafter *Microsoft II*] (findings of fact); United States v. Microsoft Corp., 87 F. Supp. 2d 30 (D.D.C. 2000) [hereinafter *Microsoft III*] (conclusions of law), *aff'd in part and rev'd in part*, United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir.) (en banc) [hereinafter *Microsoft IV*] (merits appeal), *cert. denied*, 534 U.S. 952 (2001); New York v. Microsoft Corp., 224 F. Supp. 2d 76 (D.D.C. 2002) [hereinafter *Microsoft V*] (non-settling states remedy); United States v. Microsoft Corp., 231 F. Supp. 2d 144 (D.D.C. 2002) [hereinafter *Microsoft VI*]; (Justice Department settlement approval); Massachusetts v. Microsoft Corp., 373 F.3d 1199 (D.C. Cir. 2004) [hereinafter *Microsoft VII*] (remedy appeal).

2. Error costs are the social costs of mistaken decisions. A false positive error cost occurs when a court wrongly finds liability based on conduct that is actually efficient. A false negative error cost occurs when a court wrongly finds no liability based on conduct that is actually inefficient.

3. *See* Frank H. Easterbrook, *Allocating Antitrust Decisionmaking Tasks*, 76 GEO. L.J. 305 (1987); Frank H. Easterbrook, *Comparative Advantage and Antitrust Law*, 75 CALIF. L. REV. 983 (1987) [hereinafter Easterbrook, *Comparative Advantage*]; Frank H. Easterbrook, *Information and Antitrust*, 2000 U. CHI. LEGAL. F. 1; Frank H. Easterbrook, *On Identifying Exclusionary Conduct*, 61 NOTRE DAME L. REV. 972 (1986); Frank H. Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1 (1984) [hereinafter Easterbrook, *Limits*].

as a strictly logical matter, the arguments for this view are surprisingly weak.

A second school of thought holds that error costs are simply one aspect of what courts should consider when determining whether conduct is anticompetitive. On this view, error costs should be calculated and then weighed in the balance, along with the costs and benefits to society of the conduct at issue. Uncertainty regarding such costs might cause one to discount the estimated benefits or harms of a practice, or to discount the advantages or disadvantages of a remedy, but it would not cause one to refrain from analysis. Professor Oliver Williamson advanced this view in direct response to Judge Easterbrook's claim.[4] Williamson's view treats the error cost problem as an aspect of decision theory and seeks to integrate the rationality of decision theory with the economic analysis of the case at hand. I therefore refer to Williamson's view as the integration approach to error costs.[5]

The integration approach to error costs requires close judicial scrutiny of specific business practices. The resulting analysis resembles administrative regulation by judges enforcing an unfair practices statute. Analogizing to law enforcement more generally, the integration approach is "traffic cop antitrust," which pursues even minor infractions unlikely to cause significant harm to competition. The market correction approach is "SWAT team antitrust," in which intervention is reserved only for extreme cases and takes a more extreme form. Each approach is logical relative to certain dispositions toward regulation and certain goals for antitrust policy. Neither approach logically dominates the other.

Perfect information could reconcile the two approaches. Data derived from the market correction approach could be used to set the value of variables needed to perform integrationist analysis. Information is imperfect, however, and even the notion of what constitutes a mistake is debatable. Thus, we have disagreements which cannot be explained either by data or by logic.

The error costs debate was central to several articles written shortly after the government filed suit against Microsoft. In 1999, Professors Salop and Romaine published an important defense of the Justice Department's liability theories. They defended an integrationist approach to error costs, arguing in part that the standard of liability for monopolization should re-

4. Oliver Williamson, *Delimiting Antitrust*, 76 GEO. L.J. 271, 280 (1987).

5. Both approaches to error costs seek to embody their view in doctrine, so it is useful to remember that what is integrated is decision theory and the judicial decision, not a view of error costs and a doctrinal standard.

flect the approach that best minimizes the sum of such costs.[6] Professors Cass and Hylton responded with an article that was skeptical of the government's theories and that defended the market correction approach.[7] Professor Lopatka also argued for the market correction approach in an article discussing *Microsoft* against the background of the Justice Department's 1969 monopolization suit against IBM.[8]

In this Article, I examine the record regarding the most significant monopolization claims in the case, from both the liability and remedy phases of the proceeding, to see what light the end of the case might shed on the error costs arguments made at the beginning. I conclude that the lesson of *United States v. Microsoft Corp.* is that the law should be more averse to false positives (cases finding liability for conduct that is actually welfare-enhancing) than to false negatives (not finding liable conduct that actually reduces welfare). A good way to turn this lesson into doctrine would be to reverse the D.C. Circuit's approach to causation in monopoly maintenance analysis.

I support this thesis by arguing three main points. First, the D.C. Circuit's stated approach to monopolization in *Microsoft* appears consistent with the integration approach to error costs. This fact is seen most easily in the court's four-part test for monopolization liability and in its two-tiered approach to the question of causation. In addition, the lawyers and economists on each side of the case were superb, and the court of appeals' unanimous liability opinion reflected the agreement of some economically sophisticated judges. The case therefore should present a fair test of the integration approach. If the approach did not work in *Microsoft*, something is wrong with it.

Second, the actual analysis in *United States v. Microsoft Corp.* was not as pure an application of the integration approach as one would expect from the D.C. Circuit's stated standard for monopolization liability. In reality, the court balanced little. With respect to several issues, the court engaged in the fine-grained analysis of particular practices one would expect from an integrationist approach, but it did so through relatively categorical statements that seemed better suited to a market correction approach. The tensions between these two approaches and between the court's statement of standards and its application of them explain why the court's opinion so

---

6. Steven C. Salop & Craig R. Romaine, *Preserving Monopoly: Economic Analysis, Legal Standards, and Microsoft*, 7 GEO. MASON L. REV. 617, 659 (1999).

7. Ronald A. Cass & Keith N. Hylton, *Preserving Competition: Economic Analysis, Legal Standards and Microsoft*, 8 GEO. MASON L. REV. 1 (1999).

8. John E. Lopatka, United States v. IBM*: A Monument to Arrogance*, 68 ANTITRUST L.J. 145 (2000).

often seems confused or contradictory. The net result is a sort of hybrid "third way" that appears quite logical at each step, but, like a badly executed pointillist painting, produces an unsatisfactory overall effect.

Finally, as a matter of experience rather than logic, the *Microsoft* litigation provides more support for the market correction approach to error costs than for the integration approach. The case produced a peculiar mishmash of liability and remedies, in which the acts that did the most to reinforce Microsoft's market power were found lawful while the acts found unlawful were effectively trivial. The net result was a tepid tapioca pudding of a consent decree, which almost certainly will do nothing to reduce Microsoft's market power.[9]

At the end of the day, therefore, *United States v. Microsoft Corp.* offers two lessons relative to the error cost debate. The first lesson is that a court should make a choice on the issue. Courts should either engage in the detailed balancing of costs and benefits implied by the integration approach or refuse to impose liability based on ambiguous conduct or conduct that is unlikely to preserve market power, even if the costs of that particular conduct exceed the benefits.

Even if commentators can agree that courts should get off the fence on the error cost issue, of course, there is widespread disagreement as to which way the courts should jump. The second lesson, therefore, is more controversial: *United States v. Microsoft Corp.* supports the view that experience should trump logic where the two conflict. The muddled result in *Microsoft* is consistent with the government's overall record in policing predatory conduct in markets related to general-purpose computing. The record inspires little confidence that antitrust litigation in this market has made us better off.

I. THE ERROR COST DEBATE AND APPLYING *UNITED STATES V. MICROSOFT CORP.* TO THE DEBATE

A. The Market Correction Approach

The market correction approach to error costs claims that mistaken condemnation of competitive conduct is costlier than mistaken acquittals of anticompetitive conduct. Starting with an influential article published

9. I believe this result is good because the case should not have been brought in the first place. I concede that it is perfectly logical to go precisely the other way, however, and to regret that the court of appeals acquitted Microsoft of so much and that the remedy is not stronger.

twenty years ago, Judge (then-Professor) Easterbrook advanced three reasons to accept this claim.

First, most forms of cooperation are efficient, so a judge who refuses to condemn an ambiguous practice is more likely to be right than wrong.[10] Professors Hylton and Salinger extended this point with regard to tying claims, arguing that if most instances of a practice are lawful, then legal review of such practices will enhance welfare only if the standard of review has an extremely low probability of error.[11]

Second, markets are self-correcting, but bad precedents are not. Monopoly power implies supra-competitive profits that, over time, will attract entry that erodes monopoly power.[12] Society may lose in the interim, but it will not lose forever. Mistaken condemnation of a firm's conduct, however, creates precedent that deters other firms from engaging in competitive behavior. Precedent thus creates efficiency losses that ripple through the economy.

Judge Easterbrook's key point here is that judges do not fix bad precedents as surely and as quickly as markets fix (unlawful) market power. Bad precedent does not induce the entry of good precedent, which only judges can provide. Judges are either unlikely to do so (because they like following precedent, which lowers the cost of thinking about new cases) or unable to do so (because they are bound by a superior tribunal or because as a panel they cannot reverse precedent and have to try to invoke the en banc process). Because "mistakes of *law* are not subject to competitive pressures,"[13] condemned practices are likely to stay condemned, even if there is good reason to believe they are actually efficient.

Judge Easterbrook's final argument is that the cost of monopoly wrongly permitted is likely to be smaller than the cost of efficient conduct wrongly condemned. Monopoly's true evil is the deadweight loss that affects only a portion of the demand curve, while efficient conduct may (for example) lower the production cost of every single unit.[14]

---

10. Easterbrook, *Limits*, *supra* note 3, at 15. Judge Easterbrook casts this point in terms of cooperative practices, but his general skepticism of claims of predatory conduct justifies expanding the claim to business practices more generally.

11. Keith N. Hylton & Michael Salinger, *Tying Law and Policy: A Decision-Theoretic Approach*, 69 ANTITRUST L.J. 469 (2001). As Professors Hylton and Salinger put it, "given the ubiquity of beneficial tying, a rational legal rule must have virtually no risk of a false conviction." *Id.* at 502.

12. Easterbrook, *Limits*, *supra* note 3, at 15.

13. Easterbrook, *Comparative Advantage*, *supra* note 3, at 986.

14. Easterbrook, *Limits*, *supra* note 3, at 15-16.

In addition to Judge Easterbrook's points, Professors Cass and Hylton argue that false positives in antitrust litigation produce an undesirable culture of economic political correctness in which firms are rewarded for whining to courts rather than for competing on price or quality. They also claim false positives are costlier than false negatives in dynamic markets where technological change facilitates entry, thus reducing the costs of a false negative.[15]

These are powerful but not irrefutable arguments. I am persuaded that they state the best view, but candor compels a survey of their weak spots. To begin, suppose it is true that most business practices that survive competition for any length of time are efficient. That hypothesis justifies the sort of filters Judge Easterbrook recommends, such as requiring a showing of a high degree of market power as a necessary element to most antitrust violations.

Once relatively few of these filters are in place, however, there is no need to take the set of all business practices as the baseline for measuring the probability of a mistaken decision. There is no reason to believe that tying arrangements are as likely to be efficient when the defendant has substantial power in the tying product market, and the tie might deter entry into the tied product market, for example, as they are in the general run of bundling arrangements.[16] That does not mean tying is ever *likely* to harm competition, of course, but it introduces uncertainty into the baseline probability, which means the argument suffers to some degree from the uncertainty to which it is presented as a solution.

Judge Easterbrook's timing argument also overstates his case. Markets do tend to erode monopoly market power in the long run, but that fact does not establish that they erode market power faster than a combination of market forces plus legal prohibitions on anticompetitive conduct. The ripple effect of errors cuts both ways. Condemning efficient practices in one case deters efficient conduct in the market as a whole, but allowing inefficient conduct in one case invites such conduct in the market as a whole. And if the inefficient conduct produces market power, firms will have every incentive to engage in such conduct, producing wasteful competition for monopoly rents, a point Judge Easterbrook recognizes but does not attempt to incorporate in his analysis.[17]

---

15. Cass & Hylton, *supra* note 7, at 31-33.

16. *E.g.*, Michael D. Whinston, *Tying, Foreclosure, and Exclusion*, 80 AM. ECON. REV. 837 (1990).

17. Easterbrook, *Limits*, *supra* note 3, at 16 n.31.

Nor is it entirely clear that judges are slower to reverse bad precedent than markets are to erode market power. History provides an imperfect test of the claim, because views on what amounts to a "bad" decision have shifted over time. If one wants to protect small dealers, then cases like *Brown Shoe*[18] and *Image Technical Services*[19] are not obviously wrong, though they either impede or do nothing to advance welfare.[20] It is not fair to use efficiency to condemn a process that aims at something else.

Second, though there is evidence of inefficient precedents lasting a long time,[21] there is also evidence that judges neutralize or reverse bad precedents in a reasonable period of time once the judges settle on a policy goal to pursue. As a general matter, if judges never corrected "mistakes," it would make no sense to talk of a Chicago revolution. As a particular matter, consider some examples. *Brown Shoe* was decided in 1962, and *General Dynamics* in 1974.[22] *Arnold Schwinn* was decided in 1968,[23] and *GTE Sylvania* in 1977.[24] A twelve-year or nine-year interval might seem

18. Brown Shoe Co. v. United States, 370 U.S. 294 (1962) (holding unlawful merger part of which involved forward integration of shoe manufacturer into retail distribution of its shoes).

19. Eastman Kodak Co. v. Image Technical Servs., 504 U.S. 451 (1992) (allowing plaintiff to proceed on claim that copier manufacturer tied the provision of service for its machines to the purchase of parts for its machines).

20. The same might be said of *Albrecht v. Herald-Tribune Co.*, 390 U.S. 145 (1968), a bad decision that lasted twenty-nine years, until *State Oil v. Khan*, 522 U.S. 3 (1997). The late 1960s were merciless on efficiency-based antitrust, but at that time the Court still took seriously the notion that, to borrow from Justice Harlan's dissent in *Albrecht*, "one of the objectives of the Sherman Act was to preserve, for social rather than economic reasons, a high degree of independence, multiplicity, and variety in the economic system." 390 U.S. at 158 (Harlan, J., dissenting). On the demise of *Albrecht*, see Roger D. Blair & John E. Lopatka, Albrecht *Overruled—At Last*, 66 ANTITRUST L.J. 537 (1998).

21. *E.g.*, Dr. Miles Med. Co. v. John D. Park & Sons Co., 220 U.S. 373 (1911) (declaring minimum resale price maintenance unlawful *per se*).

22. United States v. Gen. Dynamics Corp., 415 U.S. 486 (1974) (reviewing merger under effectively more lenient, efficiency-friendly approach than had been employed in *Brown Shoe*).

23. United States v. Arnold Schwinn & Co., 388 U.S. 365 (1968) (holding unlawful *per se* manufacturer's imposition of vertical nonprice restraints).

24. Continental TV, Inc. v. GTE Sylvania, Inc., 433 U.S. 36 (1977) (holding that vertical nonprice restraints should be reviewed under the rule of reason). Alan Meese points out that in reversing bad precedents courts generally leave open the possibility that a practice might be reviewed under the rule of reason, rather than declaring the practice *per se* lawful. *See* Alan J. Meese, *Intrabrand Restraints and the Theory of the Firm*, 83 N.C. L. REV. 5, 10 (2004) (arguing for rule of *per se* legality for intrabrand restraints). This point is true, though rule of reason analysis, with its market power requirement, *e.g.*, E. Food Servs., Inc. v. Pontifical Catholic Univ. Servs. Ass'n, 357 F.3d 1, 5 (1st Cir.

like an eternity to a firm pinched by precedent, but these are not intolerably long times for a bad decision to be discarded, especially when one considers that *Brown Shoe* was probably more faithful to Congress's intentions than the cases that came after it. *Eastman Kodak Co. v. Image Technical Services* was (in my opinion) a bad decision, but lower court opinions have essentially gutted it, proving that errors might not be very costly even when they are not reversed.[25]

Just as there are examples of relatively fast judicial turnabouts, there are examples of firms maintaining dominant positions for a long time. Microsoft has held its operating systems (OS) monopoly about as long as *Arnold Schwinn* was good law, and the company looks set to maintain that position for at least a while longer. One could reply, as I would, that where monopoly power is durable it is likely because of economic conditions the law cannot and should not try to remedy.[26] On this view, durable market power does not contradict the market correction thesis, because antitrust law cannot change the underlying economic facts and therefore should not try to change market positions dictated by those facts. This argument confirms the market correction view by suggesting that the market will have to undo the power by introducing new technologies that leapfrog the old. This is a cogent argument, and I believe a right one, but it is a long way from "proving" anything.

Third, antitrust law is essentially common law, and the claim that bad antitrust precedents persist runs into the claim that common-law adjudication tends toward efficient rules.[27] Firms seeking to maximize profits will bump up against inefficient decisions, which will come under renewed scrutiny as parties litigate in an effort to clear the judicial barrier to profit. It would of course be better if the barrier were not there in the first place, but one can see how market competition would put pressure on inefficient

---

2004), weeds out so many cases that the practical rule, though not the doctrine, approximates *per se* legality.

25. *E.g.*, Herbert Hovenkamp, *Post-Chicago Antitrust: A Review and Critique*, 2001 COLUM. BUS. L. REV. 257, 283-99. Cases include PSI Repair Servs., Inc. v. Honeywell, Inc., 104 F.3d 811 (6th Cir.), *cert denied*, 520 U.S. 1265 (1997); Digital Equip. Corp. v. Uniq Digital Techs., 73 F.3d 756 (7th Cir. 1996); SMS Sys. Maint. Serv. v. Digital Equip. Corp., 11 F. Supp. 2d 166 (D. Mass. 1998), *aff'd*, 188 F.3d 11 (1st Cir. 1999), *cert denied*, 528 U.S. 1188 (2000).

26. The large net economies of scale in operating system software are an example.

27. The notion that antitrust adjudication is essentially common law is discussed in many places, including David McGowan, *Innovation, Uncertainty, and Stability in Antitrust Law*, 16 BERKELEY TECH. L.J. 729, 752-54 (2001). The common-law nature of antitrust is linked to the thesis that the common law tends to efficiency in Blair & Lopatka, *supra* note 20, at 549-54.

precedents. One can think of counter-examples, of course,[28] but a long-standing inefficient precedent may be one that does not significantly impede efficient conduct. If transaction costs are low, for example, then at least with respect to vertical conduct the Coase Theorem implies that firms will be able to work around even inefficient rules, a result the Court alluded to when overruling *Albrecht*.[29] Either way, history does not justify the categorical claim that judges cannot correct or ameliorate error.

Nor can we take for granted that monopolistic activity creates losses across only a portion of the demand curve while foregone efficiencies affect the full range of production. Suppose a court can determine that a dominant firm's anticompetitive acts deter entry by a technology that is superior on the merits to an incumbent technology, or that such acts fragment the technology so that it fails to generate the positive feedback that protects the incumbent. In that case all users of the incumbent's products would be stuck with inferior technology, and losses represented by the difference in utility between the incumbent and the entrant would extend across the full range of demand.

This last point also applies to the claim that in dynamic markets, entry is likely to erode market power rapidly. Technology changes rapidly in computers, but first IBM and now Microsoft have held fairly stable market positions for relatively long periods of time. The minicomputer industry (DEC and Data General) provides a possible counter-example to this history of stability, but the history of market power persisting during periods of rapid innovation contradicts any claim that such innovation necessarily undermines market power. (That history also may suggest that such power is not socially harmful.) I agree emphatically with Professors Cass and Hylton that overly aggressive antitrust liability rules give firms an incentive to whine rather than compete, but one must acknowledge that a too-lenient liability standard might reduce welfare by, for example, deterring entry by firms that would compete if the antitrust laws more closely policed the responsive conduct of incumbent monopolists such as Microsoft.[30]

---

28. *See supra* note 20.

29. State Oil v. Khan, 522 U.S. 3, 18 (1997) ("[M]anufacturers and suppliers appear to have fashioned schemes to get around the *per se* rule against vertical maximum price fixing."). That transaction costs are low does not mean they are absent, so it would still be better to have a sensible rule than to force (many) parties to incur (even small) costs to contract around the inefficient precedent.

30. Logically possible, even though a long history of entry, incumbent responses, and more entry, casts serious doubt on whether the logic produces a sound rather than

B. The Integration Approach

The integration approach to error costs is the main competitor to the market correction approach. The integration approach views error costs as simply one variable to be included in the analysis of a particular case. On this view, a court should not resort to categorical presumptions about error costs or particular types of conduct. The court should instead analyze an allegedly anticompetitive practice on the merits, including in this analysis an analysis of the risk and magnitude of errors one way or another.

Responding to Judge Easterbrook, Professor Williamson defended this approach as a "legal process" analysis, which he distinguished from the "legal rules" approach he attributed to Judge Easterbrook. Professor Williamson felt the legal rules approach was too sure of itself and too unwilling to take into account new learning about the potential anticompetitive effects of strategic behavior. There might be limits on the theories courts could handle, he agreed, but these were not fixed. Over time the law might improve its understanding of strategic behavior, allowing judges to review conduct they previously left alone, and presumably enhance efficiency. As Professor Williamson described it,

> The approach to antitrust and economic regulation proposed herein recommends that administrability considerations be factored into the overall rationality analysis of the issues. The rules in force at each point in time would thus be required to pass an administrability test, but provision would be made to successively improve the rules upon refining the relevant theory and out understanding of complex phenomena. Rationality and the needs of the legal process are thereby joined.[31]

Moreover, in a footnote to this paragraph, Professor Williamson cited an article by Professors Michael Katz and Carl Shapiro that discusses principles that played an important role in the *Microsoft* litigation.[32] In part for this reason, *United States v. Microsoft Corp.* offers a fair test of the ap-

---

merely a valid argument. *See* F.M. Scherer, *Antitrust, Efficiency, and Progress*, 62 N.Y.U. L. REV. 998, 1017 (1997).

31. Williamson, *supra* note 4, at 280 (footnote omitted).

32. *Id.* at 280 n.38 (citing Michael Katz & Carl Shapiro, *Technology Adoption in the Presence of Network Externalities*, 94 J. POL. ECON. 822, 835, 840 (1986)). Briefly, Professors Katz and Shapiro discuss the risk that in network markets there may be too little standardization or too much (lock-in), that standardization might happen too quickly or too slowly, and that a firm holding intellectual property rights in a standard might sponsor optimal standardization (relative to open competition among standards) or suboptimal standardization.

proach Professor Williamson advocated. I will return to this point in Part III.

In context, Professor Williamson's message to Judge Easterbrook was clear. Just as there should be no "ratchet" in antitrust doctrine,[33] there should be no ratchet in antirust economics. It would be at best churlish, and at worst an admission that the Chicago School was more about ideology than efficiency, to declare that economics was useful (or even mostly useful) when it helped defendants but not when it helped private plaintiffs or the Department of Justice. As a matter of logic and economic theory, one has to concede that Professor Williamson was right, and that his not-so-implied charge was a serious one.

A little over ten years later, Professors Salop and Romaine advanced a similar argument. They maintained that error cost analysis should be understood with regard to decision theory in general, and that the risk and magnitude of error should be integrated into, and therefore expressed by, the standard of liability. If the expected net harm of false condemnation exceeds the expected net harm of false acquittal, then the liability standard should tilt against a finding of liability, and vice versa.[34] Parties would fight about error costs as they fight about everything else, and trial courts would make findings on those issues that would influence their ultimate liability decisions.[35]

This approach led Professors Salop and Romaine to endorse what they called the "unnecessarily restrictive conduct" test for monopolization liability. Under this test, courts weigh conflicting arguments regarding the motives and effect of a practice in order to determine the primary effect of the practice on consumers.[36] Salop and Romaine faulted alternative tests for being too one-sided to take into account all the relevant economic effects and the risks of both false positive and false negative decisions.[37] As

---

33. *Cf.* Frank H. Easterbrook, *Is There a Ratchet in Antitrust Law?*, 60 TEX. L. REV. 705 (1982).

34. Salop & Romaine, *supra* note 6. Professors Salop and Romaine say the standard should tilt "marginally" one way or the other. *Id.* at 654. Ideally the degree of bias in the standard would offset precisely the expected cost of error, but there is no way to know any of these magnitudes with any certainty, and no real way to convert the analysis of models into a general standard expressed in English.

35. *Id.* at 655.

36. *Id.* at 659.

37. Salop and Romaine identified these tests as 1) the "avoidable exclusionary conduct" test, which condemned conduct that had an exclusionary effect, regardless of its benefits, so long as the conduct could be avoided, and 2) the "sole purpose and effect" test, which refuses to condemn conduct for which there is any efficiency justification, even if the harms from the conduct outweigh the benefits. *Id.*

we will see in a moment, the D.C. Circuit's *Microsoft* opinion essentially adopts this view in its standard of liability for monopolization.

C. Assessing the Approaches

Arguing about error costs is a time-honored way of dealing with uncertainty. If you cannot actually demonstrate that your thesis is correct, you can always fall back to the position that it would be better to err in your favor than to err the other way. Because the debate concerns what we should do when we do not know what to do, it is no surprise that it is inconclusive.

Perhaps predictably, disputants sometimes conduct the error-cost debate by arguing about the burden of proof (arguments about error costs in the error-cost debate presumably being too obviously circular). Thus Professor Williamson, who is to a degree sympathetic with Judge Easterbrook's concern over error costs, argues that "no one has provided a demonstration that the cost differences are as Easterbrook indicates. Easterbrook has an undischarged burden of proof that the cost of false positives in the market power region where strategic behavior is implicated is similarly low."[38]

It is true that no one has proved that the market correction approach to error costs is correct. It is also true that no one has proved it incorrect, or that anyone has proved that the integration approach works better in practice. Nor is it clear where the undischarged burden of proof comes from, unless it is just the ordinary burden facing anyone asserting any proposition, in which case one could say that the integration approach has an undischarged burden of showing that it works as well in practice as it does in logic.[39] Even to the extent one relies on experience to bolster one's position, as I do here, there is always the problem (familiar from Hume and Popper) that we cannot prove that past will be prologue, we can only falsify.

Having said all that, as a purely logical matter the integration approach is obviously correct. In fact, it is hard to see why there should be a difference between the two approaches. If and to the extent the facts asserted in favor of the market correction approach are true, then it should be possible to estimate the expected cost of error those facts imply. That estimate could then be plugged into a standard cost-benefit analysis of a case as a whole, including the characteristics of the institution in which it is liti-

38. Williamson, *supra* note 4, at 289.

39. Professor Lopatka rightly says that it is hard to see how one could *prove* that the market correction approach is best. Lopatka, *supra* note 8, at 151. The same is true of the integration approach.

gated, which is the aim of the integration approach. If that analysis were performed properly, and if all else is equal, the two approaches should produce the same results. On the other hand, if information is poor across the board, so there is no evidentiary basis for favoring the market correction approach, it might be impossible to implement the integration approach. But even then the two approaches would not conflict; instead, they both would simply fail.

There is a conflict between the approaches, however, and it is significant. The balancing that the integration approach favors implies that judges will intrude more into business practices and decisions than they would if they adopted the relatively categorical risk aversion of the market correction approach. The integration approach implies traffic cop antitrust, while the market correction approach implies SWAT team antitrust. If one has faith in governmental regulation and believes judges or enforcement officials can do a good job of balancing the costs and benefits of business conduct, the integration approach will seem obviously correct. If one is skeptical of governmental regulation and the precision of cost-benefit analysis, the market correction approach will seem a wise judgment amply backed up by experience.

This debate will never end. Neither is it completely indeterminate, however. The competing views can be tested in light of actual litigation experience. *United States v. Microsoft Corp.* provides a good test case because the standard for liability the D.C. Circuit articulated in *Microsoft* is similar to the unnecessarily restrictive conduct test Professors Salop and Romaine advocated as integrating error cost analysis with economic analysis and with antitrust doctrine.

Under the D.C. Circuit's test, a defendant is liable only for acts that harm the competitive process itself, rather than acts that merely harm competitors. If a plaintiff makes out a prima facie case that the defendant has harmed competition, the defendant may rebut that case by advancing a "nonpretextual" claim that its conduct is in fact a form of competition on the merits, meaning that its conduct has efficiency benefits. The plaintiff then bears the burden of either rebutting this justification—which leaves the plaintiff's original case standing—or showing "that the anticompetitive harm of the conduct outweighs the procompetitive benefit."[40] The weigh-

40. *Microsoft IV*, 253 F.3d 34, 58-59 (D.C. Cir. 2001). Although in this portion of the opinion the court said that a plaintiff might either rebut a justification or show that the harms of a practice outweighed the benefits, the court later said that "the plaintiff bears the burden not only of rebutting a proffered justification but also of demonstrating that the anticompetitive effect of the challenged action outweighs it." *Id.* at 67. These two

ing aspect of the D.C. Circuit's test, which the court saw as a rule of reason inquiry, tracks the balancing of Professor Salop and Romaine's test.[41] The back-and-forth between the prima facie case, justification of conduct, and rebuttal of that justification may be seen as a way of reducing error costs by testing in detail claims of both harms and benefits to competition.

In addition, the Justice Department's case in *United States v. Microsoft Corp.* rested upon a highly sophisticated and logically rigorous economic analysis. The government's economics experts included Professor Franklin Fisher, who had worked for IBM against the government in the 1970s, and Professor Carl Shapiro, who did path-breaking work in the analysis of network markets. David Boies tried the case for the government, and the non-settling states hired Williams & Connolly, one of the very best litigation firms in the country, to pursue the remedies they sought. Professor Mark Lemley advised the government on antitrust and intellectual property issues. Thus, one could not ask for a better team to present any theory, including the post-Chicago theories underlying the government's case. If an approach performs poorly with such a team behind it, there is reason to question the approach itself.

United States v. Microsoft Corp. is not a perfect case for testing the competing approaches, however. As I noted at the outset, and discuss in some detail in Part III, the court's balancing test did not produce much balancing. The opinion provides a messy, hybrid, hair-splitting analysis that, I suspect, satisfies no one, and rightly so. Rather than interpreting this fact as a reason to ignore the case, however, I treat it as a test of whether the logical elegance of the integration approach can survive the messy uncertainties of real-world litigation. *United States v. Microsoft Corp.* suggests that it cannot.

This is a relatively strong claim. Because it argues for the triumph of experience over logic, I begin my argument by placing *United States v. Microsoft Corp.* in the long and unhappy context of the Justice Department's antitrust involvement in general purpose computing. I examine that history in the next Part, and turn to key aspects of the *Microsoft* litigation in Part III.

---

formulations may amount to the same thing. A plaintiff who has shown genuine harm and also rebutted a proffered justification probably will have shown net competitive harm.

41. *Id.* (contradicting Professors Cass and Hylton's view of the appropriate standard).

II. BRIEF HISTORY OF APPLYING ANTITRUST ANALYSIS TO COMPUTER AND COMPUTER SOFTWARE MARKETS

Computer and computer software markets have troubled antitrust analysis for decades. Innovation is rapid in such markets, as is the need for close integration of complementary technologies. Fixed costs are high, and most of them must be sunk before production begins, while variable costs can be low (especially for software), so economies of scale on both the supply and demand side (network effects) can be large.

Historically, these factors often have combined to produce markets dominated by a single firm. Antitrust ideology and economic understanding have changed significantly over time, but judges and government enforcers have always viewed such markets with suspicion. That suspicion has led to almost continuous antitrust challenges to the dominant firm by the Justice Department and private plaintiffs. As this Part shows, there is no particular reason to believe this history has made society better off. (It has amortized the debts of generations of law students, but that is a different thing.)

This Part discusses the government's history with regard to IBM and AT&T. It also discusses the relationship between the AT&T history and the growth of F/OSS development and production practices. We do not know yet whether those practices will play an important role in antirust litigation in markets related to general-purpose computing. F/OSS development made an appearance in the *Microsoft* remedy proceedings, however, and it may be relevant to future cases as well. I discuss its history here, and its relation to the *Microsoft* litigation in Part III. The history shows that theories of liability are very similar over time; that litigation tends to lag market developments that call those theories into question; that unintended consequences are likely but hard to assess; and that there is no reason to think we are better off because of these cases than we would be had they not been brought.

A. IBM

The Justice Department has sued IBM three times. It first attacked IBM's policy of requiring that firms that leased IBM's tabulating hardware also purchase its punchcards. That policy easily can be defended on efficiency grounds: it allowed IBM to discriminate in price between high-volume and low-volume users of its machines. Firms had leased machines and sold cards from the time tabulation machines entered the market, before IBM was even founded, which is good evidence that the practice was

efficient.[42] Nevertheless, the government alleged that IBM tied the sale of its cards to the lease of its machines, and it won the case, which concluded in 1936.[43]

IBM continued its policy of leasing tabulating machines, and the Justice Department attacked that policy directly in a suit filed in 1952. Unfortunately, the government's timing was poor. Two years earlier, IBM's old punch-card rival Remington-Rand had introduced a fully electronic computer. IBM, which the government attacked as an entrenched monopolist, was about to enter a period of robust competition in which it would, for a time, play catch-up. The government's leasing suit settled in 1956 with a consent decree requiring IBM to sell machines as well as lease them. The decree also covered the emerging computer market.[44] In 1994, IBM moved to terminate the decree. The government eventually agreed to terminate the decree over time, and termination was approved in 1998.[45]

The government filed its third suit in 1969. It alleged that IBM had monopolized the market for general-purpose computers—the market the government had not seen coming in 1952. The third case dragged on until 1982, when the Justice Department, led by Assistant Attorney General William Baxter, moved to dismiss it as having no merit.[46] From an administrative point of view, the suit was a Dickensian nightmare.[47] Judge Bork memorably called it the Justice Department's Vietnam.[48]

A particularly important feature of this history is that the Justice Department never could keep up with the market. The Justice Department went after tabulating machines the year after general purpose computers were introduced, and it doggedly pursued IBM through a period in which the fundamental structure of competition changed. For most of the history of general purpose computing, the dominant firm was vertically inte-

42. PETER SALUS, A QUARTER CENTURY OF UNIX 15 (1994).

43. IBM Corp. v. United States, 298 U.S. 131 (1936).

44. United States v. IBM Corp., 1956 Trade Cas. (CCH) ¶ 68,245 (S.D.N.Y. 1956).

45. United States v. IBM Corp., 163 F.3d 737 (2d Cir. 1998).

46. *In re* IBM Corp., 687 F.2d 591 (2d Cir. 1982).

47. Lopatka, *supra* note 8, at 145. The statistics include, as Professor Lopatka puts it, "700 trial days over the course of nearly seven years, preceded by six years of discovery; 87 live witnesses; 860 deposition witnesses (whose testimony was read aloud to an empty bench, a process that consumed 70 trial days); 104,400 trial transcript pages; 17,000 exhibits." *Id.* at 145.

48. ROBERT H. BORK, THE ANTITRUST PARADOX: A POLICY AT WAR WITH ITSELF 432 (Rev. ed. 1993).

grated.[49] IBM's series 360 and 370 machines, featuring IBM processors, operating systems, applications, and storage devices, competed with those of firms such as Sperry Univac and DEC. IBM responded to firms that sought to make IBM-compatible peripheral devices, such as disk drives, in part by integrating its own make of those devices into its machines.[50]

Things began to change in the 1970s. In the mid-1970s several firms introduced microcomputers, most notably Apple with its Apple II. Production in this market was modular, not integrated.[51] One firm made a chip, another wrote an operating system for it (such as Microsoft's initial product, Altair BASIC), and a third provided memory. One of these firms, or even the user herself, could put the pieces together. Competition shifted from a horizontal interface, in which firms sought to plug into a vertically integrated proprietary system, to a vertical interface, in which different firms sought to plug into the layers above and below them.[52] IBM was a relatively late entrant into this market, and it hastened its entry by adapting to the modular production model that already characterized this market, rather than the integrated, proprietary model the company had pursued in the past.[53]

The Justice Department fell on its sword in the 1969 monopolization case just as developments began to occur that ultimately led to Microsoft displacing IBM in many software markets. Microsoft became an operating system vendor for IBM's personal computer by purchasing from Seattle Computer a program called QDOS (Quick and Dirty Operating System), which it modified into MS-DOS.[54] Microsoft priced MS-DOS aggres-

---

49. *See, e.g.*, Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Open Access Policies: Towards a Convergence of Antitrust and Regulation in the Internet Age*, 17 HARV. J.L. & TECH. 85, 92 (2003).

50. That integration was the basis of part of the *Telex* litigation and later of the government's monopolization case. *See* Cal. Computer Prods., Inc. v. IBM Corp., 613 F.2d 727 (9th Cir. 1979); Innovation Data Processing, Inc. v. IBM Corp., 585 F. Supp. 1470, 1476 (D.N.J. 1984); ILC Peripherals Leasing Corp. v. IBM Corp., 448 F. Supp. 228, 233 (N.D. Cal. 1978), *aff'd per curiam sub nom.* Memorex Corp. v. IBM Corp., 636 F.2d 1188 (9th Cir.1980); Telex Corp. v. IBM Corp., 367 F. Supp. 258, 347 (N.D. Okla. 1973), *aff'd in relevant part*, 510 F.2d 894 (10th Cir. 1975).

51. *See* Richard N. Langlois, *External Economies and Economic Progress: The Case of the Microcomputer Industry*, 66 BUS. HIST. REV. 1, 9-18 (1992).

52. *See, e.g.*, Timothy F. Bresnahan & Shane Greenstein, *Technological Competition and the Structure of the Computer Industry*, 47 J. INDUS. ECON. 1, 1-2 (1999).

53. *Id.* at 20-23.

54. In connection with its reversal of the district court's order breaking Microsoft into an applications firm and an operating system firm, the D.C. Circuit emphasized that, unlike other firms that had been broken up through antitrust consent decrees, Microsoft had grown internally rather than through corporate acquisitions. *Microsoft IV*, 253 F.3d

sively, and most purchasers of the IBM PC chose the MS-DOS operating system. By the early 1990s, Microsoft was the dominant platform firm in the personal computer market.

It may not be fair to judge the Justice Department's IBM experience from an efficiency point of view. The Department was involved with IBM from 1952 continuously (if one counts the 1956 consent decree) through 1998, and antitrust theory and doctrine changed quite a bit during that period. To the extent one cares about efficiency, however, the Justice Department's experience offers no reason to believe that the government's suits enhanced welfare. The IBM experience does not undermine the integration approach because many of the practices the Justice Department attacked might well have come out differently under the 2001 version of the approach. Neither does the experience provide any reason for optimism, however. If only because the costs of these efforts were high, and the benefits low (and perhaps nonexistent), the IBM experience provides solid anecdotal support for the market correction approach.

B. AT&T/Unix

In 1949, the Justice Department sued AT&T and Western Electric Co. The complaint alleged "the defendants had monopolized and conspired to restrain trade in the manufacture, distribution, sale, and installation of telephones, telephone apparatus, equipment, materials, and supplies, in violation of [the antitrust statute]."[55] The complaint sought to divest AT&T of its ownership of Western Electric, and to divest Western Electric of its fifty percent interest in Bell Laboratories, a research center it co-owned with AT&T.[56] After much lobbying, the Justice Department and AT&T resolved the case by entering into a consent decree in 1956. The decree did not require divestiture or other structural relief, but it "precluded AT&T from engaging in any business other than the provision of common carrier communications services."[57]

Between 1969 and 1975, engineers at Bell Labs, most prominently Dennis Ritchie and Ken Thompson, developed a computer operating system that became known as Unix.[58] Developers were interested in Unix but,

---

34, 105-06 (D.C. Cir. 2001). That is a fair point, though one could argue that Microsoft's acquisition of rights to products such as QDOS presents a more complex picture than the court of appeals recognized.

55. United States v. AT&T, 552 F. Supp. 131, 135 (D.D.C. 1982).

56. *Id.* at 136.

57. *Id.* at 138.

58. For a brief history, see Dennis M. Ritchie, *The Evolution of the Unix Time-Sharing System* (1984), *available at* http://cm.bell-labs.com/cm/cs/who/dmr/hist.html. Unix followed from frustration with a project involving GE, Bell Labs, and MIT to de-

under the 1956 consent decree, AT&T could not enter lines of business other than providing common carrier communications services. AT&T's counsel concluded that this restriction did not allow AT&T to exploit Unix. The decree also had required AT&T and Bell Labs to license patents on a nondiscriminatory basis.[59] AT&T counsel concluded that this provision implied an obligation to license Unix.[60] AT&T therefore decided that it could not distribute Unix for commercial purposes but would license it for academic and research purposes.[61]

AT&T licensed Unix to various universities under academic licenses. The most important licensee was the University of California at Berkeley. Thompson spent 1975 teaching at Berkeley and, over the next few years, students and faculty at Berkeley developed tools (complements) for Unix. By 1977, Berkeley was distributing these complements to existing AT&T licensees. Berkeley's distributions were known as "BSD," which stood for Berkeley Software Distribution.[62] This development work produced the BSD license, which is one of the more prominent of the more than forty "open source" licenses.

Beginning around 1979, AT&T realized that Unix had commercial potential. Its lawyers presumably found a way to allay their earlier concerns, and AT&T began to restrict distribution of Unix code and charge significant license fees.[63] By 1979, however, the Unix developers at Berkeley had gone beyond the development of tools. They began modifying the operating system itself.[64] In 1980, Berkeley received a contract from the Defense Advanced Research Projects Agency (DARPA) to adapt Unix for the Agency's ARPANET project, which was a precursor to the Internet. The contract accelerated Berkeley's work on adapting Unix for networking purposes (most importantly by incorporating TCP/IP into Unix).[65] It also set BSD on a path that would eventually diverge from AT&T.

---

velop a system called Multics, which stood for "Multiplexed Information and Computing Service." "Unix" is a pun on "Multics." The standard reference work is SALUS, *supra* note 42. *See also* Marshall K. McKusik, *Twenty Years of Berkeley Unix: From AT&T-Owned to Freely Distributable*, *in* OPEN SOURCES: VOICES FROM THE REVOLUTION (Chris DiBona et al. eds., 1999).

59. *AT&T*, 552 F. Supp. at 136.

60. SALUS, *supra* note 42, at 58-59; STEVEN WEBER, THE SUCCESS OF OPEN SOURCE 28 (2004).

61. SALUS, *supra* note 42, at 59.

62. *Id.* chs.16, 18; McKusik, *supra* note 58.

63. WEBER, *supra* note 60, at 38. The license fee was $60,000 per machine for commercial installation and $7,500 per university for an academic license.

64. SALUS, *supra* note 42, at 222-23.

65. *Id.* at 163-72; WEBER, *supra* note 60, at 33-34.

In 1982, the same year it abandoned its 1969 monopolization case against IBM, the Justice Department settled its 1974 antitrust case against AT&T.[66] This settlement took the form of a modification of the final judgment from the 1956 action but, unlike the 1956 decree, this settlement fundamentally altered AT&T's structure by divesting AT&T of its Bell Operating Companies. AT&T retained Bell Laboratories, however. Freed from the line of business restriction in the 1956 decree, AT&T tried to exploit Unix harder.[67] Licensing fees went up, reaching $100,000 for commercial installations by 1988.[68] In 1989, AT&T created Unix System Laboratories (USL) to pursue its Unix business.[69]

By that time, however, the cat was out of the bag. BSD version 4.3, released in 1986, incorporated the TCP/IP stack. Demand for this system was high, and it roughly coincided with AT&T's newfound freedom to exploit Unix and its aggressive attempts to do so. Berkeley Unix still contained AT&T code, however, which gave AT&T the effective right to limit BSD distribution to entities that had the ever-costlier AT&T Unix license.[70] Berkeley responded by distributing its networking code, to which it had exclusive rights, apart from the components that still included AT&T code. In 1989, Berkeley issued "Networking Release 1," under a license that allowed users to make any use of the code they wished, including incorporating the code in commercial applications, so long as they credited Berkeley and left intact Berkeley's copyright notices.

Having built and released its own networking components, it was a logical step for Berkeley to work downward and write its own version of Unix, which did not include AT&T code and which therefore could be distributed to anyone, regardless of whether they held an AT&T license. In 1991, Berkeley issued "Networking Release 2," which was an almost-finished Unix-like operating system.[71] This release produced a complete system called 386/BSD, and eventually three variants: NetBSD, FreeBSD, and OpenBSD. It also became the basis for a commercial Berkeley Unix

---

66. Also in 1982, one of the Berkeley graduate students working on Unix, Bill Joy, moved to Sun Microsystems, which licensed Unix from AT&T and then modified that code based on BSD improvements starting with BSD version 4.1C. SALUS, *supra* note 42, at 199. Sun went on to play a key role in the Microsoft antitrust litigation as the proponent of the Java technologies, which the Justice Department cited as a potential substitute for Microsoft's Windows operating system. *See infra* Part III.A.

67. WEBER, *supra* note 60, at 38-39

68. *Id.* at 39.

69. *Id.* at 39-40.

70. SALUS, *supra* note 42, at 210, 222-23.

71. *Id.* at 222-23.

business called BSDI, which began selling its BSD-based system in 1992.[72]

In 1992, AT&T's USL sued BSDI, and then the University of California, for infringing AT&T's rights in Unix. In 1993, AT&T sold USL to Novell, which quickly settled the suit for an agreement by Berkeley not to distribute certain disputed files, and permission by AT&T to allow other disputed files to remain in the BSD distribution. In 1995, Novell sold its Unix business (and, depending on whom one believes, its rights in Unix) to the Santa Cruz Operation (SCO).[73] In 2003, SCO would later assert those rights against, of all firms, IBM, in an indirect assault on the GNU/Linux OS.[74] In 2004, Novell would re-enter the Unix business by acquiring SuSe, the leading German distributor of GNU/Linux.

The variety of Unix systems significantly reduced the value of Unix. This point is important in assessing the *Microsoft* litigation, so it is worth quoting Professor Weber's description of the problem Unix encountered in the decade after 1983:

> The proliferation of partly compatible or incompatible hardware and software was daunting. Apollo, DEC, Integrated Solutions, NSC, and other companies built further versions of BSD. AT&T, Altos, Apollo, Compaq, HP, IBM, Intel, Microsoft, Silicon Graphics, and others had AT&T System 5 derivatives. Cray, DEC, Data General, Motorola, and Unisys offered proprietary Unix systems, most of which were based on 4.2 BSD. The Unix "market" was a mess.
>
> These differences led predictably to pressures for standardization of some kind. Unix user groups were some of the earliest de facto standard bodies for operating systems. But they were fighting an uphill battle. The user groups had no official authority and inadequate market power in a traditional sense. . . . The deeper disagreements came over deciding when, in fact a particular area should be standardized—that is, when innovation was locked in

72. WEBER, *supra* note 60, at 51, 95-97.

73. SCO had originally been formed to port Unix to Intel's x86 chip, around the same time Microsoft introduced its own x86 Unix port, called Xenix. Microsoft licensed Xenix to software developers, not hardware Original Equipment Manufacturers (OEMs), and to SCO, which produced its first x86 Unix product in 1983. Microsoft later sold Xenix to SCO while retaining an interest.

74. *See* DAVID MCGOWAN, SCO WHAT? RHETORIC, LAW, AND THE FUTURE OF F/OSS PRODUCTION (Minn. Legal Studies Research Paper No. 04-9, 2004), *available at* http://ssrn.com/abstract=555851.

> to a desirable pathway and the overall Unix market should try to standardize on it.[75]

As we will see in Part III, the fragmentation issue and the tension between accepting a certain state of technology as a standard, and pushing for improvement before agreeing to standardize,[76] were central to the *Microsoft* litigation. In that case, the Justice Department complained that Microsoft had thwarted Sun's Java standardization efforts. It is therefore interesting to note that, with regard to Unix, Sun itself created significant fragmentation worries.

In late 1987, AT&T bought twenty percent of Sun and announced that Sun would receive preferential treatment as AT&T and USL developed Unix.[77] Unix licensees other than Sun worried that this alliance would mean the end of collaborative Unix development and would give Sun a leg up in the marketplace. A group of firms with a stake in Unix development banded together to defend the idea that Unix should develop as an "open" standard. This group eventually formed the Open Software Foundation (not to be confused with the Free Software Foundation, which we will encounter in a moment) to promote an open Unix standard. AT&T and Sun responded by forming Unix International. Each organization eventually had about 200 members, but each foundered in the recession of 1991 and 1992, about the time Linus Torvalds began work on Linux.[78]

Opinions might vary on whether the Justice Department's 1956 consent decree was good or bad for Unix development. One could tell a story in which Unix thrived because AT&T could not exploit it, and therefore essentially gave it to the developer community for five crucial years. On this account, the line of business restriction allowed a new product to flourish, which in turn helped build the market for general purpose computers. The moral of this story would be that antitrust may play a useful role in preventing dominant firms from leveraging their way into new markets. Alternatively, one could tell a story in which Unix fragmented because AT&T could not orchestrate its development until it was too late. On this account, AT&T might well have produced a coherent operating system in an even shorter period of time if the profit motive had been allowed free reign. The moral of this story is that antitrust intervention can

---

75. WEBER, *supra* note 60, at 44. For an even more extensive list of Unix vendors, see SALUS, *supra* note 42, at 210.

76. *See generally*, Katz & Shapiro, *supra* note 32.

77. SALUS, *supra* note 42, at 216.

78. *Id.* at 217-18; WEBER, *supra* note 60, at 44-45.

disrupt the coordination essential to complex markets characterized by virtual network effects.

Which story is right? Neither, if by "right" one means indisputably correct. Either, if one means logically coherent and factually plausible. What is clear about the AT&T/Unix experience is that the Justice Department did not intend to have the effect it had. Whichever narrative one prefers, the Justice Department did not have Unix in mind when it negotiated the 1956 consent decree, nor when it filed suit again in 1974, as Unix was making its way out of Bell Labs and into the university community. The Justice Department did not have Unix in mind when it settled with AT&T in 1982, which let AT&T know that it could begin exploiting Unix in 1984, when the line of business restrictions ended. Whether what it did was good or bad, the Justice Department did not mean to do it. This fact supports neither the market correction approach nor the integration approach. It serves only as a reminder that unintended consequences count, which is an important point that does nothing to resolve concrete disputes.

C. GNU/Linux

In 1984, the year AT&T was freed from its line of business restrictions, a programmer named Richard Stallman left MIT's Artificial Intelligence Lab and founded the Free Software Foundation. Stallman objected to commercial software practices, under which rights holders restricted the ability of licensees to modify the source code of a program (or even obtain it) and to distribute the modified versions to others. He decided to combat those practices, and he left MIT to avoid the risk that the university would own any code he wrote. Stallman's initial project, which he called the "GNU" project, was to write a version of Unix that would be free from AT&T's claims.[79]

Stallman planned to release this operating system under a "General Public License" (GPL). The GPL, now in its second version, guarantees that licensees receive the source code to a program and gives licensees the right to modify and redistribute the program, so long as they distribute it under the same terms on which they received it.[80] In other words, a GPL

79. The history is recounted in David McGowan, *Legal Implications of Open-Source Software*, 2001 UNIV. ILL. L. REV. 241, 261-62. *See also* WEBER, *supra* note 60, at 46-47. As Unix was a play on Multics, GNU was a recursive abbreviation for "GNU's Not Unix."

80. Open Source Initiative, The GNU General Public License (GPL) (1991), *at* http://www.opensource.org/licenses/gpl-license.php.

licensee may modify and distribute code so long as his modified distribution gave others the rights to modify and distribute his code, too.[81]

The GPL uses copyright law to subvert copyright's normal operation. Copyright grants rights to authors or original works fixed in a tangible medium. Commercial authors use the default right to exclude in order to extract from users payment and agreement to limitations on use the author wishes to impose. Among these rights are the rights to make works based on the original work (derivative works),[82] so the default rights structure extends to modifications, too.

As a sociological matter, a developer who releases code under the GPL invites other developers either simply to run the code or to collaborate with him in improving the code. The GPL gives the initial developer, and each subsequent improver, confidence that no one will appropriate his work for their own profit. At a minimum, people who use but do not improve the code expand the installed base of the code and therefore, to the extent the code is subject to network effects, confer positive externalities on other users, including the developers. Persons who improve the code provide their improvements and enhance the sense of community and collaboration that may be associated with a project (though it need not be).

It bears repeating that F/OSS development rests squarely on copyright law. Without copyright, there is no copyleft.[83] F/OSS development is a far cry from the property-less Nirvana it is sometimes described as being. Thus far, however, it has proved to be an effective way of producing code that competes with more conventional software in some significant commercial markets.

Free or open-source development practices entered the *Microsoft* litigation in the form of the GNU/Linux operating system. We will examine its relevance to the litigation in the Part III, but its history is better covered here. As noted above, Unix forked repeatedly as developers cloned different versions of it. One of the clones, developed by Professor Andrew Tenenbaum in 1987, was called Minix.[84] In 1991, Linus Torvalds began work in Minix on a Unix-like operating system (or, depending on whom

81. For a more detailed explanation, see MCGOWAN, *supra* note 74.

82. 17 U.S.C. § 106(2) (2000).

83. A phrase I borrow from Professor Eben Moglen, who is general counsel to the Free Software Foundation.

84. WEBER, *supra* note 60, at 100.

you believe, an operating system with Unix code in it), which eventually became known as Linux.[85]

Emerging just as the Open Software Foundation and Unix International collapsed, Linux was a breath of fresh air for Unix-style development. Torvalds licensed his code under the GPL, shared it freely, and revised it frequently, thereby giving developers rapid gratification either in terms of the evolution of the operating system, development of a developer community, reputational capital as expert programmers, or one or more of the above.[86]

This model has been tremendously successful. In 1993, a firm called Red Hat was founded to make it easier for consumers to install and configure Linux.[87] In 2000, IBM announced support for Linux, as part of a bet that Linux would alter the paradigm of computing in which Microsoft had displaced IBM as the dominant platform firm, with IBM making money as a services and consulting firm working on a commoditized, open-source operating system.[88]

At the time of the Microsoft trial, the GNU/Linux OS was credible enough to appear as a potential entrant to the operating systems market. Both sides tried to use the prospect of entry to their advantage. In Microsoft's view, that prospect constrained its market power, and thus furthered its claim that it was not a monopolist. In the Justice Department's view, GNU/Linux would be a desirable platform to which an applications firm might port Microsoft office if Microsoft were broken up into two firms.

Because the GNU/Linux operating system played a significant role in the *Microsoft* litigation, it is important to remember two things about the model that produces that system. First, notwithstanding a great deal of discussion about "distributed" production, the GNU/Linux system is maintained through a fairly narrow hierarchy.[89] If there is a dispute regarding

85. Strictly speaking, Linux comprises only the kernel, or core, of the operating system, which relies heavily on GNU programs as well. This is why I refer to it as the GNU/Linux operating system.

86. The literature discussing the motivation of open-source programmers is quite large. An important paper modeling motivation as rational within a reputational payoff game is Josh Lerner & Jean Tirole, *The Simple Economics of Open-Source* (Feb. 2000), *at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=224008. For a survey of the literature, see WEBER, *supra* note 60, ch.5.

87. WEBER, *supra* note 60, at 108.

88. *Id.* at 203.

89. McGowan, *supra* note 79, at 268-69.

whose code goes into the Linux kernel, ultimately one person can resolve it.[90]

Second, because anyone may reproduce and make derivative works of GPL'd code, such as the Linux kernel, there is no legal impediment to the proliferation of different versions of Linux (known as "forking" the code base), presenting a risk of the fragmentation that plagued Unix. On at least a couple of occasions, Torvalds faced a serious threat of a fork in Linux development.[91] At present, GNU/Linux development is increasingly being driven by firms that produce goods or services complementary to the operating system.[92] These firms would like a commoditized complement to their business model, and this rational competitive desire may constrain forking, but whether it will do so as the GNU/Linux OS gains market share remains to be seen.

III. THE *MICROSOFT* CASES

So much has been written about *United States v. Microsoft Corp.* that the case seems drearily familiar already. In this Part, I will try to demonstrate that there is still something useful to learn from the experience. I relate the D.C. Circuit's standard of liability for monopolization, and in particular its discussion of causation, to the error cost debate. I emphasize the record developed on remand, which served as the basis for the district court's decision to impose only modest remedies that were largely agreed upon between Microsoft and the Justice Department and to reject the more drastic remedies proposed by certain states. That record has not received as much attention as some of the evidence adduced during the liability phase of the trial. With regard to the error cost debate, however, the record in the remedy phase is as significant as the record developed in the liability phase of the proceeding.

I focus on four aspects of the case: Sun's Java technologies; Microsoft's "commingling" of browser code with operating system code; license restrictions on Original Equipment Manufacturer (OEM) modifications to the Windows user interface; and the role of open source licensing in both

---

90. For an example involving the incorporation of the TCP/IP stack into the kernel, see WEBER, *supra* note 60, at 104-05.

91. *Id.* at 117.

92. The reader is invited to contemplate the logos of firms supporting Open Source Development Labs, which begins to look like a NASCAR endeavor when its sponsorship is on parade. On open-source development as part of the business strategy of various firms, see Michelle Levesque & Greg Wilson, *Women in Software: Open Source, Cold Shoulder*, SOFTWARE DEV., Nov. 2004, *available at* http://www.sdmagazine.com/documents/s=9411/sdm0411b.html (subscription site).

the proposed breakup of Microsoft and, later, in the proposed open-source release of its browser code. In each aspect, the D.C. Circuit's analytical approach led it to hold lawful conduct that actually harmed competitors, and could conceivably have harmed competition (though I do not believe such harm was proven), while condemning conduct that caused little or no harm to competitors, much less to competition in general.

A. Sun Microsystems' Java Technologies

United States v. Microsoft Corp. began as a contempt proceeding based on Microsoft's integration of its web browser into its operating system. The theory of the government's case was that Netscape's Navigator browser might develop into a platform of its own, so Navigator was at least a potential substitute for Microsoft's Internet Explorer.[93] On this theory, Navigator represented a potential path of innovation in which the dominant firm moved up the computing hierarchy from the operating system to an application that interfaced with the Internet and other applications.

The browser theory had its problems. The browser was not a comprehensive platform. There was no realistic prospect that it would become a meaningful operating system substitute in the foreseeable future. When Netscape's CEO testified at trial that Netscape itself did not envision that the browser would become a substitute for Internet Explorer,[94] it was hard to see how Netscape would have eroded Microsoft's market power significantly, even if one accepted that Microsoft harmed Netscape by leveraging that power. Although whether that fact should matter to antitrust policy is a separate question, it is fair to say that the browser theory lost its appeal in rough proportion to the degree of information available about Netscape. In addition, it was not clear how far this theory advanced the idea of competition rather than the welfare of Netscape as a competitor. The theory did not imply that there would be perfect competition, or anything close to it, but that there would be a new platform consisting of technology in which a different firm—but still only one firm—held the rights.

As the case progressed, and the browser became a less plausible substitute for Internet Explorer, the government shifted its emphasis to Sun's

---

93. *Microsoft I*, 147 F.3d 935 (D.C. Cir. 1998) (affirming district court finding that Microsoft was not in contempt of the settlement decree; reversing district court order preventing Microsoft from demanding that OEMs license its web browser if they wanted to license its operating system).

94. *See* McGowan, *supra* note 27, at 787 n.199 (citing Testimony of Jim Barksdale, United States v. Microsoft Corp., 97 F. Supp. 2d 59 (D.D.C. 2000) (Nos. 98-1232, 98-1233), *available at* http://www.usdoj.gov/atr/cases/f1900/1999.htm).

Java technologies.[95] I will describe the technologies in more detail in a moment. For now, it is enough to say that Java technologies were designed to (eventually) allow application developers to write a program that would run on any operating system that also employed the technologies. If this vision became reality, Java would introduce competition into the operating systems market by lowering the applications barrier to entry into that market. An operating system vendor would only need to write a Java Virtual Machine (JVM) for its operating system, and any application written with the Java technologies would run on that system. Any number of operating systems could employ this strategy, so Java could lead to robust competition in the operating systems market.

Like the browser theory, the Java theory did not promise to eliminate all market power. On this theory, Java would replace Windows as the dominant computing technology, and Sun (rather than Netscape) would replace Microsoft as vendor of the *de facto* platform standard. There would be robust competition above and below the Java layer, but that layer would be owned by Sun. Sun would have supplanted Microsoft, but then Sun would be the monopolist. It is worth pausing to ask whether, as a matter of antitrust policy, such a goal is worth pursuing. Monopoly power at one level of a vertical stack of tight complements will affect the stack as a whole, and competition at each level. There was no reason to believe that competition in personal computer markets would approximate the model of perfect competition just because Netscape or Sun supplanted Microsoft as the platform monopolist. Why should the government spend millions of dollars, and risk setting precedent that might deter welfare-enhancing acts, just to move the monopoly position up one level in the computing hierarchy?

There are two possible answers to this question. First, barriers to entry might be lower farther up the stack, producing a reduction in market power and a corresponding consumer benefit. If true, that is at least a cogent theory. Second, it might not matter whether there is a platform monopolist in the computing hierarchy, much less who it is or at what level it supplies products. Even if competition in this market followed a Schumpeterian model, the antitrust law might still ensure that the fight is fair. On

95. The browser subsequently became significant more in terms of its role a distribution channel for Java than for any potential that it would become a platform substitute for Windows. *See* David McGowan, *Has Java Changed Anything? The Sound and Fury of Innovation Litigation*, 87 MINN. L. REV. 2039, 2041 (2003). The discussion in this portion of this Article draws heavily upon my earlier analysis, though it also revises that analysis to account for weaknesses in the appellate court's opinion that I should have given more weight the first time around.

this view, by acting as a referee to ensure a fair fight antitrust might increase confidence that the best fighter won. Appealing as it is, however, the referee metaphor promises more than reality delivers. Two aspects of the Java portion of the proceeding deserve close attention: Microsoft's modification of the Java technologies and Microsoft's alleged confusion of software developers.

1. Microsoft and Java Fragmentation

The most interesting aspect of the Java allegations was the Justice Department's claim that Microsoft maintained its monopoly by attempting (with some success) to fragment Java as a technology standard. The evidence calls to mind the fragmentation of Unix, which undermined its potential to become a coherent operating system standard. As noted above, Sun's alliance with AT&T encouraged the fragmentation of Unix. In the *Microsoft* case, the shoe was on the other foot.

a) District Court Findings

Much of the evidence regarding the fragmentation theory had to do with Microsoft's development of Java technologies. The relevant technologies include the Java programming language and the Java class libraries, which contain application programming interfaces (APIs). Developers could use these technologies to write Java programs, and these programs would run on a Java compiler, which would translate them into Java "bytecode." The bytecode in turn would run on a JVM, which are software programs written for specific operating systems. JVMs translate Java bytecode into instructions the particular operating system can execute.[96]

The gist of the government's charge was that Microsoft maintained its monopoly power by developing technologies that allowed Java to function well as a set of development tools for Windows but impeded the degree to which Java could serve as new platform. I will briefly summarize the principal allegations.

At the time relevant to the litigation, the Java class libraries did not provide enough interfaces to allow developers to write sophisticated programs without invoking some code from an underlying operating system.[97] The manner in which developers invoked this code was important to the path of Java's development. The district court found that Sun recognized this problem and "sponsored a process for the creation of a software method that would allow developers writing in Java to rely directly upon" code specific to particular operating systems "in a way that would never-

96. *Id.*

97. *Microsoft II*, 84 F. Supp. 2d 9, 105 (D.D.C. 1999).

theless allow them to port their applications with relative ease to . . . different operating systems."[98]

Notwithstanding the Sun-sponsored process, Microsoft produced its own set of Java development tools and complementary products such as a Microsoft JVM.[99] The net result of these efforts was a distinctive "Java runtime environment" for Windows that was incompatible with the environment Sun sponsored. In part, Microsoft developed its own methods for software developers to invoke "native" operating system code (meaning code specific to a particular operating system running underneath the JVM).

Microsoft's methods were incompatible with Sun's, and they created a trade-off. The district court found that programs written with Microsoft's methods ran slightly faster on Windows than did programs written with Sun's methods. On the other hand, a program written with Microsoft's methods was harder to port to other operating systems than programs written with Sun's.[100] The district court found it would have been inexpensive for Microsoft to implement Sun's methods along with its own, but that Microsoft refused to do so until ordered to do so by a court.[101]

In addition, Microsoft created two "keywords" and "compiler directives" that added to the functions Java could perform but which also worked only with Microsoft's JVM. Microsoft shipped its Java developer tools with these keywords and compiler directives enabled by default; if a developer did not wish to use them, she would have to turn them off, using a menu option available for that purpose.[102] If a developer used these tools, the developer's program would run only on Windows and would be incompatible with other JVMs.

Microsoft also developed a high-performance JVM for Windows, which ran programs faster than other JVMs. Because Java technologies comprised a translation system, in which Java bytecode was translated into code an underlying operating system could execute, Java programs necessarily ran somewhat slower than programs written directly for an operating system. The faster a JVM ran, the less of a difference there would be between a Java program written for the JVM and one written for the underly-

98. *Id.*

99. It did so under a license from Sun, which Sun later alleged Microsoft had breached.

100. *Id.*

101. *Id.* at 106.

102. *Id.* at 106-07.

ing operating system, so this faster JVM tempted developers to write Java for Windows rather than cross-platform Java.[103]

Microsoft's strategy of developing Windows-specific Java technologies also formed the basis of the charge that Microsoft impeded Java's progress by fooling developers into writing Windows-specific programs when they in fact wished to write cross-platform programs. The district court found that Microsoft failed to warn developers that its Windows-specific Java extensions were enabled by default, and that the default mode of Microsoft's Java development tools would therefore produce Windows-specific programs rather than cross-platform programs.[104] The court concluded that Microsoft intended to deceive developers,[105] but it did not find that any developers actually had been deceived.[106]

b) The Court of Appeals' Fragmentation Analysis

The D.C. Circuit treated Microsoft's development of its JVM and Java developer tools as product development decisions. With respect to such decisions generally, the court said that courts should view skeptically "claims that competition has been harmed by a dominant firm's product design changes."[107] The court reasoned that firms in competitive markets innovate to attract or increase demand, and innovation sometimes produces incompatible products. It worried that imposing liability for the design of incompatible products would deter innovation. On the other hand, it also said that "judicial deference to product innovation . . . does not mean that a monopolist's product design decisions are per se lawful."[108] As a formal matter, the court opted for balancing, holding that the proper test is that, "in order to violate the antitrust laws, the incompatible product must have an anticompetitive effect that outweighs any procompetitive justification for the design."[109]

At first glance, the court's balancing rhetoric seems to fit nicely with the integration approach. A closer look invites doubts, however. To say that innovation generates demand is to say only that either false positive or false negative findings would be costly. The integration approach seems to require straightforward cost-benefit analysis of all relevant effects, includ-

---

103. *Id.* at 108.

104. *Id.* at 106-07.

105. *Microsoft III*, 87 F. Supp. 2d 30, 43 (D.D.C. 2000), *aff'd in part and rev'd in part*, 253 F.3d 34 (D.C. Cir.) (en banc), *cert. denied*, 534 U.S. 952 (2001).

106. McGowan, *supra* note 95, at 2045.

107. *Microsoft IV*, 253 F.3d 34, 64 (D.C. Cir. 2001) (en banc).

108. *Id.* at 65.

109. *Id.* at 75.

ing the effects of product design decisions, and the court's balancing rhetoric is consistent with this view. However, the court's skepticism presumably is rooted in the sort of experience on which the market correction approach stakes its claims. In fact, the court's analysis was so deferential that what it did came closer to a rule of per se legality for design decisions than to true cost-benefit balancing.

The court of appeals first held that Microsoft could not be found liable for developing its own JVM, even though that JVM was incompatible with Sun's Java implementation. The court reasoned that Microsoft's JVM ran programs written for Windows faster than did other JVMs and "[did not] itself have any anticompetitive effect."[110] This ruling shows how willing the court was to break down Microsoft's conduct into individual parts and then analyze each part separately. The court acknowledged that Microsoft's development of its JVM was related to its development of Java development tools, and to its alleged deceit regarding those tools, but the court considered the JVM development effort on its own and acquitted Microsoft of liability for that effort.

The court's fine-grained analysis is what one would expect from the integration approach, but its willingness to accept that faster is better without weighing the degree to which Microsoft's JVM may have fragmented a nascent standard is not consistent with that approach. It is market-correction-style analysis, plain and simple. The court's analysis combined very detailed review of individual aspects of conduct with relatively categorical standards of review, producing an odd result that endorsed balancing in the abstract but avoided it in the particular.

The court's treatment of the development tools issue strengthens the sense of tension in the court's analysis. The court held that Microsoft could not be found liable for producing Java development tools that were incompatible with Sun's Java implementation. Its explanation for this acquittal was breathtakingly brief. Here it is, in full: "The District Court found that, not only were these tools incompatible with Sun's cross-platform aspirations for Java—no violation, to be sure—but Microsoft deceived Java developers . . . ."[111]

It is hard to know what to make of this language. It no doubt follows from the court's general skepticism of allegations regarding design incompatibilities. The court said that skepticism of such allegations did not imply a per se rule that design decisions are lawful, however, while the "to be sure" aside at least leaves room for such a rule. The court made no ef-

110. *Id.* at 70.
111. *Id.* at 76.

fort to weigh the effects of Microsoft's development of such tools. The court also treated the development of tools in isolation, as it did with development of the JVM, even though it recognized that the two developments were closely related.[112]

The court's use of broadly skeptical assumptions in its examination of particular conduct obscured the strength of the Justice Department's charges and the district court's conclusions. It was perfectly logical for the Justice Department to allege that Microsoft acted in an anticompetitive manner by trying to fork Java. The history of Unix shows that fragmentation may stall the adoption and development of a platform. It is hard to believe that the lessons of that history were lost on either Microsoft or Sun, and therefore it is perfectly plausible to conclude that the Justice Department and the district court's more contextual approach captured an aspect of Microsoft's conduct—call it synergistic anticompetitiveness—that the appellate court's approach missed.

Because both logic and experience suggest that forking a nascent standard may undermine adoption and development of the standard, it is plausible to assert that the anticompetitive effects of Microsoft's Java development efforts exceeded the benefits of that work. I happen to agree with the appellate court's conclusions on these issues, and in fact would have gone farther than it did in exculpating Microsoft, but one has to admit that the court's analytical approach did not adequately refute the arguments it rejected.[113]

The court's acquittal of Microsoft's work on Java developer tools was qualified somewhat by its affirmation of the district court's conclusion that Microsoft engaged in monopolization by deceiving software developers. Unfortunately, that affirmation was probably the low point of the opinion. To understand why, one must bear in mind that, earlier in its opinion, the court had cautioned against using a party's intention as a basis of liability.[114] The court there warned that intention was relevant to monopolization only insofar as it "helps us understand the likely effect of the monopolist's conduct."[115]

The court's treatment of the developer deceit claims was inconsistent with these principles. The court of appeals treated the district court as having found that Microsoft actually had deceived developers, but the district

112. *Id.* at 75.

113. For a similar analysis of the court of appeals' treatment of these allegations, see Einer Elhauge, *Defining Better Monopolization Standards*, 56 STAN. L. REV. 253, 318-19 (2003).

114. *Microsoft IV*, 253 F.3d at 59.

115. *Id.*

court had found no such thing. The lower court found only that Microsoft hoped and intended to deceive developers, which under the D.C. Circuit's own standard is quite different. Other than one vague, hearsay reference from a combative Sun witness,[116] there was no record evidence that any developer had ever been deceived into writing Windows-specific programs when she intended to write cross-platform programs. If it existed, evidence of actual developer confusion would have been so easy to obtain that the government's failure to show even a single instance of such confusion supports a strong inference that there was none. Developers write code for a living, after all, and it seems highly unlikely that they would be unable to work through a menu option and disable Microsoft's Windows-specific tools if they wanted to write cross-platform programs.

The court of appeals said Microsoft had offered no justification for its "campaign to deceive developers,"[117] but that characterization is too slanted for a case based on nondisclosure rather than misrepresentation. It also makes no sense. The court itself said the tools helped developers write programs that would run faster (through the Microsoft JVM) than programs written with Sun's technology. If this aspect of Microsoft's conduct was procompetitive, as the court held it was, then why would it be anticompetitive for Microsoft to make these competitive products the default option in its Java development distribution? What purpose would be served by making developers incur even the trivial marginal cost of enabling these supposedly competitive tools? Why would the antitrust laws force Microsoft to make an efficient option costlier to implement by requiring that developers opt into it rather than requiring that developers who do not wish to use the technologies opt out? If the answer is that it would minimize fragmentation, why was development of incompatible tools (which one could plausibly if not inarguably conclude were designed to fragment Java) lawful in the first place? Perhaps there are answers to these questions, but I doubt that they are very good. Most likely they would take the form of assertions about a balance of costs and benefits for which there was little if any data, imprecisely weighed.

The court's application of its liability standard is both more abstract and categorical than the language of the standard suggests. Acts the court sees as part of a set of good activities, such as product design, are lawful, without the need to balance actual costs and benefits. Acts the court sees as part of a set of bad activities, such as deceit, are unlawful regardless whether they actually worked. The result is analysis that is at once oddly

116. *See* McGowan, *supra* note 95, at 2045-46 (citing *Microsoft IV*, 253 F.3d at 77).
117. *Microsoft IV*, 253 F.3d at 77.

categorical and nuanced to the point of hair-splitting. The court's analysis is detailed enough to distinguish the JVM from developer tools, and the tools from knowledge about them, and categorical enough to treat the first two subjects very differently from the third, while being with respect to all three charges relatively indifferent to the actual balancing of real-world costs and benefits.

Under the liability standards it articulated, the court should have rejected the developer deceit allegations on the ground that the deceit caused no harm. Under those same standards, however, the court should have taken more seriously the charge that Microsoft's Java development work was unlawful because it fragmented a possible substitute standard. At the end of the day, I agree with the court's rejection of those claims, and in fact think the court should have rejected all the claims related to Microsoft's development of a Windows-oriented Java runtime environment. However, the court's hybrid approach made a mess of the allegations, condemning the trivial acts and approving the significant ones, and doing so by using categorical analysis that is at odds with the balancing approach it was supposed to have employed.

c) The Causation Argument

The Justice Department's allegations concerning Java provided the occasion for the court of appeals' discussion of the question of causation. That discussion relates directly to the question of error costs, and to the larger question whether antitrust should be enforced through the SWAT team model or the traffic cop model. A bit of background is necessary to place the discussion in context.

Adopting a remedy proposed by the Justice Department, the district court ordered that Microsoft be broken up into two companies. One would produce operating systems, such as Windows. The other would produce applications, such as Microsoft Office.[118] The court of appeals reversed this order, noting in part that the order rested on a number of liability findings it had rejected.

The court of appeals also chose the remedy question as the appropriate way to deal with one of Microsoft's main arguments regarding liability for its Java-related conduct. Microsoft insisted that it should not be held liable for monopolization with regard to that conduct because the government failed to prove that, but for its allegedly anticompetitive conduct, Microsoft would have lost any market power. In other words, the government

118. United States v. Microsoft, 97 F. Supp. 2d 59, 64 (D.D.C. 2000) (final judgment).

had not shown that Microsoft's Java-related conduct caused it to maintain its market power beyond the point at which such power would have eroded had it not engaged in that conduct.[119]

The court of appeals rejected this argument as it applied to liability. It found no precedent for the proposition that, in an equitable enforcement action, the government had to "present direct proof that a defendant's continued monopoly power is precisely attributable to its anticompetitive conduct."[120] The court said that to require plaintiffs to present such proof would only "encourage monopolists to take more and earlier anticompetitive action."[121] The court was therefore willing to infer causation "when exclusionary conduct is aimed at producers of nascent competitive technologies" in part on the ground that this inference made the defendant "suffer the uncertain consequences of its own undesirable conduct."[122]

Two aspects of the court's reasoning are significant. First, if one believes the monopolization offense is concerned with anticompetitive effects, rather than the regulation of conduct as such, this reasoning is circular. It infers anticompetitive effects from the badness of an act, rather than condemning an act as bad because of its effects. Second, this inference forms the front end of an error-cost argument that, though it does not proclaim itself as such, is a perfect example of the integration approach. Though the court was not concerned with causation at the liability phase, it was concerned about causation when it came to choosing a remedy. The court said that, "absent some measure of confidence that there has been an actual loss to competition that needs to be restored, wisdom counsels

---

119. *Microsoft IV*, 253 F.3d at 78-80.

120. *Id.* at 79. It is not clear why the court thought an equitable enforcement action should be treated differently from a private action. Earlier in the opinion, the court said that "[i]n a case brought by a private plaintiff, the plaintiff must show that its injury is of 'the type that the statute was intended to forestall'" and that "no less in a case brought by the Government, it must demonstrate that the monopolist's conduct harmed competition, not just a competitor." *Id.* at 59. Section 28 of 15 U.S.C., which the court did not cite, might support the idea that the United States could enjoin conduct that harmed no one simply because it was unlawful, but that interpretation of the statute both assumes that the antitrust laws condemn conduct that does not harm competition and ignores the court's statement that the government must prove harm to competition.

121. *Id.* at 79.

122. *Id.* ("[T]he question in this case is not whether Java or Navigator would actually have developed into viable platform substitutes, but (1) whether as a general matter the exclusion of nascent threats is the type of conduct that is reasonably capable of contributing significantly to a defendant's continued monopoly power and (2) whether Java and Navigator reasonably constituted nascent threats at the time Microsoft engaged in the anticompetitive conduct at issue.").

against adopting radical structural relief" in part because the long-term effects of divestiture are uncertain.[123]

From the perspective of the integration approach to error costs, this two-tiered approach to causation is perfectly logical. If there is some reason to believe an act is not good, but little reason to believe it has caused harm, then the act should be enjoined rather than made the basis for significant structural reform. A sliding scale connecting confidence in causation with the severity of remedy connects the probability of error to the magnitude of error, thus lowering the expected cost of error. It is just what the doctor ordered. Just as the court's two-tiered approach to causation may be read as a device to limit error costs, one's view of the relationship between causation and error costs expresses a position on the larger question whether the monopolization offense should be viewed as SWAT team antitrust or traffic cop antitrust. The integration approach to error costs, and the sliding scale relating confidence in liability to severity of remedy that implements it, leans strongly toward a traffic-cop view of antitrust.

In fact, unlike the market correction approach, which insists on fairly strict limits but has only anecdotal, backward-looking justifications for them, the integration approach admits of no limits on antitrust review at all. The integration approach is an extension of a logical, decision-theoretic process, and there are no logical limits to that process. Practical limitations might arise, such as the cost of acquiring the marginal unit of information needed to take analysis one step further, but the approach itself has no boundary. On this view, it makes perfect sense to speak of antitrust violations that do not harm competition (the acts in question may produce net losses) or of monopolies maintained through acts that do not prolong monopoly.

In contrast, advocates of a market correction approach to error costs, of which I am one, would reject the D.C. Circuit's command that "in devising an appropriate remedy, [on remand] the District Court also should consider whether plaintiffs have established a sufficient causal connection between Microsoft's anticompetitive conduct and its dominant position in the OS market."[124] We would not say, as the court of appeals did, that "if the court on remand is unconvinced of the causal connection between Microsoft's exclusionary conduct and the company's position in the OS market, it may well conclude that divestiture is not an appropriate remedy."[125] Rather, the court should dismiss the case altogether, because if there is no

123. *Id.* at 80.
124. *Id.* at 106.
125. *Id.* at 107.

connection between Microsoft's conduct and its market position, then there was no violation in the first place.[126]

The court of appeals' reluctance to employ strict causation standards at the liability phase of the proceedings sheds some light on its odd use of seemingly categorical propositions to engage in a particularized review of discrete acts. The court's analysis is too detailed for the market correction approach, which would advise against judges intruding so far into business decisions and inferring liability from ambiguous or poorly explained acts. The analysis is also too general for a true integrationist approach, which would substitute true cost-benefit balancing for the court's skeptical, categorical balancing.

It was clear after the court of appeals' decision that Microsoft would not be broken up. The court of appeals did not preclude such a remedy, but its position was apparent. What remained for the district court on remand was to choose a remedy "tailored to fit the wrong creating the occasion for the remedy."[127] Like the integration approach, this command had the compelling rational appeal of the idea of equilibrium. It was aesthetically pleasing and logically unimpeachable. It had no truck with off-the-rack presumptions. It was a complete pipe dream.

d) The Remand Proceedings and the (Lack of a) Remedy

Insofar as the JVM and Java development tools arguments were concerned, the court of appeals handed the district court a complete mess. Microsoft's development of the JVM and its Java developer tools had been lawful, and therefore had to be disregarded in tailoring a remedy. Microsoft's alleged deception of developers was unlawful, and therefore could be the basis for a remedy, even though no developers were deceived.

On remand, the evidence showed that the worst thing Microsoft had done to Java was to fork the technology by producing a Microsoft-specific Java runtime environment. Carl Ledbetter, a witness for the nine states that did not accept the consent decree agreed to by Microsoft and the Justice Department, testified that Microsoft's development of a Java runtime environment that deviated from Sun's authorized implementation was "crippling" to Java's position in the market, and a substantial cause of

126. *See* Alan J. Meese, *Monopolization, Exclusion, and the Theory of the Firm*, 89 MINN. L. REV. 743, 770 (2004) (referring to D.C. Circuit's imposition of liability without requiring a finding of causation as "entirely circular").

127. *Microsoft IV*, 253 F.3d at 107.

Java's failure to become a viable substitute for Windows.[128] The court of appeals had, of course, held that this conduct was legal, so there was no basis to try to remedy it.

That left the alleged deception of developers. There was no evidence that anyone had been deceived, but what if there had been? Would it have mattered? Professor Carl Shapiro was the principal causation witness for the non-settling states.[129] He testified that he knew of no evidence that any developers actually had been confused, and that he did not regard the alleged deception as "having the most significant effects."[130] (Because there was no evidence of confusion this testimony actually overstates matters.) Not surprisingly, the district court found that developer deception amounted to a "single, very specific incident of anticompetitive conduct," which Microsoft had ceased, and which posed no threat of continuing harm from developer deception.[131] The district court therefore ordered no remedy relating to this conduct, a ruling the D.C. Circuit affirmed.[132] For those keeping score, the final result on the technology-related Java allegations is: what was harmful (at least to competitors, though not, in my view, to competition) was not illegal, and what was illegal was not harmful, so no remedy was required.

No matter how you look at it, litigation of the Java allegations produced a mess. It would have been better for the court to follow the market correction approach to its logical conclusion, and reverse liability with regard to all these charges. Alternatively, the court could have followed the integration approach to its logical conclusion and balanced the costs and benefits of the development work through which Microsoft fragmented Java, without viewing that work categorically and in isolation as almost *per se* lawful. I believe the latter approach would lead to an intolerable amount of judicial involvement in product design, in pursuit of chimerical litigation equilibriums, so that the former approach is by far the better one. Because the court's stated approach essentially tracks the integration approach to error costs, the Java mess stands as anecdotal evidence in support of the market correction approach. Even if one is inclined to criticize

---

128. Transcript of Trial Record at 1700-01, *Microsoft V*, 224 F. Supp. 2d 76 (D.D.C. 2002) (No. 98-1233) [hereinafter Trial Transcript] (testimony of Carl Ledbetter on March 27, 2002).

129. *Microsoft V*, 224 F. Supp. 2d at 149.

130. Trial Transcript, *supra* note 128, at 3439 (testimony of Carl Shapiro on April 11, 2002). Professor Shapiro also testified that his view of the effect of Microsoft's fragmentation of Java was consistent with the view expressed in Mr. Ledbetter's testimony.

131. *Microsoft V*, 224 F. Supp. 2d at 265.

132. *Microsoft VII*, 373 F.3d 1199, 1213 (D.C. Cir. 2004).

the court as having been too intrusive rather than too lenient, however, one has to concede that the disposition of the Java allegations in the *Microsoft* litigation stand as an example of how hard it is for even a very able court to follow the integration logic where it leads.

2. Exclusive Agreements

Both the district court and the court of appeals found Microsoft liable under section 2 of the Sherman Act for entering into certain agreements with Independent Software Vendors ("ISVs"). Under the relevant agreements, Microsoft would provide technical benefits to ISVs only if they agreed to use Microsoft's JVM as their default JVM.[133] The district court thought developers had to use Microsoft's Java development tools to ensure compatibility with its JVM, implying that developers would write Windows-only programs.[134] If the developers wrote Windows-specific programs, they would have no incentive to distribute any JVM other than Microsoft's, meaning any JVM that complied strictly with Sun's standards rather than adding to them (as with Microsoft's additional keywords). Thus, Microsoft's agreements might raise the cost to Sun of distributing its authorized JVMs, possibly impeding the advance of cross-platform Java.[135]

The court of appeals affirmed the district court's liability findings with respect to these agreements. The court agreed that effective foreclosure of an efficient distribution channel was an anticompetitive act, and it rejected Microsoft's justification for that act, which the court described as being only that the "ISV agreements reflect an attempt 'to persuade ISVs to utilize Internet-related system services in Windows rather than Navigator.'"[136] The court thought this explanation meant nothing more than that keeping developers focused on Windows would preserve Microsoft's monopoly, which it characterized as a "competitively neutral" goal.[137] The court therefore found that Microsoft had not justified its conduct, so once again it did not reach the balancing portion of its liability analysis.

It is not clear what the court of appeals meant by saying that Microsoft's efforts to promote distribution of its Java technologies were com-

133. *Microsoft II*, 84 F. Supp. 2d 9, 108 (D.D.C. 1999).

134. *Id.* The record showed that Microsoft's JVM ran programs written in pure Java (not employing the Microsoft extensions) very well, so it is not clear why the district court thought this. For the evidence, see McGowan, *supra* note 95, at 2048-49.

135. I say "possibly" because during the periods relevant to the litigation (and to date), Java never actually lived up to the promise that a developer could write a program once and have it run on any platform. *See id.*

136. *Microsoft IV*, 253 F.3d 34, 72 (D.C. Cir. 2001) (en banc).

137. *Id.*

petitively neutral. The court found that Microsoft's *development* of those technologies was a form of competition, so what was wrong with actively promoting their use by developers and their distribution? Why would such acts not be a form of competition, too? The court seems to have avoided balancing the harms of Microsoft's agreements against the benefits of such distribution, which the last stage of its liability test would require. Instead, the court simply restated Microsoft's argument in a less sympathetic way while ignoring its own holdings regarding development of the technologies distributed under the agreements.

The entire set of Java-related allegations depicted a standards battle in which Microsoft and Sun competed to sponsor the dominant version of the Java technologies. It would make sense to say that the battle was anticompetitive because Java belonged to Sun, which had the exclusive right to define the parameters of the standard, so that Microsoft's fragmentation was unlawful.[138] If the court was unwilling to go so far, however, its rejection of Microsoft's argument seems another example of judicial hair-splitting that deems the truly significant conduct lawful while tinkering at the margins of distribution costs that meant little in light of the fragmentation of the underlying technology.

B. Browser Integration

The district court found Microsoft liable for preventing OEMs from serving as a viable distribution channel for competing browsers. It found that Microsoft did this in part by integrating its web browser into its operating system.[139] By integrating browser code with operating system code, Microsoft allegedly made it harder for OEMS to be sure they would not impair Windows' performance if they installed a competing browser before shipping a computer. In addition, OEMs incur most of the cost of supporting users who need help with their machines. A novice user might be confused if more than one web browser icon appeared on a desktop, leading to higher support costs; OEMs could avoid these costs by hiding those icons or otherwise concealing access to Microsoft programs.[140] However, because OEM margins were tight, the court believed they were

138. Such a ruling would have raised some questions—why, for example, should Sun have the exclusive right to control Java when Microsoft was found liable for exerting too much control over the desktop—but there were plausible answers for those questions.

139. *Microsoft III*, 87 F. Supp. 2d 30, 39-40 (D.D.C. 2000), *aff'd in part and rev'd in part*, 254 F.3d (D.C. Cir.) (en banc), *cert. denied*, 534 U.S. 952 (2001).

140. *Microsoft II*, 84 F. Supp. 2d 9, 49-50 (D.D.C. 1999).

unlikely to incur such additional costs, making Microsoft's commingling of code prohibitive of OEM installation of competing browsers.[141]

The court of appeals analyzed the general integration claim as three individual acts. First, Windows would sometimes launch Microsoft's browser automatically even if a user had chosen another default browser. The examples of such cases introduced at trial involved users accessing Microsoft's own computers, as they might do in obtaining help or updates for their operating system. The court found that Microsoft had demonstrated valid technical reasons for launching its own browser in such cases, and that Microsoft therefore could not be held liable for this design choice.[142]

The court affirmed liability for the remaining two acts, however. Microsoft made it impossible for users to employ Windows' "add/remove" utility to remove the browser from the operating system,[143] and it commingled browser code with files that contained code that performed other functions.[144] As to these claims, the court found that Microsoft was liable for monopolization because the government demonstrated how those decisions might shore up Microsoft's market power and because Microsoft had not justified the decisions.[145] Microsoft had made "some general claims regarding the benefits of integrating the browser and the operating system" but it had not specified or substantiated those claims.[146] One wonders how much substantiation the court would require, given that this allegation concerned a design decision, and why Microsoft's testimony that its integration of code was consistent with its overall design philosophy, which one would be hard-pressed to say had no competitive benefits, was not enough.

On remand, Microsoft and the government agreed to allow OEMs and end-users to employ means such as the add/remove utility to remove Microsoft's browser from Windows.[147] Proceedings regarding commingling were more controversial. The district court rejected demands from various sources that Microsoft be ordered to separate the unlawfully integrated

141. *Id.* (Findings of Fact 159); *id.* at 60-61 (Findings of Fact 210); *id.* at 63 (Findings of Fact 217).

142. *Id.* at 67.

143. *Microsoft IV*, 253 F.3d 34, 66-67 (D.C. Cir. 2001) (en banc).

144. *Id.*

145. *Id.*

146. *Id.*

147. *Microsoft VI*, 231 F. Supp. 2d 144, 176 (D.D.C. 2002).

code.[148] Instead, it approved the remedy to which Microsoft and the government had agreed: allowing OEMs to remove user access to Microsoft programs that might become OS substitutes (middleware) and to feature competing middleware instead. Under this remedy, the code could remain mingled, so long as the user did not have to know about it.[149]

The district court offered several reasons for not ordering Microsoft to separate the integrated code. The court noted that the government had never sought a remedy requiring the actual separation of code, but had instead focused its attention on the perceptions of end users.[150] It found that the non-settling states had not offered a coherent definition of operating system code, and thus no reliable way to separate the code from other code, a distinction necessary to enforce any remedy requiring the separation of code.[151] The non-settling states also failed to present any economic justification for a separation remedy; Professor Shapiro declined to opine on that aspect of the states' request.[152]

The district court's reasoning went beyond such deficiencies, however. It credited testimony that an order requiring Microsoft to remove code from Windows would harm the industry and consumers.[153] The court's analysis is interesting because it sheds light on the reasons for finding liability for bundling in the first place, as well as on some other aspects of the government's case. Regarding harm to the industry, the court reasoned that ordering Microsoft to separate integrated code would impede Microsoft's development of Windows and pose a risk that the Windows platform would fragment.[154] In addition, if Microsoft had to remove middleware code from Windows, the platform code base would vary depending on whether the code was replaced and, if so, what replaced the code.[155] The court also credited testimony that separation of integrated code could

148. *Id.* at 28-30. The court specifically rejected a proposed remedy submitted by the non-settling states, which would have required that Microsoft unbundle versions of any program that bundled middleware code with operating system code. *Microsoft V*, 224 F. Supp. 2d 76, 157-59 (D.D.C. 2002).

149. *Microsoft V*, 224 F. Supp. 2d at 157-59; *id.* at 268 (App. B, Final Judgment ¶ III.C); *id.* at 270-70 (App. B, Final Judgment ¶ III.H); *see* Second Revised Proposed Final Judgment ¶¶ III.C, III.H, *available at* http://www.usdoj.gov/atr/cases/f10100/10146a.htm.

150. *Microsoft VI*, 231 F. Supp. 2d at 29; *Microsoft V*, 224 F. Supp. 2d at 158.

151. *Microsoft V*, 224 F. Supp. 2d at 157, 247-48.

152. *Id.*

153. *Id.* at 158.

154. *Id.*

155. *Id.* at 252-53.

produce over 1,000 variations of Windows.[156] Not surprisingly, the court believed that fragmentation would increase substantially the cost to developers of writing Windows-compatible programs.[157]

The court of appeals affirmed the district court's remedy. It too rejected demands that Microsoft be ordered to separate the relevant code, and its reasons also make one wonder why the court held that the integration of code was a basis for liability in the first place. The court of appeals agreed with the district court that the anticompetitive effects of the integration allegations centered on user perceptions, not on code aggregation.[158] It said the district court's remedy "went to the heart of the problem Microsoft had created, and it did so without intruding itself into the design and engineering of the Windows operating system. We say, Well done!"[159]

The D.C. Circuit was so complimentary in part because it took seriously the risk that a more intrusive order would fragment the Windows standard. Massachusetts, the lone state to press an appeal from the district court's rejection of its proposed remedies, claimed that the fragmentation worry was simply a different version of Microsoft's claim (in defense of effectively exclusive agreements with Internet access providers) that it needed to keep developers focused on Windows APIs rather than the APIs of some other system.[160]

This argument was not quite right—there is a difference between condemning acts that favor one's own platform to the exclusion of others and refusing to order acts that would fragment an existing platform—but there is something to this analogy. If fragmentation is bad, however, isn't standardization good? If so, efforts to increase the strength of a standard should be good, too, if for no other reason than that they reduce the risk of (bad) fragmentation. And if there is a benefit to have a firm "sponsor" the

156. *Id.*

157. *Id.* at 254. Some commentators are skeptical that fragmentation of the OS code base would be a significant problem even if several firms offered versions of Windows, rather than simply having Microsoft offer multiple versions. *See* Robert J. Levinson et al., *The Flawed Fragmentation Critique of Structural Remedies in the* Microsoft *Case*, 46 ANTITRUST BULL. 135, 138 (2001). Some of the reasons for this view would support a skeptical view of any fragmentation, though they do not demonstrate that fragmentation is a trivial risk. One implication of this argument is that developers would absorb some of the (possibly modest) cost of making sure an application ran on all OS variants, which may explain evidence of developer opposition to the proposed un-mingling remedy in *Microsoft*. Levinson et al., *supra*, at 153.

158. *See Microsoft VII*, 373 F.3d 1199, 1207-08 (D.C. Cir. 2004).

159. *Id.* at 1210.

160. *See id.* at 1211.

adoption and maintenance of a standard,[161] then why should the sponsor be held liable for policing and strengthening the standard?

Massachusetts might answer that one may not keep developers focused on one's platform by effectively foreclosing the most efficient distribution channel, thus raising the distribution costs of potential rival platforms. It is not clear that this answer should be conclusive, however. If acts that strengthen a standard also increase rivals' distribution costs then, so long as strengthening the standard is good, the acts present a tradeoff between effects that, under the court of appeals' liability rule, should be weighed in the balance. Raising the distribution costs of rival platforms would reduce the supply of such platforms, but if fragmentation is bad then increased distribution of a rival platform is not obviously a net gain to competition. That is probably why the D.C. Circuit rejected the view that "fragmentation [is] merely competition by another name."[162]

Nevertheless, in affirming liability with regard to Microsoft's API agreements, the court of appeals said that keeping developers focused on Microsoft's Windows API's was a competitively *neutral* argument, not a procompetitive justification for the agreements.[163] This conclusion allowed the court to affirm liability without balancing the costs and benefits of the agreements, but the arguments presented above suggest that the court of appeals should have engaged in such balancing. Those arguments suggest the court judged Microsoft's defense too harshly, a point that in turn calls into question the court's decision that Microsoft could be held liable for its platform design decisions (as well as for the agreements for which the defense was presented).

Before leaving this topic, it is worth noting that one argument fell through the cracks as litigation dragged on. The district court's remedy of hiding access to Microsoft programs addressed only one aspect of the original liability findings—that novice users with access to multiple icons might generate higher OEM support costs. The "hide the icon" remedy did not address the finding that integrating code increased the testing costs of

161. To use a phrase Professors Katz and Shapiro employ in the article Professor Williamson cited as showing an economic model beyond the power of a court to apply. Katz & Shapiro, *supra* note 32. Alan Meese suggests that keeping developers focused on Windows APIs might be a way for Microsoft to recoup costs it incurs in helping developers write for Windows. *Cf.* Meese, *supra* note 126, at 824-25 (making the suggestion in the context of exclusive dealing relationships). This idea builds on Howard Marvel's argument that exclusive dealing arrangements mimic property rights and allow one firm to recoup its investment by cooperating with another. Howard P. Marvel, *Exclusive Dealing*, 25 J. LAW & ECON. 1, 6 (1982).

162. *Microsoft VII*, 373 F.3d at 1212.

163. *Microsoft IV*, 253 F.3d 34, 71 (D.C. Cir. 2001) (en banc).

OEMs that wished to install multiple programs.[164] It would be pointless to try to assess the marginal loss (if any) from failing to address this aspect of costs. More to the point, if the district court was right about the risk of fragmentation, then the costs of an unbundling remedy would almost certainly exceed the marginal OEM testing cost. Again, this result calls into question why an increase in OEM testing costs should have created liability in the first place.

The final verdict on commingling resembles the final verdict on Java. That which was unlawful did not have to be undone, because undoing it would likely create greater costs than benefits. Those costs, however, called into question whether liability should have been found in the first place. It is very difficult to consider this result as anecdotal evidence in favor of the integration approach to error costs. It is very easy to consider it as anecdotal evidence in favor of the market correction approach.

C. OEM License Restrictions

The district court found Microsoft liable for monopolization in part because Microsoft restricted the ability of OEMs to modify Windows in ways that promoted competitive products.[165] The court of appeals grouped these restrictions into three categories; I reduce them to two here.

In the court of appeals' first and third categories were Microsoft restrictions that prohibited OEMs from removing access (such as by desktop icons) to Microsoft's web browser and from altering Microsoft's "Active Desktop." The district court found that these restrictions were anticompetitive because they raised the cost to OEMs of pre-installing additional browsers, and thereby deterred OEMs from doing so, thus protecting Microsoft's operating system monopoly. The court of appeals agreed.[166]

Microsoft argued that these restrictions were simply valid assertions of the copyrights it held in its code. As applied to these restrictions, the D.C. Circuit said the copyright argument "borders upon the frivolous."[167] Microsoft also argued that its restrictions maintained a uniform appearance so

---

164. *Microsoft II*, 84 F. Supp. 2d 49-50 (D.D.C. 1999) (Findings of Fact 159). The court of appeals mentioned the OEM cost issue but did not distinguish between user support costs and testing costs. *Microsoft VII*, 373 F.3d at 1209.

165. Specifically, Microsoft forbid OEMs from (1) removing certain things (such as icons or folders) from the Windows "start" menu; (2) modifying the initial Windows boot sequence; (3) installing software that launched automatically upon completion of the boot sequence; (4) installing icons dissimilar in size or shape to Microsoft icons; and (5) using Microsoft's "Active Desktop" to display third-party brands. *Microsoft II*, 84 F. Supp. 2d at 61.

166. *Microsoft IV*, 253 F.3d at 357.

167. *Id.* at 359.

users would know how to use Windows on any machine. The D.C. Circuit rejected this argument, as had the district court, on the ground that OEMs had an incentive not to confuse consumers because OEMs bore the brunt of user support costs.[168]

The court of appeals also said Microsoft had not "substantiated" its claim that it needed to control the user interface to avoid consumer confusion. However, it is not clear what sort of substantiation the court sought, nor why Microsoft would have to substantiate consumer user interface confusion when the government did not have to substantiate actual professional software developers' confusion, a far less intuitive claim that the court of appeals nevertheless upheld.[169] And even more disturbing, the court also said that whether icons do or do not appear did not "self-evidently affect either the 'stability' or the 'consistency' of the platform,"[170] a comment that conflates the code itself with user perceptions of the code. As noted above, in the remedy phase, the court was quite content to focus on user perceptions of the code rather than actual commingling of the code, so its elevation of code over perception on this issue is hard to understand.

As the court noted, OEMs have an incentive not to confuse consumers. At the same time, however, OEM changes to the user interface still could lower the value of that interface as a work. A consumer who works on machines from different OEMs and spends time figuring out different configurations of Windows is likely to value Windows less than a consumer who knows how Windows works every time. OEM support costs would be an imperfect proxy for diminution in the value of the interface because not all confused consumers would call an OEM for support. Some would just blame the operating system. Nothing would stop an OEM itself from blaming a problem on Windows.[171]

The court's questionable analysis on this point relates both to its stated test for liability and to the question of error cost. The court of appeals seems to have balanced the benefits of the uniformity argument against the costs of reduced OEM distribution of potential platform substitutes, but it did not say it was doing so. The court instead simply rejected a perfectly plausible (though not necessarily compelling) argument as if it was incoherent. As with allegations we have examined previously, the court used relatively categorical reasoning to dispose of arguments pertaining to very

168. *Id.*

169. *See supra* notes 114-17 and accompanying text.

170. *Microsoft IV*, 253 F.3d at 370.

171. Trial Transcript, *supra* note 128, at 3636 (testimony of Carl Shapiro on April 15, 2002).

specific acts, thus avoiding an overt balancing of the costs and benefits of those acts.

It may not be fair to fault the court for the limits of its explanation, however, because there are fairly significant limits to the degree to which anyone may explain balancing. These limits were apparent in the court of appeals' treatment of Microsoft's second type of restriction, regarding Windows' initial boot sequence. These restrictions prevented OEMs from substituting their own user interfaces for Microsoft's or from using the initial boot sequence to promote products (such as Internet access from a provider that distributed Netscape's browser) that might threaten Microsoft's monopoly.[172] Both the district court and the court of appeals held these restrictions were anticompetitive because the restrictions effectively denied to firms such as Netscape access to the most efficient means of distributing browsers.

For the most part, Microsoft made the same arguments regarding these restrictions as it did regarding the first type of restrictions. And, for the most part, the court of appeals treated these arguments the same way: Microsoft lost. As to one argument, however, the court of appeals agreed with Microsoft's position, and it did so through explicit balancing. The court said

> We agree that a shell that automatically prevents the Windows desktop from ever being seen by the user is a drastic alteration of Microsoft's copyrighted work, and outweighs the marginal anticompetitive effect of prohibiting the OEMs from substituting a different interface automatically upon completion of the initial boot process. We therefore hold that this particular restriction is not an exclusionary practice that violates § 2 of the Sherman Act.[173]

This conclusion seems very sensible at first glance. Still, it is hard to see exactly why the court reached it. The court does not explain how it has measured the "marginal anticompetitive" effect of substitute interfaces, nor how one could weigh that effect against Microsoft's legal interest in avoiding "drastic alteration" of Windows. The court seemed to give weight to precedents holding that substantial alteration of copyrighted works is infringement, but it also distinguished those cases on the ground that they did not involve antitrust claims.[174] That distinction applies across the board, however, and it is hard to see how a court could possibly weigh

172. *Microsoft IV*, 253 F.3d at 61-62.
173. *Id.* at 63.
174. *Id.*

a legal interest in the integrity of a work against the economic concept of an increase in the distribution costs of rival firms.

In addition, the court's distinction between "drastic" alterations of a work, which even a platform monopolist may prevent, and non-drastic alterations, which a monopolist may not prevent on pain of section 2 liability, is unclear. We know that OEM substitution of interfaces is a drastic alteration and that prohibitions on meddling with icons or start menu folders are not drastic, but we do not really know why. The court's rejection of Microsoft's user interface uniformity argument might imply that the notion of a drastic alteration refers to alterations to the code itself, rather than what the user perceives, but the court did not say that substitution of interfaces altered the Windows code, and the most obvious consequence of interface substitution is that consumers will not know whose operating system they are using—an issue of perception, not code.

The remand proceedings provided a more concrete basis for questioning this aspect of the decision. The non-settling states proposed that Microsoft be forbidden from preventing a third party supplier of computers (such as AOL) from displaying a non-Microsoft desktop so long as the third party provided on that desktop a way for consumers to display the Windows desktop.[175] The district court rejected this proposal, citing the portion of the court of appeals' decision quoted above, but neither court ever really analyzed whether the costs of allowing other firms free reign to substitute desktops, thus allowing them to market a wide variety of computers, exceeded the gains from allowing such freedom.[176] The court of appeals simply proclaimed that it had balanced seemingly incommensurable considerations and found that Microsoft's legal interests outweighed whatever benefits might come from OEM interface substitution. The district court simply cited this proclamation. QED.

Against all this, one could argue that it is unfair to criticize the court's balancing because it is obvious that Microsoft should have the legal right to prevent an OEM from such drastic alteration of Microsoft's work. Common sense seemingly compels the conclusion that Microsoft should be allowed to insist that users look at an interface that tells them whose

175. *Microsoft V*, 224 F. Supp. 2d 76, 195 (D.D.C. 2002); Plaintiff Litigating States' Remedial Proposals at 9 (proposed text for ¶ 2.c.iii), *Microsoft IV* (No. 98-1233) [hereinafter States' Remedial Proposals], *available at* http://www.naag.org/issues/microsoft/pdf/011207-states-remedy.pdf.

176. This proposal could be criticized on the ground that it contemplated competition within the Windows standard, as opposed to competition in the operating systems market generally, so that any gain to competition would probably be slight. However, that is not the way the court of appeals framed the argument.

operating system they are using. The point is not wholly obvious, however. Microsoft could still pay OEMs not to substitute other interfaces for its own, producing an auction in which OEMs rather than Microsoft reaped returns from the valuable real estate of the first desktop screen. Maybe such bidding would constrain Microsoft's ability to foreclose the OEM distribution channel (though if rivals have to bid against Microsoft for desktop space their costs might be as high under a bidding regime as they were in light of the conduct for which Microsoft was held liable). Such bidding might only benefit OEMs at Microsoft's expense, of course, but one could say the same of the remedy adopted on remand.

In fact, the consent decree Microsoft and the Justice Department negotiated allows Microsoft to give OEMs consideration[177] to promote Microsoft products so long as the consideration is commensurate with the absolute amount of OEM support and does not discriminate against OEMs that promote rival technologies.[178] This provision is of questionable benefit to the non-settling states. The nondiscrimination restriction tempers the effect of such benefits somewhat, but even though Microsoft could not discriminate among OEMs based on their support for rival technologies, an OEM that wanted to support a rival instead of supporting Microsoft would have to forgo Microsoft's compensation. An OEM making that choice would incur an economic cost in the amount of the compensation it otherwise would have received. A rival that wanted to use an OEM to distribute its products presumably would have to compensate the OEM for that cost, which would in turn raise the rival's costs relative to a world in which Microsoft was forbidden from offering such inducements.[179]

In explaining why this provision was acceptable, the district court noted that the court of appeals had insisted that even a monopolist could compete by offering its product at an attractive price, a concept that included technical information and other support in addition to cash.[180] Threats to withhold such benefits, however, were unlawful. Thus under the decree the court approved,

---

177. A term defined broadly to include "any monetary payment or the provision of preferential licensing terms; technical, marketing, and sales support; enabling programs; product information; information about future plans; developer support; hardware or software certification or approval; or permission to display trademarks, icons or logos." Second Revised Proposed Final Judgment, *supra* note 149, § VI.C.

178. *Id.* § III.A.

179. Indeed, in the liability phase the district court found as a fact that some OEMs view desktop positioning as an opportunity to reap bounties from software firms. *Microsoft II*, 84 F. Supp. 2d 9, 60 (D.D.C. 1999).

180. *Microsoft V*, 224 F. Supp. 2d at 164.

> Microsoft will be permitted to provide compensation for OEM action which promotes or supports Microsoft products, but Microsoft cannot withhold such consideration or other consideration based upon OEM action which tends to favor non-Microsoft products. While these two goals may appear to be somewhat at odds, the liability in this case all but demands this level of hair-splitting.[181]

The court of appeals did not dispute that, from an economic point of view, this distinction is very fine indeed. Instead, the court described the provision as a matter of choice. The remedy gave OEMs the choice "either to distribute non-Microsoft middleware or to get a discount from Microsoft."[182] One could question how much of a choice this really was. In the liability phase, both the district court and the D.C. Circuit had stressed that OEM profit margins were very small, which was one reason why Microsoft's integration of code and restrictions on altering the desktop were unlawful.[183] The courts also stressed that installing more than one version of a program such as a browser tended to increase OEM costs.[184]

These facts suggest OEMs would be hard pressed to turn down Microsoft's offers, and would be unlikely to pre-install duplicative programs unless the producer of such programs offered similar incentives (raising the producer's distribution costs relative to a world in which OEMs might not demand such payments). The choice the court emphasized therefore may be more apparent than real, leading again to the inference that the conduct should either be declared lawful (my view) or subjected to stronger remedies.

As Professors Salop and Romaine suggested,[185] balancing is the method that best implements the integration approach to error costs. However, analysis of the OEM licensing arguments demonstrates that implementing the integration approach through balancing leads more to arbitrariness than precision or, if one prefers, achieves the appearance of precision through subjective decisions about whether to balance or how to balance. To the extent these allegations call into question the integration approach, they support the market correction approach.

---

181. *Id.*
182. *Microsoft VII*, 373 F.3d 1199, 1227 (D.C. Cir. 2004).
183. *See supra* text accompanying notes 139-41.
184. *Id.*
185. Salop & Romaine, *supra* note 6.

D. The F/OSS Remedies Arguments

For the most part, the claims discussed so far looked backwards at things that had already happened. The debates were about whether courts should condemn what had happened as a violation of the antitrust laws and, if so, what measures would best remedy the violation. This retrospective analysis did occur in the context of theories predicting how competition might have developed had the conduct not occurred, and in that sense one might say there was a prospective element to them. However, the nub of the claims involved the type of retrospective analysis that characterizes most litigation.

In this Section, I examine some explicitly forward-looking claims the government advanced in connection with its proposed remedy. These claims rest in part on aspects of F/OSS production the government saw as desirable. I relate these claims to the error cost debate by testing the soundness of the government's reasoning about F/OSS development.

As noted, the government asked Judge Jackson to split Microsoft vertically into two firms, an applications firm and an operating systems firm, which he did. Though the remedy did not survive appellate review, the government's justification for its proposed remedy is notable for the appearance of the GNU/Linux OS. According to the government, the separate applications firm would have had an incentive to make its applications work on any operating system, thus broadening its market and creating a situation where a consumer's desire to have a full range of applications did not limit the consumer to a single operating system such as Windows.

Professor Carl Shapiro agreed that such a breakup would help increase competition in the operating systems market. He cited Apple's MacOS and GNU/Linux as examples of operating systems that might compete more robustly with Windows if Microsoft were split in two.[186] The MacOS is, of course, owned and controlled by Apple, which would have to earn a profit on its work in order to keep producing it. As noted in Part II.C, GNU/Linux is licensed under the GPL, which means any licensee may distribute the code, modify it, and distribute the modifications, so long as the licensee allows others to do the same with its own work. GNU/Linux's status as F/OSS, licensed under the GPL, distinguishes it from the MacOS in two potentially significant ways. The first distinction has to do with pricing, and the second with the risk of strategic behavior.

Professor Shapiro noted the first distinction in his declaration. One could object to the government's vertical divestiture proposal on the

186. Declaration of Carl Shapiro April 28, 2000 at 3, *Microsoft V* (No. 98-1233).

ground that if we wound up with one dominant application firm and one dominant operating system firm, the government would have created a double-marginalization problem, thus raising prices relative to a world in which the two firms were integrated (as is the case with Microsoft). An antitrust remedy that raises consumer prices is on shaky ground. Professor Shapiro anticipated this objection, however. He testified in his declaration that "as a theoretical matter this concern is very likely outweighed by the lowering of entry barriers into operating systems that the reorganization will cause, especially when one considers nonprice as well as price considerations, specifically the innovation that will be stimulated by the reorganization."[187] He footnoted the point by saying "consumers stand to benefit as a cheaper operating system, namely Linux, becomes more attractive."[188]

Certainly the cost of operating systems could go down if the GNU/Linux OS became a viable substitute for Windows. The GPL gives any licensee the right to redistribute the code, and that is bound to constrain the price any vendor can charge. Moreover, the reduction in the price of operating system code could be greater than any price increase due to double marginalization, thus yielding a net benefit to consumers.

This analysis fails to capture some aspects of the F/OSS model, however. For example, the GPL allows anyone to modify GNU/Linux code and distribute modified versions of the code. If the GNU/Linux code base were to fork into different versions, users would face several important questions. Is the code backwards compatible with previous versions, for example, or has the copy you acquire been altered in a way you are not competent to detect? The risk of forking has parallels in Unix development, the Java allegations in the case itself, and near-misses in the history of GNU/Linux.[189]

Uncertainty, and the rational reluctance of ordinary consumers to monitor F/OSS development practices, has led to the emergence of firms such as Red Hat, which sell branded versions of Linux, employing trademark rather than copyright as the intellectual property right most relevant to competition.[190] Importantly, however, these firms also sell consulting services and support that help firms tailor Red Hat Linux for particular purposes and maintain their Linux installations.[191] The GPL does not apply to the sale of consulting services or support. Nothing stops a firm such

187. *Id.* at 15.
188. *Id.* at 15 n.25.
189. *See supra* Parts II.B and II.C.
190. *See, e.g.*, WEBER, *supra* note 60, at 108, 199-201.
191. *Id.*

as Red Hat or SuSe from selling such services on per-server basis, which is in fact Red Hat's business model.[192] On this model, a customer that purchased one copy of (for example) Red Hat Linux and copied it over to 9,999 machines, as Google does, might receive support for only one machine. It would have to pay for the other 9,999 copies to obtain support for the full installation. It is therefore possible that even if GNU/Linux entered the Windows market, and the proportion of the total cost of computing attributable to the code itself went down, the total cost of computing including complementary services might go up (or down, or stay the same).

One might argue that this possibility is so far-fetched that, properly discounted, it does not amount to much. The history of antitrust intervention in markets related to general purpose computing gives us no reason to be sanguine on this score, however. As Part II shows, the government has consistently been behind the curve of technology, and its most significant effects have been unintended. Because the government's track record gives no reason to have confidence in such predictions, it is fair to ask why that track record should be disregarded with respect to this prediction. If it should not be discarded, it justifies a fair degree of skepticism.

The government's somewhat tentative endorsement of GNU/Linux as a cure for the double marginalization risk did not discuss the risk of a fork in the GNU/Linux code base. The risk of forking was discussed, however, in connection with the second appearance of F/OSS development practices. The non-settling states proposed a remedy requiring Microsoft to release the source code to Internet Explorer and license that code on open-source terms.[193] The states argued that much of Microsoft's alleged misconduct was designed to make Internet Explorer the dominant browser, so a remedy for that misconduct should deprive Microsoft of the dominance it had unlawfully acquired.[194]

Microsoft attacked the remedy on the ground that open-source projects are vulnerable to forking and fragmentation, which could reduce the utility they offer both developers and consumers. Professor Shapiro agreed that

---

192. Red Hat's 2004 10K states:

> We have created a business model based on a suite of enterprise software products and technology-based systems management services (Red Hat Enterprise Linux and Red Hat Network), which are developed as open source technologies. We sell these technologies and services to our customers in the form of annual subscriptions on a per-server basis.

Id. at 3.

193. States' Remedial Proposals, *supra* note 175, at 17 (proposed text for ¶ 12). The proposal did not specify which of the licenses the states had in mind, if any, but the parties argued as if the GPL were the model.

194. *Id.* at 28 (proposed text for ¶ 19).

open-source projects need to be managed and coordinated to combat fragmentation. He testified that it would be desirable for Microsoft to act as the maintainer or coordinator of the proposed Internet Explorer open-source project,[195] though the remedy proposed by the non-settling states did not provide for such a role. This testimony captured the difficulties the government encountered in *United States v. Microsoft Corp.* The question of maintaining Internet Explorer as an open-source project exemplified the degree to which the government's case rested on an attempt to calibrate precisely the manner in which Microsoft could serve as a maintainer of a standard that, as a standard, was socially valuable, without going too far, as that notion might be measured by some metric as to which opinions will necessarily vary. Good luck.

The district court refused to order that Microsoft release Internet Explorer under an open-source license. It believed this remedy was not tied to Microsoft's liability because the remedy aimed to help operating systems vendors rather than middleware providers.[196] That argument missed the idea that it was platform competition that mattered, not the level at which that competition took place. The court of appeals was not concerned with this point, however, concluding that it was within the district court's discretion to be satisfied with re-opening distribution channels. Referencing its own two-tiered approach to causation, the court of appeals said the district court properly applied more stringent standards to what it reasonably viewed as a structural remedy (effective divestiture of Internet Explorer as a program) than it would apply to other sorts of remedies.[197]

The second difference between the GNU/Linux OS and ordinary proprietary systems such as the MacOS has to do with the risk of strategic behavior. As the Netscape allegations began to wither, the Microsoft litigation turned into a kitchen-sink style examination of Microsoft's allegedly strategic behavior. The costs and benefits of such behavior can be very hard to assess, leading some to favor measures that eliminate the need for such assessment by eliminating the risk of such behavior. One could argue that the rights the GPL grants to licensees would be one method by which strategic behavior could be constrained.[198] After all, if

195. Trial Transcript, *supra* note 128, at 3539-40 (testimony of Carl Shapiro on April 15, 2002).

196. *Microsoft V*, 224 F. Supp. 2d 76, 242-43 (D.D.C. 2002).

197. *Microsoft VII*, 373 F.3d 1199, 1228-29 (D.C. Cir. 2004).

198. In congressional testimony regarding remedies in the *Microsoft* case, for example, Professor Lessig opined that free or open-source software "is unlikely ever to pose any of the same strategic threats that closed source software does." *The Microsoft Settlement: Hearing Before the Senate Committee on the Judiciary*, 108th Cong. (2001) (testi-

anyone can modify and redistribute your code, how could you use it strategically?

It is possible that open-source projects would turn out to present fewer risks of strategic behavior than proprietary projects, but the point is not obvious. The history of competition in markets related to general-purpose computers and the D.C. Circuit's acquittal of Microsoft's Java development efforts both give reason to suspend judgment on this claim. To take the second point first, if it was legal for Microsoft to fork Sun's Java technologies, then presumably it would be legal for Microsoft to fork an open-source project such as the GNU/Linux OS. Under the GPL, Microsoft would have the right to modify and redistribute the GNU/Linux code, so long as it provided the source code to its modifications and allowed others to modify and redistribute them, too. Because licensees could copy and redistribute Microsoft's improvements, Microsoft would not make much money on them. Still, the history of Unix and Java suggests that the risk of a significant fork in the code would itself deter adoption of the entrant system.

No doubt one could make various arguments to distinguish this hypothetical case from Microsoft's modifications of Sun's Java technologies, but it is not clear whether such arguments would work, especially if Microsoft could point to efficiencies in its version of the GNU/Linux OS. The court of appeals' rule seems too weak to combat such behavior, while its rulings on related issues, such as distribution agreements, seem too strong. There is no reason to expect the unhappy equilibrium of the Java aspects of *United States v. Microsoft Corp.* to get any better just because an open-source project comes into the picture.

Perhaps more to the point, development of the GNU/Linux OS *is* currently driven by strategic concerns. That development is being backed by firms, such as IBM and Red Hat, which would profit from a commoditized operating system platform.[199] From this perspective, the government's endorsement of GNU/Linux in connection with its initial remedy proposal constitutes endorsement of a particular strategy, not a path to avoid strategic behavior. It is ironic that this strategy is being pursued aggressively by IBM, the firm the Justice Department spent half the 20th Century worrying about, and that the origins of the strategy can be traced to a development model that emerged in part as the unintended consequence of a consent decree aimed at different conduct. But it is a strategy just the same.

---

mony of Prof. Lawrence Lessig, Stanford Law School), *available at* 2001 WL 1580448 (F.D.C.H.).

199. *E.g.*, WEBER, *supra* note 60, at 199-200.

The history of governmental involvement in these markets provides little reason for optimism regarding what seems to have been its implicit endorsement of F/OSS production methods.

IV. CONCLUSION

Holmes famously said the life of the law has been experience, not logic.[200] Canonical though it is, something about this comment grates on the academic mind. First, experience is varied and subjective, not rigorous and scientific. Different people experience different things differently, and they draw different lessons from similar experiences. Law professors and judges certainly have no monopoly on experience. It is not even clear that they have a comparative advantage over ordinary people in distilling useful lessons from experience. Decisions grounded only in such a highly subjective and debatable concept as experience will appear biased—at least when compared to modus ponens or modus tollens—and unsophisticated, compared to a mathematical proof. Second, and perhaps most damning of all, experience looks backwards, not ahead, making it seem a weak foundation for law, which must change as the society law regulates changes.

For these and related reasons, it is difficult to mount an analytically rigorous defense of the claim that experience should trump logic. In the context we have assessed here, Professor Williamson was right to assert that the law should look forward and do its best to evolve, rather than look backward and refuse to change. He was right to imply that such a refusal had to be explained by something other than economic logic, making it vulnerable to attack as more an ideological than an economic stance. Professor Williamson and Professors Salop and Romaine were quite right to insist that logic both demands and accommodates the integration of error cost analysis into the more general analysis of allegedly anticompetitive conduct.

If one can defend Holmes on normative grounds as well as historical ones, however, then the logical shortcomings of the market correction approach to error costs do not compel its rejection. The analysis of this Article suggests a few lines of defense, which in turn imply a recommendation for improving the D.C. Circuit's standard for monopolization liability.

The first line of defense is to deny that there is a non-ideological position in this debate.[201] It is true that a certain degree of libertarian thought

200. OLIVER WENDELL HOLMES, JR., THE COMMON LAW 1 (1881).

201. *See* Andrew Gavil, *Exclusionary Distribution Strategies by Dominant Firms: Striking a Better Balance*, 72 ANTITRUST L.J. 3, 30 (2004) ("How one strikes the balance

runs through the arguments for the market correction approach, but it is also true that a certain degree of faith in governmental (including judicial) administration and regulation of markets runs through the integration approach. Indeed, some of the appeal of the integration approach probably stems from the mathematical aesthetics of equilibrium analysis, which is a far stronger force in economic analysis of the law than is commonly admitted. One might argue about how to characterize this appeal, but for present purposes we may treat it as simply a variant ideology with a disturbingly utopian view of how well judges can run an economy.

Even when courts say they are balancing costs and benefits, the almost-always-lacking data are replaced with something, generally with bets and presumptions, which are not derived from abstract logic. For this reason, it would be wrong to cast the error cost debate as being between logic and ideology. It is a debate over how to interpret antitrust experience and the degree to which such experience should temper abstract logic. As noted earlier, as an abstract matter, the domain of decision-theoretic cost-benefit analysis is unlimited. No rule of logic forbids antitrust courts from stamping out market power by determining the price of every transaction. To the extent cost-benefit analysis or decision theory precludes such a result, it is only because experience provides information that, when plugged into the equation, draws boundaries.

Debates over decision theory, and thus the integration approach, therefore will either be too abstract to be useful or will be influenced by debates over what lessons should be drawn from history and how those lessons should influence future analysis. Because the integration approach and the market correction approach differ even though it is theoretically possible for them to coincide, one must explain the difference. Logic does not do the job, suggesting that ideology must play a role.

If I am right about these points, then the market correction approach is not discredited by the fact, which I concede, that there is an ideological aspect to it. That aspect is a constant in this debate, and therefore cannot distinguish between competing positions. On what ground, then, should the choice be made? Here I claim two comparative advantages for experience over logic.

First, though one cannot prove the future will be like the past, when one looks at both the history in Part II and the opinions in *United States v. Microsoft Corp.*, one sees a fairly damning indictment of the Justice Department's monopolization theories. With respect to the general history of

---

between error costs—between over- and under-deterrence, for example—may be more a product of viewpoint than hard economics.").

this market, the Justice Department either accomplished nothing or, depending on how one views the Unix experience, accomplished nothing that was intended, which amounts to the same thing for present purposes.

With respect to the *Microsoft* litigation, the verdict is that it produced nothing good. At tremendous expense, it produced analytically shaky precedent (the foundations of which were called into question by the evidence in the remand proceedings) that satisfies no one. On my view, the opinions condemn conduct without providing any reason to believe that the conduct extended Microsoft's market power. That the remedies were toothless only mitigates the harm. On the view of analysts with more interventionist instincts, the court condemned some conduct that harmed competitors but excused conduct that caused them far more harm, and exacerbated that failing by accepting a toothless remedy. Either the case should not have been brought or it should have been pursued to the end of the premises that justified bringing it.

Second, when the past looks like the history surveyed in Part II, it is fair to ask what reason there is to expect it to get better. The natural answer is that economic analysis is more sophisticated than in the past, as is our understanding of the limitations of the courts. Exactly the same answer, in other words, which would have been given in 1949, 1956, 1969, or 1982. There is no particular reason to have faith that this answer implies good future results, and significant reason (based in experience) to discount it severely. Experience trumps logic because law is applied, not abstract. One cannot get by in the real world by ignoring what happens there.

The problem with such arguments, of course, is that they do not meet the most serious objection to the market correction approach, which is Professor Williamson's point that one cannot make progress if one looks only at the past. He was right, and no amount of quibbling will change that. If we are to reconcile this fact with the unhappy history that goes along with it, we need a doctrinal tool that leaves room for new learning but demands a high degree of certainty that applying that learning will enhance competition in the real world.

My suggestion is to reverse the D.C. Circuit's causation analysis in *United States v. Microsoft Corp.* That is to say, at least where the government or a private plaintiff does not challenge the legality of an initial market position, monopoly maintenance liability should not attach unless the government or a private plaintiff can prove that the acts in question prolonged a defendant's market power past the point at which that power would have eroded significantly had the acts not been taken. Once that point is proved, however, the defendant should suffer real pain. I am will-

ing to bet that evidence sufficient to meet such a rigorous standard will also provide guidance for adopting a cogent remedy.

Notions of finely balanced liability findings and finely tailored remedies are as impractical in the real world as they are elegant in theory. The traffic cop should be retired, and the SWAT team should be called in only when strong medicine is plainly warranted. That is the lesson of *United States v. Microsoft Corp.*

EXCLUSIONARY CONDUCT UNDER THE ANTITRUST LAWS: BALANCING, SACRIFICE, AND REFUSALS TO DEAL

By A. Douglas Melamed†

TABLE OF CONTENTS

Important recent antitrust cases involving allegations of monopolization and exclusionary conduct, including the Supreme Court's decision in *Trinko*[1] and lower court decisions in *Microsoft*,[2] *LePage's*,[3] and *Xerox*,[4] have focused attention on the evolving and uncertain standards regarding

---

© 2005 A. Douglas Melamed

† Partner, Wilmer Cutler Pickering Hale and Dorr LLP, Washington, D.C. I am grateful to Steve Salop for a continuing and illuminating dialogue about the issues addressed in this Article, and to James Frost and Barbara Blank for research assistance. A prior draft of this Article was presented at an antitrust panel, Evolving Antitrust Treatment of Dominant Firms, at the Association of American Law Schools (AALS) 2005 Annual Meeting.

1. Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398 (2004).

2. United States v. Microsoft Corp., 253 F.3d 34 (D.C. Cir. 2001) (en banc).

3. LePage's Inc. v. 3M Co., 324 F.3d 141 (3d Cir. 2003) (en banc), *cert. denied*, 124 S. Ct. 2932 (2004).

4. CSU, L.L.C. v. Xerox Corp. (*In re* Independent Serv. Orgs. Antitrust Litig.), 203 F.3d 1322 (Fed. Cir. 2000), *cert. denied*, 531 U.S. 1143 (2001).

exclusionary conduct. Because of the *Trinko* case in particular, refusal to deal issues have been especially controversial.

This Article discusses a possible approach to exclusionary conduct, which is sometimes called the "sacrifice test." Part I describes the problem of exclusionary conduct in general. Part II explains that one's approach to exclusionary conduct depends to a large extent on whether one views antitrust as law enforcement or as regulation. Parts III and IV discuss pros and cons of different approaches to exclusionary conduct. Part V discusses refusals to deal in greater detail. It argues that refusals to deal do not warrant special antitrust rules and explains how general principles applicable to exclusionary conduct can be applied to refusals to deal. Part VI sets forth some concluding comments.

I. THE PROBLEM OF EXCLUSIONARY CONDUCT

Broadly speaking, the antitrust laws are concerned with two types of anticompetitive conduct. One is collusion: conduct in which two or more firms agree to reduce rivalry between them in order to enable one or both to exercise market power. The other is exclusion: conduct by a firm or group of firms that weakens rivals or excludes them from the market and can thereby enable the firm or firms engaging in the conduct to gain market power.[5]

Although problems of collusion are not trivial, antitrust law has generally found them much easier to deal with than exclusion problems. There appear to be two reasons for this. First, naked collusion—that is, conduct like price fixing that has no welfare-enhancing properties—is not uncommon[6] and provides a focal point for the development of a jurisprudence

5. Sections 1 and 2 of the Sherman Act, 15 U.S.C. §§ 1, 2, are directed at these types of anticompetitive conduct. Section 1 prohibits anticompetitive agreements and encompasses both collusion (for example, price fixing by competitors) and exclusionary agreements (for example, exclusive dealing agreements). Section 2 applies to single firm conduct and encompasses only exclusionary conduct. Section 7 of the Clayton Act, 15 U.S.C. § 18, which prohibits anticompetitive mergers, can also be said to encompass both types of anticompetitive conduct.

6. *See generally* Scott D. Hammond, An Overview of Recent Developments in the Antitrust Division's Criminal Enforcement Program, Address Before the American Bar Association Midwinter Leadership Meeting 5 (Jan. 10, 2005) (noting that, since 1997, over $2 billion in criminal fines have been imposed by the Justice Department with "well over 90 percent of this total . . . obtained in connection with the prosecution of international cartel activity"), *at* http://www.usdoj.gov/atr/public/speeches/207226.pdf; Scott D. Hammond, A Review of Recent Cases and Developments in the Antitrust Division's Criminal Enforcement Program, Address Before the 2002 Antitrust Conference: Antitrust Issues in Today's Economy (Mar. 7, 2002) (discussing recent Justice Department crimi-

dealing with collusion issues. Second, and more important, while some forms of collusion, such as horizontal mergers and joint ventures, can have important procompetitive or efficiency-generating attributes, determining whether collusion is on balance procompetitive or anticompetitive is conceptually straightforward. Although the broad language regarding the rule of reason sometimes obscures this point, the issue in a collusion case is simply whether the collaboration among the competitors is likely to increase or decrease the output of the parties to the agreement. The focus is on the defendants.[7]

Antitrust law has had much more difficulty deciding how to determine whether exclusionary conduct is anticompetitive. Occasionally, antitrust law has confronted naked exclusion, such as fraud on the Patent Office in *Walker Process*[8] or damaging a rival's property and falsely disparaging the rival in *Conwood*.[9] These cases are distinctive and easy because they entail conduct that excludes rivals and that has no efficiency or welfare-enhancing properties. These cases, however, are very rare, perhaps because the conduct involved usually also violates other, non-antitrust laws.

In the vast majority of cases, exclusion is a result of conduct that has both efficiency properties and the tendency to exclude rivals. This is true of predatory pricing, exclusive dealing, tying, many types of bundling, and countless other forms of exclusionary conduct.

The challenge in exclusion cases is how the law should treat conduct that has both efficiency benefits and exclusionary harm. The benefits are usually realized at least in part by the defendants, but the exclusionary harm is experienced by rivals and indirectly by consumers. In other words, by contrast to collusion conduct, which reduces the defendants' output, exclusionary conduct reduces the output of the defendant's excluded rivals.

The law has struggled uneasily with exclusion cases because it has not yet embraced a conceptual framework for dealing with these competing considerations. The several recent exclusion cases have provoked a renewal of a long-standing and unsettled debate. Critics of cases like *Trinko*, in which the defendant prevailed, complain that the legal principles apparently embraced by the Supreme Court will have too many false nega-

---

nal convictions for price-fixing, bid rigging, and international cartels), *at* http://www.usdoj.gov/atr/public/speeches/10862.pdf.

7. To be sure, collusion cases assessed under the rule of reason require proof of market power; and that requires an examination of the market, not just of the defendants. But the market power requirement is a separate element of the collusion offense.

8. Walker Process Equip. v. Food, Mach. & Chem. Corp., 382 U.S. 172 (1966).

9. Conwood Co. v. U.S. Tobacco Co., 290 F.2d 768 (6th Cir. 2002).

tives—too many cases in which conduct that diminishes economic welfare escapes antitrust condemnation.[10] In varying ways, these critics urge an antitrust test that gives more weight to the prospect of long-run monopoly power.

Cases like *Microsoft* and *LePage's*, in which the plaintiffs prevailed, have been criticized on the very different ground that the principles embraced by the courts are likely to lead to too many false positives—to condemning and thus deterring efficient conduct.[11] Critics of these and other cases complain that antitrust courts either cannot be counted on to get difficult economic and technical issues right or they have failed to articulate intelligible standards that might give suitable guidance for future conduct.

II. ANTITRUST—REGULATION OR LAW ENFORCEMENT?

Unstated in the debate is an implicit disagreement about the nature of antitrust. Critics on the left, for example, who are concerned about false negatives, focus almost entirely on the desired end state and complain whenever welfare-reducing conduct is permitted.[12] Critics on the right often complain that antitrust is not worth the cost, in part because antitrust has failed, they say, to remedy perceived wrongs effectively and because

10. *See, e.g.*, Andrew I. Gavil, *Exclusionary Distribution Strategies By Dominant Firms: Striking a Better Balance*, 72 ANTITRUST L.J. 3, 36-51 (2004) (arguing that "the price of false negatives, of under-deterrence, in exclusionary conduct cases . . . has been seriously understated"); Marina Lao, *Reclaiming A Role for Intent Evidence in Monopolization Analysis*, 54 AM. U. L. REV. 151, 190-91 (2004) (criticizing the sacrifice test in *Trinko* on the ground that, despite the fact that defendant did not sacrifice short-term profits, its conduct hindered competition); Steven C. Salop & R. Craig Romaine, *Preserving Monopoly: Economic Analysis, Legal Standards, and Microsoft*, 7 GEO. MASON L. REV. 617, 654-55 (1999) ("Permitting the monopolist to exclude rivals may deprive consumers of the lower prices and increased innovation that would have occurred absent the exclusionary conduct."); Salop & Romaine, *supra*, at 662.

11. *See, e.g.*, Ronald A. Cass & Keith N. Hylton, *Preserving Competition: Economy Analysis, Legal Standards and Microsoft*, 8 GEO. MASON L. REV. 1, 31-32 (1999); Frank Easterbrook, *The Limits of Antitrust*, 63 TEX. L. REV. 1, 6, 15-16 (1984) [Easterbrook, *Limits*]; Frank Easterbrook, *When Does Competition Improve Regulation?*, 52 EMORY L.J. 1297, 1306-07 (2003) [hereinafter Easterbrook, *Competition*]; Kenneth G. Elzinga et al., United States v. Microsoft*: Remedy or Malady?*, 9 GEO. MASON L. REV. 633, 651 (2001) ("In antitrust cases, false positives are serious because less efficient firms will benefit and consumers will be harmed as a result of legal intervention.").

12. *See, e.g.*, Gavil, *supra* note 10, at 36-40; Lao, *supra* note 10, at 190-91.

antitrust intervention into complicated industries like computers and telecommunications has been problematic.[13]

The debate often appears to gloss over the unstated premises of the participants about the nature of antitrust. Critics on both the left and the right often seem to regard antitrust as in large part a kind of regulation, in which courts should decide cases in order to achieve the right result (that is, the procompetitive market outcome) in the particular case, or in which courts should abstain if that outcome seems beyond their remedial powers.

Antitrust is better and more accurately understood to be a form of law enforcement, not regulation. Antitrust is a form of law enforcement because it depends on courts, not regulators, both for its enforcement and for its doctrinal evolution through a common-law type process. Unlike regulation, antitrust does not specify end states or required market conditions, and it does not entail affirmative commands or require prior government approval as a condition of private conduct.

Accepting the premise of antitrust as law enforcement not only is compelled by its nature, but also is important in order for antitrust to serve its contemporary substantive purposes. Antitrust rests on the premise that a decentralized market is most likely to create incentives for, and to take advantage of multiple sources of, creativity and entrepreneurship, thereby maximizing economic welfare. Antitrust thus presumes that government intervention should, as a general matter, be modest and should be undertaken only when the rules are clear and understandable so that uncertainty about the rules does not inhibit competitive and entrepreneurial forces that antitrust is intended to encourage. Core principles of a law enforcement regime—that law should give clear notice to affected parties so that they will know what is required of them, that legal decisions should turn on tractable factual issues, and that like cases should be treated alike—help make law predictable, thereby further robust conduct by economic actors, and thus promote antitrust objectives.

The idea of antitrust as law enforcement has important implications. First, it means that antitrust law's principal focus should be on identifying conduct that should be prohibited. Issues of remedy are second-order considerations. One would not, for example, suggest a diminished role for laws prohibiting murder on the ground that, once the murder has been committed, there is no good remedy.

---

13. *See, e.g.*, Frank Easterbrook, *When Is It Worthwhile To Use Courts To Search for Exclusionary Conduct?*, 2003 COLUM. BUS. L. REV. 345, 348-50, 352-54, 357-58. *See generally* Elzinga et al., *supra* note 11.

Second, the idea of antitrust as law enforcement means that, while antitrust rules should be fashioned with an eye to the competitive consequences of the prohibited and permitted conduct, selection of antitrust rules depends critically on their administrability. Administrability has two basic components. It refers both to the ability of courts and antitrust enforcement agencies to administer the rules after-the-fact and to the ability of businesses to know what conduct is permitted and what is prohibited.

The principal function of an antitrust rule in a law enforcement regime is to create appropriate incentives for the avoidance of welfare-reducing conduct. Creation of such incentives requires articulation of rules with which businesses can comply without excessive transaction costs or uncertainty. To this end, sound antitrust rules will identify factors that are both likely to have welfare-reducing or anticompetitive consequences and that firms in real time can recognize and use as guideposts to inform their conduct. In other words, while the law in a regulatory regime might focus on desirable and undesirable conduct and require the former while prohibiting the latter, the law in a law enforcement regime needs to temper its enthusiasm for theoretical precision with an appropriate accommodation for the practical limitations upon firms that must comply with the law and courts and agencies that must enforce it.

As a crude analogy, the law has posted speed limits on roads, not because welfare is always reduced by those who drive faster or never enhanced by those who drive more slowly, but because a simple, posted speed limit is easier to administer fairly ex post, because it gives more useful guidance to drivers in real time (ex ante) and because it identifies readily recognizable types of conduct that are likely to create an unreasonable hazard.[14] Much of the discussion that follows emphasizes the important constraints on antitrust law that arise from its law enforcement character.

III. PROBLEMS WITH BALANCING TESTS FOR IDENTIFYING EXCLUSIONARY CONDUCT

As a matter of pure theory, the problem of exclusion in antitrust is not obscure. Although most exclusionary conduct has some efficiency benefits, determining whether the conduct is on balance anticompetitive is straightforward: Calculate the magnitude of the benefits—increased consumer welfare from lower prices or improved quality, and perhaps increased total welfare from cost savings—and calculate the amount of welfare loss attributable to the exclusion of rivals—reduced consumer surplus

14. *See generally* FTC v. Superior Court Trial Lawyers Ass'n, 493 U.S. 411, 433-34 (1990).

or perhaps the deadweight loss attributable to market power maintained or created by the conduct. If the latter is greater than the former, the conduct is anticompetitive.[15]

Here is how such a balancing test might work.[16] Suppose that the defendant builds a better mousetrap, that consumers prefer the mousetrap and buy it in large quantities, and that, as a result, the defendant's rivals exit the business. Thereafter, the defendant uses its resulting monopoly power to increase price by a large amount. Suppose further that the price increase endures because the rivals and other potential entrants know that the defendant could profitably undersell them with its more valuable mousetrap, and they therefore do not enter. Depending on the magnitude of the quality improvement offered by the defendant's mousetrap and the length of time before the price increase, compared to the magnitude and duration of the monopoly price increase, the development and marketing of the better mousetrap might be found to reduce economic welfare and therefore to be anticompetitive.

One could come up with examples that seem to be more sympathetic to the antitrust plaintiff, perhaps examples that involve conduct that does not just increase the attractiveness of the defendant's products or reduce its costs but also raise its rivals' costs. Suppose, for example, that development of the better mousetrap requires the defendant to have an assured and regular supply of a critical input and that the contracts entered into by the defendant to obtain that supply have the effect of making it harder for the defendant's rivals (who are producing a less valuable mousetrap) to obtain access to needed supplies of the input. To make plaintiff's case even more attractive, assume that a number of the rivals' customers are perfectly content with the lower value, lower cost mousetrap offered by the rivals but are now disadvantaged by the fact that the rivals' costs and thus their prices have been increased as a result of the defendant's better mousetrap and the related supply agreements. In principle, antitrust law could weigh the benefits to consumers from the new and better mousetrap (or the contracts necessary to its development) against the harms to rivals' consumers resulting from their increased costs.

If economic actors and legal fact finders were omniscient, such balancing tests would make good sense. Courts and agencies could readily identify anticompetitive conduct, and firms in real time would be able to pre-

15. *See generally* Nat'l Collegiate Athletic Ass'n v. Bd. of Regents, 468 U.S. 85 (1984).

16. *See generally* Salop & Romaine, *supra* note 10, at 31-36, 42-48.

dict whether the possibly long-run competitive harms would outweigh the efficiency benefits and, thus, to avoid anticompetitive conduct.

The problem, however, is that neither economic actors nor law enforcement entities are omniscient. Given real world limitations, market-wide balancing tests that seek to assess the benefits and competitive harms of exclusionary conduct are intractable for courts and antitrust agencies, and even more so for firms trying to decide in real time what conduct is permitted and what is prohibited. Prospective defendants cannot be expected to know in real time, ex ante, whether their efficiency-generating conduct will cause disproportionate harm to their rivals or consumers because, in order to know that, the defendants would have to know more than they can be expected to know about consumer demand, their rivals' costs and prospects for innovation and for mitigation of harm, future entry conditions, and the like. From the perspective of the defendants, therefore, a balancing test would likely either be ignored, impose excessive transaction costs (a kind of tax on entrepreneurship), or result in excessive caution. There is little reason to expect that a balancing test would create optimal ex ante incentives for marketplace behavior.

A balancing test might also create perverse incentives for the defendant's rivals. Exclusionary conduct can reduce welfare only to the extent that it weakens or excludes rivals. The likelihood that, and extent to which, exclusionary conduct will weaken or exclude rivals often depends a great deal on how the rival responds to the conduct—on whether he responds creatively, efficiently, and effectively in the marketplace. But if the determination over whether a defendant's conduct is anticompetitive depends in large part on the impact of the conduct on the defendant's rivals and their customers, the incentive of the rivals to respond to suspected exclusionary conduct by aggressive and creative marketplace conduct of their own will be diminished. This is because effective marketplace responses might weaken the rivals' antitrust claims. In effect, the net cost of such responses would be increased, and expected net returns from such conduct would be reduced. This theoretical concern, however, is unlikely to be of practical importance, except perhaps in unusual circumstances.

Some commentators, most famously Judge Easterbrook, have noted these and other problems of market-wide balancing tests and have drawn a pessimistic inference.[17] They urge a much more modest role for antitrust based largely on a concern about the costs of false positives. Broadly

17. *See, e.g.*, Easterbrook, *Limits*, *supra* note 11; Easterbrook, *Competition*, *supra* note 11; *see also* David McGowan, *Between Logic and Experience: Error Costs and* United States v. Microsoft Corp., 20 BERKELEY TECH. L.J. 1185, 1188-89 (2005).

speaking, these costs include both specific costs that these critics have found in what they regard as misguided antitrust intervention in particular cases and general costs resulting from what they fear to be overdeterrence of aggressive, innovative conduct. These critics urge a far more limited role for antitrust, at least in the context of exclusion offenses.[18]

There is, however, an alternative to both the confidence implied by market-wide balancing tests and the extreme skepticism suggested by these critics. It is a middle ground that rests on the premise that antitrust is about law enforcement and that the principal criterion by which antitrust rules should be judged is how well the rules deter welfare-reducing conduct without reducing welfare-enhancing conduct.

IV. THE SACRIFICE TEST: A PRACTICABLE SOLUTION

The middle ground alternative is often called "the sacrifice test," although the terminology can be misleading because it connotes a short-term sacrifice in search of a long-term, anticompetitive payoff.[19] As will be seen, properly understood, the sacrifice test discussed here does not require any such temporal dimension.

Instead of balancing market-wide costs and benefits of the conduct at issue, the sacrifice test asks a different question, one that—like the question asked by antitrust law in determining whether collaborative conduct is anticompetitive collusion—focuses on the defendants. Specifically, the sacrifice test asks whether the allegedly anticompetitive conduct would be profitable for the defendant and would make good business sense even if it did not exclude rivals and thereby create or preserve market power for the defendant. If so, the conduct is lawful. If not—if the conduct would be unprofitable but for the exclusion of rivals and the resulting market power—it is anticompetitive.[20]

Although words like "profitable" and "business sense" can have more than one meaning, the sacrifice test rests on particular meanings of those terms and is not ambiguous. Specifically, the sacrifice test, or at least the version suggested here, compares the benefits and costs of the conduct in

---

18. *See, e.g.*, Easterbrook, *Limits*, *supra* note 11, at 3, 6, 9-10, 13, 15-16; McGowan, *supra* note 17, at 1188-89.

19. *See, e.g.*, Einer Elhauge, *Defining Better Monopolization Standards*, 56 STAN. L. REV. 253, 292-93 (2003).

20. Conduct that fails the sacrifice test is deemed to be anticompetitive for antitrust purposes, but it might not be unlawful. The exclusion offenses require, in addition to proof of anticompetitive conduct, proof that the conduct enabled the defendant to gain or maintain market power that it would otherwise not have.

question to the defendants as follows. The costs of the conduct are the incremental or avoidable costs incurred by the defendant as a result of the conduct, including opportunity costs. In other words, they include the costs that the defendant would not incur but for the conduct. The benefits of the conduct are variable cost savings realized by the defendant as a result of the conduct, revenues from additional units of goods or services sold by the defendant as a result of the conduct, increased revenues attributable to quality improvements, and the resulting increase in demand for the defendant's goods or services. Benefits do not include the ability to charge higher prices or to shift the variable cost curve downward (because, for example, of a diminished need to provide customer services) as a result of the exclusion of rivals.

A. Benefits of the Sacrifice Test

As a general principle for assessing exclusionary conduct, the sacrifice test has several benefits.[21] First, conduct will fail the sacrifice test only if it generates incremental costs for the defendant that exceed the incremental revenues or cost savings that the conduct creates for the defendant. In other words, the sacrifice test condemns only conduct that reduces welfare in a static sense. It does not condemn conduct that enhances welfare in a static sense on the ground that it might lead to a long-run increase in market power and a resulting welfare reduction. The sacrifice test thus presents far less risk of false positives than do market-wide balancing tests.

Second, the sacrifice test embodies a somewhat Schumpeterian intuition that courts and commentators have repeatedly expressed—the idea that firms are entitled to reap the fruits of their "skill, foresight and industry,"[22] even if those fruits include market power, and the corresponding idea that antitrust condemns only conduct that is not "competition on the merits."[23] This kind of rhetoric directs attention, not to market-wide effects, but to the nature of the defendant's conduct. As will be seen in the discussion below of refusals to deal, this rhetoric also has important implications for dynamic efficiency.

---

21. Many of the remaining points summarized in this section are addressed at greater length in A. Douglas Melamed, *Exclusive Dealing Agreements and Other Exclusionary Conduct—Are There Unifying Principles?*, 73 ANTITRUST L.J. (forthcoming 2005) (manuscript, on file with author).

22. United States v. Aluminum Co. of Am., 148 F.2d 416, 430 (2d Cir. 1945).

23. 3 PHILLIP E. AREEDA & HERBERT HOVENKAMP, ANTITRUST LAW: AN ANALYSIS OF ANTITRUST PRINCIPLES AND THEIR APPLICATION ¶ 651c (2d ed. 2002); *see also* Spectrum Sports, Inc. v. McQuillan, 506 U.S. 447, 458 (1993).

Third, by condemning only conduct that makes no sense apart from exclusion and resulting market power, the sacrifice test ensures that the antitrust laws condemn only conduct from which an anticompetitive intent can unambiguously be inferred. The test does not condemn conduct that makes good sense, regardless of resulting market power, simply on the ground that it also has the effect of creating market power.

Fourth, and most important, the sacrifice test provides simple and meaningful guidance to firms to enable them to know how to avoid antitrust liability without steering clear of procompetitive conduct. If antitrust law explicitly embraced the sacrifice test for exclusionary conduct, firms would be able to comply with the law simply by determining whether their contemplated conduct would make good business sense even if the conduct did not increase their market power.

Courts as well as firms would find the focused inquiry of the sacrifice test easier to apply than the more open-ended inquiry suggested by market-wide balancing. Indeed, even courts that have not explicitly embraced the sacrifice test appear to have sought means of avoiding market-wide balancing. The *Microsoft* case for example, which famously articulated a four step test culminating in balancing, did not actually engage in market-wide balancing.[24] To the contrary, when it came to deciding the case on the facts before it, the court did what antitrust courts usually do: it deemed anticompetitive only those aspects of the defendant's conduct that seemed to make no business sense except as a means of excluding rivals.[25]

B. Criticisms of the Sacrifice Test

In spite of these benefits, the sacrifice test has been criticized by numerous commentators who are concerned that it will result in false negatives. Commentators have also suggested other, less substantial criticisms of the sacrifice test. The most important of these criticisms are addressed below.

1. False Negatives

The most basic criticism is that the sacrifice test will have false negatives, in the sense that it does not condemn all conduct that might reduce welfare overall.[26] Proponents of this criticism complain that the sacrifice

---

24. United States v. Microsoft Corp., 253 F.3d 34, 59 (D.C. Cir. 2001) (en banc).

25. Courts typically get to this conclusion by defining the conduct at issue narrowly, and either rejecting the proffered efficiencies on the facts or accepting the efficiencies but finding that the conduct at issue was not necessary to achieve them.

26. *See, e.g.*, Elhauge, *supra* note 19, at 255, 268-72, 280-82 (criticizing sacrifice test on ground that exclusionary conduct need not require sacrifice); Gavil, *supra* note 10,

test does not weigh, against the efficiencies or profitability of the conduct for the defendant, the costs of the conduct to defendant's rivals or the costs to consumers if the conduct proves to be so successful in the marketplace that it creates market power for the defendant.

The premise of this criticism is correct. The sacrifice test does not purport to condemn all conduct that might create market power or reduce economic welfare. Rather, the test rests on the judgment that market-wide balancing tests, which in theory could condemn all welfare-reducing conduct, will in practice prove to be an inferior legal standard because of their greater difficulty in administration and their perverse incentive effects. Whether the costs of false negatives from the sacrifice test exceed the costs of false positives, increased administration costs and increased uncertainty from market-wide balancing are ultimately an empirical question. The choice between balancing tests and the sacrifice test is one of legal policy.

2. Predatory Pricing

Some have suggested that, while the sacrifice test appears to be a generalization of principles regarding predatory pricing, it is inconsistent with predatory pricing law.[27] The idea is that selling below a profit-maximizing price but above cost is not regarded as anticompetitive conduct as a matter of predatory pricing law, even though the conduct involves a profit sacrifice and makes sense only as a means of excluding rivals and facilitating recoupment through the resulting market power.

Although this criticism finds some support in the verbal formulation of the sacrifice test, there is little substance to it. Predatory pricing rules are consistent with the sacrifice test in the following three respects.

First, the sacrifice test condemns only conduct that is inefficient in a static sense, and low prices are not inefficient as long as they are above cost. To the contrary, as compared to a profit-maximizing price, a lower but above cost price increases output and consumer welfare and reduces deadweight loss.

Second, the sacrifice test is intended to give meaningful guidance to courts and firms so that they can identify anticompetitive conduct. Cost-based predatory pricing rules serve this same purpose, but rules that pro-

---

at 56-58 (arguing that the "principal flaw of a sacrifice test is its assumption that 'predation' is never of concern to the antitrust laws if it is costless," thus ignoring the numerous examples of exclusionary conduct that "involved either little or no sacrifice to profit").

27. *See, e.g.*, AARON S. EDLIN & JOSEPH FARRELL, THE AMERICAN AIRLINES CASE: A CHANCE TO CLARIFY PREDATION POLICY (Competition Policy Ctr., Paper CPC02-033, 2002), *available at* http://repositories.cdlib.org/iber/cpc/CPC02-033.

hibited above-cost pricing would not because they would turn on the elusive question of what is the profit-maximizing price. That question is elusive for courts in hindsight, and even more so for firms in real time, because the profit-maximizing price depends in large part on the response of competitors to any particular pricing strategy.

Third, the sacrifice test compares revenues with costs. The divergence in pricing cases between the profit-maximizing price and the level at which price becomes below cost is attributable to revenues from inframarginal purchases that the defendant gives up in order to gain the additional unit sales sought by the lower price. The foregone revenues from inframarginal purchases resulting from prices that are above cost but below profit-maximizing levels are not ordinarily regarded as a cost because they reflect only a wealth transfer from producers to consumers and do not require the use or consumption of any economic resources. Such foregone revenues are thus not associated with a reduction in static welfare, and they can therefore properly be disregarded in applying the sacrifice test.

3. False Positives

Professor Salop has hypothesized a situation in which the sacrifice test might result in a false positive.[28] He imagines a defendant whose innovation reduces its own costs but would so readily be copied by rivals that it would not ordinarily be profitable for the defendant to implement the innovation. But, Salop supposes, if implementing the innovation breaks an industry standard, it might impose partially offsetting cost increases on the rivals, give the defendant a competitive advantage, and thereby be profitable. Salop reasons that, under these circumstances, the innovation is profitable only because it handicaps rivals and enables the defendant to exercise market power. The innovation would thus fail the sacrifice test, even though it enhances welfare.

This criticism seems farfetched. First, if the innovation is easily emulated, it is not proprietary, and the defendant is likely to fear that others will adopt the innovation and get a first-mover advantage if the defendant does not do so. Therefore, the but-for world against which to determine whether the innovation would be profitable for the defendant without exclusion is probably not the status quo ante, but rather some world in which the defendant would be at a competitive disadvantage if it did not implement the innovation.

---

28. *See* Steven C. Salop, Section 2 Paradigms and the Flawed Profit-Sacrifice Standard 32-33 (Mar. 4, 2005) (unpublished manuscript, on file with author). A draft of Professor Salop's paper was also presented at the same AALS antitrust panel in which a prior draft of this Article was presented.

Second, it is hard to see how the defendant avoids bearing at least some of the costs of the broken industry standard. His implementation of the innovation would presumably be able to break the industry standard only if the defendant is a party to important industry complementarities. In that event, the defendant would be a beneficiary of the complementarities and would likely incur costs when the standard is broken. In other words, it is unlikely that the defendant could get much of a competitive advantage from breaking a standard.

Third, if the innovation results in a real asymmetry in costs, as posited by Salop, the defendant's best course would likely to be to drive the now higher-cost rivals out of business and then raise price. In that event, there is probably no consumer welfare gain; and condemning the innovation is probably not a false positive.

4. Costless Exclusion

Some commentators have suggested that the sacrifice test fails to condemn exclusion that is not costly to the defendant, such as fraud on the Patent Office or damaging a rival's products.[29] These critics reason that the sacrifice test requires identification of costs incurred by the defendant (the "sacrifice") and that there are no or virtually no costs in these kinds of cases.

This criticism seems mistaken in three respects. First, the sacrifice test focuses principally, not on the defendant's costs, but on the source of the benefit to the defendant from the conduct. Conduct like damaging a rival's property benefits the defendant only by weakening or excluding rivals. Such conduct would thus fail the sacrifice test.

Second, the purpose of the sacrifice test is to address the difficult issue of exclusion that arises whenever conduct both has efficiency benefits and tends to exclude rivals. Conduct that has no efficiency benefits, which is sometimes called "naked exclusion," can be condemned as anticompetitive conduct without need for the sacrifice test, market-wide balancing, or any other elaborate inquiry.

Third, almost all conduct entails some costs, even if only the relatively modest costs incurred in order to plan, execute, and presumably attempt to conceal tortious conduct. If those costs would make no sense for the de-

29. *See, e.g.*, Susan A. Creighton, Cheap Exclusion, Address at the Charles River Associates 9th Annual Conference (Feb. 8, 2005) (unpublished manuscript, on file with author); Salop, *supra* note 28, at 32-33.

fendant but for the exclusion of rivals, they would both entail a literal "sacrifice" and fail the sacrifice test.[30]

5. Administrability

Professor Salop also argues that the sacrifice test is often difficult to implement properly.[31] He has suggested various ways in which conduct that simultaneously confers benefits upon the defendant and increases rivals' costs could pose difficult challenges for a court or a firm applying the sacrifice test because it would require estimation of the profitability of the conduct in the counterfactual hypothetical world in which the rivals' costs were not increased. This criticism does not apply to conduct like price reductions or product changes that exclude rivals by reducing demand for their products, rather than by increasing their costs.[32]

While it is true that application of the sacrifice test can be difficult in cases involving simultaneous benefits for the defendant and cost increases for rivals, the sacrifice test is still easier to apply than a market-wide balancing test in at least the vast majority of such cases. This is especially likely in cases involving improvements to the quality of the defendant's goods or services. In these cases, market-wide balancing requires analyzing the same hypothetical presented by the sacrifice test—determination of what the price of the improved product would be if the rivals' costs were not increased. Moreover, market-wide balancing requires an additional and very difficult inquiry into the magnitude of the welfare impact caused by the resulting exclusion of rivals; that inquiry is not required by the sacrifice test.[33]

In cases in which the benefit to the defendant involves only cost savings and not improvements to the quality of its goods or services, the relative advantage of the sacrifice test over market-wide balancing is less. But market-wide balancing is likely to be inferior even in these cases. In these

30. The only exceptions are cases involving fraud on the patent office or other forms of theft. The reward to the defendant in those cases might be ownership of the wrongfully obtained property, regardless whether it gives the defendant market power. It is noteworthy in this respect, however, that in *Walker Process Equipment, Inc. v. Food, Machine & Chemical Corp.*, 382 U.S. 172 (1966), the antitrust violation was seeking to exercise market power by enforcing the wrongfully obtained patent, not the fraudulent conduct—the theft—itself.

31. *See* Salop, *supra* note 28, at 31.

32. Conduct that reduces demand for rivals' products could, by reducing their output below efficient scale, increase their average variable costs. But this reflects simply a movement along the rivals' cost curves. Conduct that raises rivals' cost differs in that it shifts rivals' cost curves upward.

33. *See* Melamed, *supra* note 21 (manuscript at 27-30).

cases, the sacrifice test requires determining whether the conduct is likely to be profitable even if rivals are not excluded. By contrast, market-wide balancing would require firms wishing to obey the law to undertake the always difficult task of estimating the effect of the conduct on their rivals' costs and the overall effect on consumer or total welfare of both the cost increases for rivals and the cost savings for the defendant.

V. APPLYING THE SACRIFICE TEST TO REFUSALS TO DEAL

In addition to these general criticisms, application of the sacrifice test to unilateral refusals to deal has been the subject of particular controversy. A refusal to deal is a refusal by the defendant to sell its property to, or share its property with, a rival.

A refusal to deal can, of course, cause the costs of the rival that wanted to deal with the defendant to be higher than if the defendant did not refuse to deal. But refusal to deal cases are best regarded not as a kind of raising rivals' cost case, but as a distinct kind of potentially anticompetitive conduct. As will be seen, the economic and analytical issues of refusal to deal cases differ from those in other raising rivals' costs cases. Moreover, those other cases involve conduct (such as exclusive dealing) that makes rivals' costs higher than they would be if the defendant had not acted at all. Refusal to deal cases, by contrast, are cases in which the rival wants the defendant to act so that the rivals' costs will be lower than if the defendant does not act. Refusal to deal cases are best thought of as lowering rivals' cost cases, not raising rivals' cost cases. This is an area about which there has been much confusion, and it warrants more extended discussion.

A. Conflicting Views

Numerous conflicting views have been expressed about the treatment of refusals to deal. The Solicitor General suggested in his amicus brief in the *Trinko* case that the sacrifice test is especially well suited to refusal to deal cases.[34] The Federal Circuit in the *Xerox* case suggested that there ought never be a duty to deal in patents, evidently because it believed that any such duty would prevent inventors from getting sufficient rewards for their inventions.[35] Professor Elhauge has suggested that refusals to deal

34. *See* Brief for the United States and the Federal Trade Commission at 14-17, Verizon Communications, Inc. v. Law Office of Curtis V. Trinko, LLP, 540 U.S. 398 (2004) (No. 02-682).

35. *See* CSU, L.L.C. v. Xerox Corp. (*In re* Independent Serv. Orgs. Antitrust Litig.), 203 F.3d 1322 (Fed. Cir. 2000), *cert. denied*, 531 U.S. 1143 (2001).

are the great loophole in the sacrifice test. As he describes it, as long as rivals offer to pay more than the defendant's incremental cost of dealing, the rivals will be entitled under the sacrifice test to share the fruits of the defendant's commercial success.[36] By contrast, Professor Salop argues that the sacrifice test is not well specified and might entitle the defendant to refuse to deal at prices on which dealing would increase welfare.[37]

None of these views seems compelling. The suggestion that the sacrifice test is especially well suited to refusal to deal cases might make sense from a property rights perspective in which special rules are sought for a firm's refusal to share its property with rivals. But it does not make sense as a matter of antitrust policy.[38] From the perspective of the defendant, all conduct (other than a refusal to deal) that passes the sacrifice test increases static efficiency. As will be seen, however, a refusal to deal at a price that exceeds the incremental cost of dealing can both reduce static efficiency and pass the sacrifice test. In addition, refusal to deal cases often raise difficult questions about the terms on which dealing might be required. Thus, far from being especially suited for the sacrifice test, refusals to deal involve unique complications.

B. Application of the Sacrifice Test to Refusals to Deal

The fallacies of the three other criticisms can be seen by explaining how the sacrifice test is properly applied to unilateral refusals to deal.[39] The key is to think of a refusal to deal as a make-or-buy decision.

Assume the defendant is a monopoly supplier of widgets and makes end products—call them "gidgets"—for which widgets are a necessary input. He sells the gidgets for $100. He also sells widgets in a competitive market for other uses for $25. A plaintiff, call him "Trinko," concludes that he can manufacture gidgets at a cost of less than $75 (say, $65) plus the cost of the widgets. So, he buys widgets for $25, makes gidgets, and

---

36. *See* Elhauge, *supra* note 19, at 275-76.

37. *See* Salop, *supra* note 28, at 39-45.

38. The Supreme Court's decision in *United States v. Colgate & Co.*, 250 U.S. 300 (1919), is sometimes said to establish a property right to refuse to deal. But the Court actually said that the antitrust laws "do[] not restrict the long recognized right of trader or manufacturer . . . freely to exercise his own independent discretion as to the parties with which he will deal" only "[i]n the absence of any purpose to create or maintain a monopoly." *Id.* at 307. The Court thus made clear that any such property right is tempered by antitrust policy.

39. The Federal Circuit's decision in *CSU, L.L.P. v. Xerox Corp.* is analyzed more extensively in A. Douglas Melamed & Ali M. Stoeppelwerth, *The* CSU *Case: Facts, Formalism and the Intersection of Antitrust and Intellectual Property Law*, 10 GEO. MASON L. REV. 407 (2002).

sells them for $90. Consumers are better off. The defendant reacts by refusing to sell to Trinko and banning transshipping from his other widget customers. The refusal to deal plainly reduces consumer welfare in a static sense. Prices would be lower and output greater if Trinko could buy widgets at $25 and sell gidgets for $90.

It might be suggested that, in these circumstances, antitrust law requires the defendant to sell widgets to Trinko at $25. But antitrust law rejects that conclusion as a general matter because the law permits firms to enjoy the fruits of their "skill, foresight and industry," including market power, in order to preserve ex ante incentives for innovation and entrepreneurship.[40] If the defendant had never sold widgets for $25, antitrust law would plainly not require him to sell widgets at that price to Trinko, even if Trinko were able to demonstrate that $25 exceeded the incremental costs incurred by defendant in selling widgets to Trinko. Otherwise, firms would rarely, if ever, be able to charge monopoly prices because would-be rivals would routinely be able to buy at a price at or above incremental cost and resell, in competition with the supplier, at less than a monopoly price.

The fact that the defendant has sold to others at $25 should not change this result. To be sure, defendant's decision to sell widgets for $25 in a different market does give the court a basis to know that $25 is a remunerative price.[41] But it says little about whether welfare is, on balance, enhanced by requiring the defendant to sell at that price in the gidget market, in which the widget might be worth far more.

This does not mean that there are no antitrust limits on the defendant's ability to refuse to sell widgets to Trinko. Instead, the sacrifice test thinks of defendant as having made a make-or-buy decision regarding the other—non-widget—inputs needed for the gidget. If the defendant is more efficient with respect to these inputs than Trinko, he will likely refuse to deal because self-manufacture is the optimal way to make gidgets. That decision is efficient and should be permitted by the antitrust laws, even though it would enable the defendant to charge a monopoly price for the gidgets.

---

40. *See, e.g.*, Verizon Communications, Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 407 (2004) (noting that the rewards of monopoly create incentives to invest and "induce[] risk taking that produces innovation and economic growth").

41. Without benchmarks offered by prior dealing, however, the court might be unable to determine the price at which dealing should be required or even whether dealing should be required. *See generally id.* at 408 (stating that courts are "ill-suited" as a general matter to determine "the proper price, quantity and other terms of dealing").

But, if Trinko is more efficient, the defendant would ordinarily maximize his profits, not by making gidgets himself, but by in effect buying the other (non-widget) inputs from Trinko. He might literally buy the other inputs from Trinko pursuant to some kind of contractual arrangement. Often, however, it will be more efficient for the parties to structure the transaction in a different form; instead of buying the other inputs, the defendant would sell his widgets to Trinko and let Trinko make the gidgets.

If he chose this option, the defendant would sell the widgets to Trinko at a monopoly price. There is no reason to expect the monopoly price for widgets when used for gidgets to be the $25 that the monopolist charges when he sells the widgets used for other purposes in a competitive market. Indeed, in this example, assuming that the monopoly price of gidgets is $100, the monopoly price of widgets used for gidgets is $35.[42] Antitrust law permits the defendant to charge such a monopoly price for the widget (or, in the example described in the preceding paragraph, for the gidget) in order to provide appropriate ex ante incentives both for the development of the widget and for optimal make-or-buy decisions regarding manufacture of gidgets.

Suppose, however, that the defendant passes up this more efficient and profitable alternative in order to gain market power through exclusion of Trinko—perhaps by monopolizing gidget manufacturing and thereby raising the entry barriers protecting his widget monopoly by making two-level entry necessary.[43] In that event, the refusal would entail choosing a less efficient means of making gidgets in order to gain market power.[44] The refusal would thus be anticompetitive because it would be unprofitable and would make no business sense, except as a means of creating or maintaining market power.

---

42. To avoid a double marginalization problem that would arise if Trinko became the monopoly supplier of gidgets, the defendant would probably hope to find multiple efficient gidget manufacturers. Failing that, the transactions between the defendant and Trinko would likely become more complicated. This Article ignores these complications because they are not material to the antitrust issue addressed here.

43. This would be a rational strategy only if, among other things, an entrant could not expect to achieve efficient scale in the manufacture of widgets if he were able to sell the widgets only in the competitive market for uses other than in the manufacture of gidgets.

44. For a discussion of the several kinds of circumstances in which a monopolist might have an incentive to pass up the most efficient source of complements, see Joseph Farrell & Philip J. Weiser, *Modularity, Vertical Integration, and Other Access Policies: Towards a Conveyance of Antitrust and Regulation in the Internet Age*, 17 HARV. J.L. & TECH. 85, 105-19 (2003).

The sacrifice test is, therefore, coherent and readily applicable to refusal to deal cases. It permits firms to earn monopoly profits from their assets and thus adequately preserves ex ante incentives. Furthermore, it both prohibits inefficient refusals to deal and provides a sound, principled basis for rejecting claims of those who seek access to others' property where such dealing is not efficient in the sense described above.

VI. CONCLUSION

The sacrifice test as a sensible middle ground between the more interventionist market-wide balancing tests and the less interventionist approach urged by those who are concerned about the uncertainties, false positives, and perverse incentives that such tests can generate. The sacrifice test is easier for both courts to administer ex post and firms to comply with ex ante than market-wide balancing tests and promises more effective antitrust deterrence of welfare-reducing conduct than more laissez-faire approaches.

The relative superiority of the sacrifice test varies depending on the type of conduct at issue. Naked exclusion—conduct that excludes rivals and has no efficiency benefits—can be readily condemned without any complicated balancing or sacrifice test. In most instances, such conduct in any event would violate the sacrifice test because its only reward would be the resulting market power.

The sacrifice test is unambiguously superior to market-wide balancing in all cases in which the conduct does not raise rivals' costs, but rather excludes rivals only by reducing the defendant's prices or costs and/or increasing the value of his goods or services. In these cases, balancing tests require calculation of the welfare costs caused by the defendant's conduct in a subsequent "recoupment" period, after the rivals have exited or been weakened. That calculation is likely to be beyond the capability of courts, and it is certainly beyond the capability of firms trying to decide ex ante whether their contemplated conduct is permissible. Any balancing test that requires such a calculation fails to give adequate guidance to firms.[45]

Balancing tests could be streamlined in raising rivals' costs cases to avoid some of these problems by ignoring the long-run or dynamic welfare costs of market power and, instead, calculating only the overall static welfare effects of the conduct in light of its benefits to the defendant and

45. I have elsewhere called such balancing tests "dynamic market-wide balancing." *See* Melamed, *supra* note 21 (manuscript at 8-9).

its harm to rivals.[46] But while the calculation required by such static balancing would be less daunting than that required for dynamic balancing, it is still likely in most cases to be much more difficult for the defendant than that required by the sacrifice test. This is because the calculation would require the defendant to estimate the impact of the conduct, both on himself and his customers and also on his rivals and their customers. Moreover, such a test would not completely eliminate the false negatives from the sacrifice test, and adoption of such a test would entail two additional costs: the costs of complicating antitrust jurisprudence by having a special rule for raising rivals' costs cases; and the cost imposed on firms that would be required, simply in order to know what legal rule applies to their conduct, to determine whether the conduct will materially increase their rivals' costs.

In refusal to deal cases, a balancing test would have the additional complication of requiring calculation of the costs to innovation incentives and dynamic efficiency of a duty to deal under the circumstances. A test that required such a calculation would plainly not be administrable by courts or firms. The sacrifice test avoids this complication by incorporating the ordinary antitrust presumption that the dynamic benefits of encouraging innovation outweigh the costs of permitting firms to charge monopoly prices for their lawfully obtained monopolies.

The sacrifice test can provide a sound unifying antitrust principle for analyzing all exclusionary conduct that has efficiency benefits. The sacrifice test is superior to balancing tests for at least most kinds of exclusionary conduct, and it can avoid most of the false positives and uncertainties that motivate those who argue that antitrust cannot constructively deal with problems of exclusion.

---

46. I have called such abbreviated tests "static market-wide balancing." *See id.* (manuscript at 12-13).